Cross-phaseogram: Objective neural index of speech sound differentiation

Erika Skoe[a,b], **Trent Nicol**[a,b], and **Nina Kraus**[a,b,c,+]

[a]Auditory Neuroscience Laboratory (www.brainvolts.northwestern.edu)

[b]Department of Communication Sciences and Disorders; Northwestern University

[c]Departments of Neurobiology & Physiology; Otolaryngology; Northwestern University

Abstract

We introduce a new approach, the cross-phaseogram, which captures the brain's ability to discriminate between spectrotemporally dynamic speech sounds, such as stop consonants. The goal was to develop an analysis technique for auditory brainstem responses (ABR) that taps into the sub-millisecond temporal precision of the response but does not rely on subjective identification of individual response peaks. Using the cross-phaseogram technique, we show that time-varying frequency differences in speech stimuli manifest as phase differences in ABRs. By applying this automated and objective technique to a large dataset, we found these phase differences to be less pronounced in children who perform below average on a standardized test of listening to speech in noise. We discuss the theoretical and practical implications of our results, and the extension of the cross-phaseogram method to a wider range of stimuli and populations.

Keywords

auditory brainstem response; speech; phase; subcortical; speech-in-noise perception

1.0 Introduction[1]

The ability to distinguish speech sounds is a fundamental requirement for human verbal communication. Speech discrimination relies critically on the time-varying features of the signal (Shannon et al., 1995), such as the rapid spectrotemporal fluctuations that distinguish the stop consonants [b], [d] and [g]. Because these fine-grained acoustic differences fall within very brief temporal windows, stop consonants — sounds created by a momentary stop and then rapid release of airflow in the vocal tract — are vulnerable to misperception

[+]Corresponding author Nina Kraus, Ph.D.
Addresses: Erika Skoe, Northwestern University, 2240 Campus Drive, Evanston IL, 60208 USA, eeskoe@northwestern.edu
Trent Nicol, Northwestern University, 2240 Campus Drive, Evanston IL, 60208 USA, tgn@northwestern.edu
Nina Kraus, Ph.D., Northwestern University, 2240 Campus Drive, Evanston IL, 60208 USA, nkraus@northwestern.edu, Phone: 847-491-3181, Fax: 847-491-2523

[1]Abbreviations: AM (amplitude modulation), AN (auditory nerve), cABR (auditory brainstem response to complex sounds), CF (characteristic frequency), F0 (fundamental frequency), F1 (first formant), F2 (second formant), F3 (third formant), HINT (Hearing in Noise Test), SIN (speech in noise)

(Tallal, 2004), especially in background noise (Nishi et al. 2010) The susceptibility of these, and other sounds with rapid acoustic transitions, to confusion can often lead to frustration during communication in noisy environments (e.g., busy restaurants and street corners), even for normal hearing adults. Developing methods to access the biological mechanisms subserving phonetic discrimination is vital for understanding the neurological basis of human communication, for discovering the source of these misperceptions, and for the assessment and remediation of individuals with listening and learning disorders. Here we introduce one such method, the cross-phaseogram, which compares the phase of the auditory brainstem response (ABR) evoked to different speech sounds as a function of time and frequency. By being objective and automated, and by producing results that are interpretable in individual subjects, this method is a fundamental advance in both scientific and clinical realms.

The ABR is a far-field electrophysiological response recorded from scalp electrodes that reflects synchronous activity from populations of neurons in the auditory brainstem (Chandrasekaran and Kraus, 2009; Skoe and Kraus, 2010). In its current clinical applications, ABRs provide objective information relating to hearing sensitivity and general auditory function (Hood, 1998; Picton, 2010). While ABRs are elicited traditionally using rapidly presented clicks or sinusoidal tones (Hall, 2007), the use of complex sounds (e.g., speech syllables, musical notes, etc.) to elicit ABRs has revealed the precision with which temporal and spectral features of the stimulus are preserved in the response. Importantly, auditory brainstem responses to complex sounds (cABR) have helped to expose the dynamic nature of the auditory system, including its malleability with language and musical experiences, its relationship to language ability (e.g., reading and the ability to hear speech in noise), as well as its potential to be altered by short-term auditory remediation (reviewed in (Kraus and Chandrasekaran, 2010; Krishnan and Gandour, 2009). For the cABR to be adopted broadly by the scientific and clinical communities, objective and convenient analysis methods must be developed.

Brainstem timing is exceptionally precise, with deviations on the order of microseconds being clinically relevant (Hood, 1998). Therefore quantifying the temporal properties of the cABR is central to all auditory brainstem analyses. The timing of brainstem activity has been analyzed traditionally by evaluating the latency of individual response peaks, defined as the time interval between the onset of sound and the elicited peak. Peak identification can present a technical challenge because it is subjective, time consuming, and because it relies on assumptions as to which peaks are relevant for a given stimulus. Moreover, in clinical groups, ABR waveform morphology is often poorly defined. Together, these factors hamper the clinical utility of cABRs, especially when multiple peaks must be identified and/or when responses to multiple complex stimuli are compared.

In light of these factors and the potential clinical applicability of recent findings from our laboratory (Hornickel et al., 2009; Johnson et al., 2008), we were motivated to develop a more objective method for extracting temporal information from cABRs. Here we apply our new method in the analysis of cABRs recorded to the stop consonant syllables [ga], [da], and [ba]. These speech sounds are distinguished by the trajectory of their second formant (F2) during the initial 50 ms of the sounds ([ga]: 2480-1240 Hz; [da]: 1700-1240 Hz, [ba]: 900–1240 Hz). Because phase locking in the brainstem is limited to a relatively narrow range (<~1500 Hz) (Aiken and Picton, 2008; Liu et al., 2006), much of the phonemic information, including the frequencies differentiating these voiced stop consonants, is outside these limits. However, as predicted from the tonotopicity of the auditory system and ABRs to tone bursts (Gorga et al., 1988), the F2 of these speech sounds is captured in the timing of the response, with the peak latency systematically decreasing with increases in stimulus frequency. Johnson *et al.* (2008) and Hornickel *et al.* (2009) found that the [ga]-

[da]-[ba] formant trajectories are preserved in the timing of multiple individual peaks of the cABRs, with [ga] response peaks occurring first, followed by [da] and then [ba] peaks (Figure 1). This timing pattern ([ga] < [da] < [ba]) is most apparent at four discrete peaks within the response to the formant transition. Timing differences are initially large, but diminish over time as the stimuli approach the steady-state region corresponding to the sustained [a] sound. Through a labor-intensive manual identification process, Hornickel *et al.* (2009) found that the [ba]-[da]-[ga] subcortical timing pattern is less pronounced, or even absent in children with language-related disabilities, including those who have difficulty listening to speech in a noisy background.

An alternative method for describing brainstem timing is to quantify frequency-specific phase information from the cABR (Gorga et al., 1988; John and Picton, 2000). In digital signal processing, phase can be extracted from the output of the discrete Fourier transform, along with the amplitude of each frequency. For an oscillatory signal, phase denotes where the waveform is in its cycle relative to an arbitrary zero (sine or cosine) or relative to a second signal. For example, two sine waves can be described as being either in phase with each other or out of phase by up to 360 degrees or 2π radians. If signal 1 is further in its cycle than signal 2 (e.g., 180 vs. 0 degrees), it *phase leads* signal 2. A simple analogy for this is two race cars driving at the same speed (analog of frequency) but with staggered starting positions around a race track (analog of phase). The cars will pass by the same tree but at different time points, with car 1 arriving before car 2 (car 1 phase leads car 2). If the cars were driving at different speeds, then their phase would not be comparable. In auditory neuroscience, phase has been used primarily as a surrogate of the mechanical movement of the basilar membrane (van der Heijden and Joris, 2003), or to determine the stability of the neural response over the course of an experiment or the synchronization between brain regions (Weiss and Mueller, 2003).

Based on the latency and phase measurements made by John and Picton (2000) for brainstem responses to amplitude modulated sinusoids, we hypothesized that the timing patterns glimpsed in the time-domain cABR waveforms to [ga], [da], and [ba] reflect underlying differences in response phase. By deconstructing responses into their component frequencies and extracting phase on a frequency-by-frequency basis, we expected to uncover continuous phase differences between pairs of responses (i.e., [ga] vs. [ba], [ga] vs. [da], and [da] vs. [ba]) during the response to the formant transition region and an absence of phase differences during the response to the steady-state region, a region where the stimuli are identical. In accordance with the pattern observed among individual peaks, we predicted that [ga] would phase lead both [da] and [ba], with [da] also phase leading [ba]. Given that brainstem phase locking cannot be elicited to frequencies above ~1500 Hz we also expected to observe differences in response phase at frequencies below this cutoff, even though the stimulus differences are restricted to the range of the F2s of the stimuli (900–2480 Hz, Figure 1A). Based on Hornickel *et al.*'s (2009) findings, along with other evidence that the cABR provides a neural metric for speech-in noise (SIN) perception (Anderson et al., 2010a; Anderson et al., 2010b; Kraus and Chandrasekaran, 2010; Parbery-Clark et al., 2009; Song et al., 2010); (Hornickel et al., 2010) we also predicted that phase differences between responses to different stimuli would be minimized in children who perform below average on a standardized SIN task.

Our predictions are tested here in two separate experiments. In Experiment 1, we demonstrate the cross-phaseogram method in a single subject and show that it can be used to objectively measure differences in subcortical responses to different speech sounds (in this case contrastive stop consonants). In Experiment 2, we apply this method to a dataset of 90 children. By using a large dataset, we evaluate the cross-phaseogram method on its potential

to provide rapid and objective measurements of response differences both across a set of speech stimuli and between two groups defined by their SIN perception ability.

2.0 Materials and methods

2.1 Methods common to experiments 1 and 2

Electrophysiological procedures and stimulus design followed those used by Hornickel *et al.*, 2009.

2.1.1 Stimuli—Because the cABR is sensitive to subtle acoustic deviations (Skoe and Kraus, 2010), we used synthetic speech to create stimuli that were identical except for the trajectory of the second formant. By isolating the stimulus differences to a single acoustic cue, something that is not possible with natural speech, we aimed to capture how the nervous system transmits this particular stimulus feature, a feature that is important for differentiating stop consonants.

The syllables [ba], [da], and [ga] were synthesized using a cascade/parallel formant synthesizer (SENSYN speech synthesizer, Sensimetrics Corp., Cambridge MA) to differ only during the first 50 ms, during which each had a unique frequency trajectory for the second formant. The stimuli were otherwise identical in terms of their duration (170 ms), voicing onset (at 5 ms), fundamental frequency (F0), as well as their first (F1) and third-sixth formants (F3–F6). The F0 was constant throughout the syllable (100 Hz) as were F4–F6 (3300, 3750, and 4900 Hz, respectively). During the formant transition region (0–50 ms), F1–F3 changed in frequency as a function of time: F1 rose from 400–720 Hz and F3 fell from 2850 to 2500 Hz. Across the three syllables, F2 began at a different frequency at time zero but converged at the same frequency at 50 ms (1240 Hz), reflecting a common vowel. The F2 trajectories were as follows: [ga] 2480-1240 Hz (falling); [da] 1700-1240 Hz (falling); [ba] 900–1240 Hz (rising). A schematic representation of the F2 formant trajectories is found in Figure 1A. During the steady state region associated with [a], all formants were constant (F1=720 Hz, F2=1240 Hz, F3 = 2500 Hz, F4 = 3300 Hz, F5 = 3750 Hz and F6 = 4900 Hz).

2.1.2 Electrophysiology and stimulus presentation—Auditory brainstem responses were collected in Neuroscan Acquire 4.3 (Compumedics, Inc., Charlotte, NC) with a vertical montage (active electrode placed at Cz, reference on the earlobe ipsilateral to ear of stimulus presentation, with the ground electrode on the forehead) at a 20 kHz sampling rate. Electrode impedances were kept below five kOhms.

Using Stim[2] (Compumedics, Inc.), stimuli were presented to the right ear at 80 dB SPL through an insert earphone (ER-3, Etymotic Research, Inc., Elk Grove Village, IL). The stimuli [ga], [da] and [ba] were presented pseudo-randomly along with five other syllables that had different temporal and/or spectral characteristics. Stimuli were presented using the alternating polarity method in which a stimulus and its inverted counterpart (shifted by 180 degrees) are played in alternating fashion from trial to trial. By averaging responses to the two stimulus polarities, it is possible to minimize contamination by the cochlear microphonic and stimulus artifact which both invert when the stimulus is inverted (see Aiken and Picton, 2008 and Skoe and Kraus, 2010 for more information). In addition, this adding process emphasizes the envelope-following component of the cABR (Aiken and Picton, 2008).

During testing, participants sat comfortably in a reclining chair in a sound attenuating room and viewed a movie of their choice. The movie sound track, which was set to <=40 dB SPL,

was audible to the left ear. This widely-employed passive collection technique enables the subject to remain awake yet still during testing and it facilitates robust signal to noise ratios.

2.1.3 Data reduction—A bandpass filter (70–2000 Hz, 12 dB/oct) was applied to the continuous EEG recording using Neuroscan Edit (Compumedics, Inc.) to isolate activity originating from brainstem nuclei (Chandrasekaran and Kraus, 2009). Averaging was performed over a 230-ms window to capture neural activity occurring 40-ms before through 190-ms after the onset of the stimulus. Separate subaverages were created for 3000 artifact-free trials (trials whose amplitudes fell within a −35 to +35 μV range) of each stimulus polarity; subaverages were then subsequently added.

2.1.4 Cross-phaseograms—Cross-phaseograms were generated in MATLAB 7.5.0 (Mathworks, Natick, MA) by applying the cross-power spectral density function (cpsd function in MATLAB) in a running-window fashion (20-ms windows) to each response pair ([ga] vs. [ba], [ga] vs. [da], and [da] vs. [ba]). The cpsd function, which is based on Welch's averaged periodogram method, was chosen for its ability to reduce noise from the estimated power spectrum. In total, 211 windows were compared (per response pair): the first window began at −40 ms (40 ms before the onset of the stimulus) and the last window began at 170 ms, with 1 ms separating each successive 20 ms window. Before applying the cpsd function, windows were baseline corrected to the mean amplitude (detrend function) and response amplitudes were ramped on and off using a 20-ms Hanning window (hann function; 10-ms rise and 10-ms fall). For each of the 211 comparisons, the cpsd function produced an array of numbers, representing the estimated cross-spectral power (squared magnitude of the discrete Fourier transform) of the two signals as a function of frequency (4 Hz resolution). These power estimates are the result of averaging eight modified periodograms that were created by sectioning each window into eight bins (50% overlap per bin). The process of averaging these eight individual periodograms reduces noise in the power estimate by limiting the variance of the power output. To obtain phase estimates for each window, power estimates were converted to phase angles (angle function), with jumps greater than π (between successive blocks) being corrected to their $2*\pi$ complement (unwrap function).

By concatenating the phase output of the 211 bins, a three dimensional representation of phase differences was constructed (cross-phaseogram), with the x-axis representing time (reflecting the midpoint of each time window), the y-axis representing frequency, and the third dimension, plotted using a color continuum, reflecting phase differences between the pair of signals being compared. If the two signals do not differ in phase at a particular time-frequency point, this is plotted in green. If signal 1 is further in its phase cycle than signal 2, this is plotted in warm colors, with red indicating the greatest difference. Cool colors indicate the opposite effect of signal 2 being further in its phase cycle than signal 1.

2.2 Methods for experiment 2

2.2.1 Participants—Ninety children (52 males, ages 8–13; mean, 10.93 years; s.d = 1.5) were tested as part of an ongoing study in our laboratory. Study inclusion criteria include normal audiometric thresholds (for octave intervals from 250 to 8000 Hz), no history of neurological disorders, normal cognitive function (as measured by the Wechsler Abbreviated Scales of Intelligence, The Psychological Corporation, San Antonio, TX), and normal click-evoked ABRs. Northwestern University's Institutional Review Board (IRB) approved all experimental procedures. In accordance with IRB guidelines, children provided informed assent and parents/guardians gave their informed consent.

2.2.2 Behavioral measures—Participants were divided into two groups based on their performance on HINT (Hearing in Noise Test, Bio-logic Systems, a Natus corporation,

Mundelein, IL), a widely-used clinical test that evaluates speech perception in noise (Nilsson et al., 1994). HINT measures the accuracy with which sentences can be correctly repeated as the intensity of the target sentences changes relative to a fixed level of background noise (speech-shaped noise, 65 dB SPL). The signal-to-noise ratio (SNR) is adjusted adaptively until 50% of the sentences are repeated correctly. The target sentences and noise are presented in free field from a single loudspeaker placed one meter in front of the participant. Target sentences were constructed to be phonetically balanced, using simple vocabulary and syntactic structure that are appropriate for children at the first grade level. SNRs were converted to percentiles using age-appropriate norms.

Our population of children demonstrated a wide range of HINT scores (min = 0th percentile, max = 100th percentile, mean = 46.055, s.d., = 32.935, interquartile range = 62.000, skew = 0.117, kurtosis = −1.33). Following the procedures employed in Anderson *et al.*, 2010a, participants performing at or above the 50th percentile were categorized as TOP performers ($n = 40$, 23 males (57%)), and those performing below average, we categorized as BOTTOM performers ($n = 50$; 30 males (60%)). The two HINT groups did not differ statistically in terms of the latency of the click-ABR wave V, a gauge of peripheral hearing and central auditory pathway integrity (Hood, 1998) (TOP mean = 5.704 ms, s.d. = 0.185; BOTTOM mean = 5.682 ms, s.d. = 0.182; $t_{89} = 0.55$, $P = 0.58$).

2.2.3 Statistical Analyses—To compare phase shifts between the two HINT-based groups and across the three stimulus pairings, the cross-phaseogram matrices were split into two time regions (formant transition: 15–60 ms; steady state: 60–170 ms) and three frequency ranges (low: 70–400 Hz, middle: 400–720 Hz, high: 720–1100 Hz) based on the stimulus and recording parameters, as well as the qualitative appearance of the cross-phaseograms (Supplemental Figure 2). Each time region was analyzed separately in a $2 \times 3 \times 3$ analysis of variance (ANOVA) in SPSS (SPSS Inc., Chicago, IL), using 1 between subject factor (Group: TOP vs. BOTTOM HINT performers) and 2 within-subjects factors (RANGE: 70–400 vs. 400–720 vs. 720–1100 Hz, CONTRAST: [ga] vs. [ba], [da] vs. [ba], [ga] vs. [da]).

3.0 Results

Auditory brainstem responses were recorded to three 170-ms stop consonant speech syllables ([ga], [da], and [ba]) using scalp-electrodes. Responses were compared by calculating the phase coherence between pairs of responses as function of time and frequency. Cross-phaseograms represent the phase shift between the two signals in a time-frequency plot, with color signifying the extent of the shift. In all cases, the response to the stimulus with the higher F2 served as the first member of the pair ([ga] vs. [ba]; [da] vs. [ba]; [ga] vs. [da]). Consequently, phase shifts were expected to have positive values during the formant transition period but be near zero during the steady-state region. Separate analyses were performed for the response to the formant transition (15–60 ms) and steady-state vowel (60–170 ms) components of the stimulus.

3.1 Experiment 1: Demonstrating the cross-phaseogram method on an individual subject

As predicted from the peak timing analyses performed by Hornickel *et al.* (2009) using identical stimuli, there are clear cross-response phase differences during the formant transition. This is illustrated in Figures 2 and 3 for an individual subject (male, age 12) with normal audiometry (pure tone average of 1.6 dB HL) and language ability. In accordance with the stimulus characteristics, the cross-phaseograms indicate that the response to [ga] phase leads the other responses, with [da] also having phase-lead over [ba]. Cross-response

phase differences are minimal during the response to the steady-state vowel portion of the syllables where the stimuli are identical.

3.2 Experiment 2: Applying the cross-phaseogram method to a large dataset

In Experiment 2, cross-phaseograms were calculated on a large dataset ($n = 90$) to evaluate the prediction that cross-response phase shifts are reduced in children (mean = 10.93 years) who perform below average on a clinical test of speech perception in noise (the Hearing in Noise Test (HINT) (Nilsson et al., 1994)). Children performing at or above the 50th percentile on HINT were assigned to the TOP group ($n = 40$; mean percentile = 78.71, s.d. = 15.73), and those performing below this threshold were assigned to the BOTTOM group ($n = 50$; mean percentile 21.15; s.d. = 16.52). For the formant transition and steady state analyses, we first report within-subjects comparisons to validate the results in Experiment 1, and then proceed to the between-subjects comparisons to compare the two HINT groups. All phase-shift values are reported in radians.

3.2.1 Formant transition (15–60 ms)

3.2.1.1 Within-subjects comparisons: When collapsing across the entire dataset, and all three frequency ranges, we find unique phase shift signatures for each response pairing ($F_{2,176} = 3.875$, $P = 0.023$). The extent of the phase shifts is in agreement with the degree of stimulus differences: the greatest phase shifts are found for the [ga]-[ba] pairing (mean = 0.317 radians, s.e. = 0.040), and the smallest are found for the [ga]-[da] pairing (mean = 0.208, s.e. = 0.028), with the [da]-[ba] pairing falling in the middle (mean = 0.288, s.e. = 0.031) (Figure 4).

Given the non-uniform nature of the phase differences across the frequency spectrum, average phase-shift values were calculated for three ranges: 70–400, 400–720 and 720–1100 Hz. A strong main effect of frequency range was observed ($F_{2,176} = 42.011$, $P < 0.0009$), with the middle range having the largest cross-response phase shifts (mean = 1.305, s.e. = 0.112), and the lower and higher frequency ranges having smaller, yet non-zero phase shifts (mean = 0.474, s.e. = 0.070; mean = 0.601, s.e. = 0.092, respectively). After correcting for multiple comparisons, the middle range is statistically different from the high and low ranges ($P < 0.0009$, in both cases), whereas the high and low ranges do not differ statistically from each other (Figures 4 and 5). The phase variables all found to follow a normal distribution (as measured by the Kolmogrov Smirnov test of Normality; $P>0.15$ for all variables)

3.2.1.2 Between-subjects comparisons: In line with our predictions, children performing above average on the hearing in noise test have more distinct brainstem responses to contrastive speech syllables compared to those performing below average (Figures 4 and 5) ($F_{1,88} = 6.165$, $P = 0.013$; TOP mean = 0.328, s.e. = 0.035; BOTTOM mean = 0.213, s.e. = 0.031). The group differences are not driven by one stimulus pairing ($F_{2,176} = 1.970$, $P = 0.143$). This extent of these differences is also not equivalent across different frequency bands ($F_{2,176} = 3.323$, $P = 0.038$); When collapsing across stimulus pairings, the groups are most different in the lowest ($t_{88} = 3.299$, $P = 0.001$; TOP mean = 0.7196, s.e. = 0.096; BOTTOM mean = 0.2772, s.e. = 0.092) and mid-frequency bands ($t_{88} = 2.480$, $P = 0.015$; TOP mean = 1.607, s.e. = 0.137; BOTTOM mean = 1.063, s.e. = 0.163), and least different during the highest frequency band ($t_{88} = 0.260$, $P = 0.796$; TOP mean = 0.629, s.e. = 0.129; BOTTOM mean = 0.581, s.e. = 0.130). Taken together these findings indicate that the stimulus contrasts are more robust over a wider range of frequencies in the TOP group's responses, suggesting that acoustic differences might be represented in a more redundant fashion in the neural responses of children who have better performance on a speech in noise task.

3.2.2 Steady-state region (60–170 ms)

3.2.2.1 Within-subject comparisons: While there were no differences between the three frequency ranges ($F_{2,176} = 2.559$, $P = 0.085$) during the response to the steady-state vowel, a significant main effect of stimulus pairing was found ($F_{2,176} = 7.967$, $P < 0.0009$). This finding, which is inconsistent with our initial predictions, indicates that responses do differ during the steady-state region despite the stimuli being identical. However, because the phase differences during the steady-state region are close to zero, this significant effect may be the consequence of extremely small variances ([ga] vs [ba] mean = −0.055, s.e. = 0.009; [da] vs. [ba] mean = −0.014, s.e. = 0.009; [ga] vs. [da] mean = −0.021, s.e. = 0.007) and the non-normal distribution (Kolmogrov-Smirnov test of Normality; $P<0.05$ for all variables except the [ga] vs. [da] comparison in the low frequency range). Interestingly, the average cross-response phase shifts are all negative, indicating that they occur in the opposite direction from what would be predicted if they represented a "*bleed over*" from the formant-transition period.

3.2.2.2 Between-subject comparisons: As predicted, the two HINT groups do not differ during the steady-state region ($F_{1,88} = 0.981$, $P = 0.325$; TOP mean = −0.024, s.e. = 0.009 ; BOTTOM mean = −0.036, s.e. = 0.008).

3.3 Summary

Our initial predictions were largely validated when we compared the cross-phaseograms to different response pairings and when we compared the cross-phaseograms between the TOP and BOTTOM HINT groups. Compatible with Purcell *et al.*'s work (Purcell et al., 2004), we found that the phase of the cABR tracked the linearly ramping frequency trajectory of the stimuli. In comparing response pairs, we found that frequency differences between stimuli manifest as phase-differences, with the higher frequency stimuli producing responses that phase lead responses to the lower frequency stimuli. Moreover, we found that children, who have difficulty hearing in noise, as measured by HINT, tend to have smaller phase shifts between responses than children who perform above average on HINT.

4.0 Discussion

We describe procedures that permit large-scale analysis of auditory brainstem responses to spectrotemporally complex sounds such as speech. We interpret our finding as a *proof of concept* that cross-phaseogram analysis can provide an objective method for accessing the subcortical basis of speech sound differentiation in normal and clinical populations. Moreover, unlike traditional peak picking methods, the cross-phaseogram is a highly efficient method for assessing timing-related differences between two responses. Using the cross-phaseogram algorithm, we show that the time-varying frequency trajectories that distinguish the speech syllables [ba], [da] and [ga] are preserved in the phase of the auditory brainstem response. Furthermore, we provide evidence that stimulus contrasts tend to be minimized in the brainstem responses of children who perform below average on a standardized test of listening to speech in noise but have normal hearing thresholds.

4.1 Technological and theoretical advance

While the measurement of neural phase has long been employed in the field of auditory neuroscience (Weiss and Mueller, 2003), our approach constitutes a technological advance on multiple fronts. To the best of our knowledge, this is the first time a phase-coherence technique has been used to analyze ABRs to spectrotemporally-dynamic acoustic signals. In so doing, we have developed a method for tracking how the brainstem represents (or fails to represent) minute changes in stimulus features, including those that are important for distinguishing different speech contrasts. While this kind of cross-response comparison was

previously possible (Hornickel et al., 2009); Johnson et al., 2008), clinicians and scientists now have access to this information in an objective and automated manner without the need to subjectively identify individual response peaks. In addition to being eminently more timesaving than manual peak picking methods, the use of the cross-phaseogram may also help to streamline data collection time by limiting the number of stimulus presentations that need to be collected in order to perform reliable temporal analysis on the cABR. Improving the efficiency of the recording and analysis procedures is essential for clinical applications of cABRs.

While the cross-phaseogram method is intended to supplement and not supplant existing analysis techniques, it also offers advantages over other automated techniques. For example, cross-correlation methods (e.g., stimulus-to-response correlation and response-to-response correlation) including running-window versions (see Skoe and Kraus, 2010), offer a more limited view into how nervous system represents different stimulus contrasts. Although they can reveal the extent to which two signals are correlated and the time-shift at which the maximum correlation is achieved, unlike the cross-phaseogram method they do not provide frequency-specific information. Also, because of the non-linear way in which the formant frequencies are captured in the cABR, the stimulus-to-response correlation may, therefore, be a less successful method for tracking how the formant transition manifests in the response. Using the cross-phaseogram method, we discovered that the formant frequencies (in the range of 900–2480 Hz) that distinguish [ba], [da], and [ga] are "transposed" to lower frequencies in the response, as evidenced by frequency-dependent phase-shifts below 1100 Hz in the three stimulus pairings we examined. This provided a unique insight into the non-linear nature of auditory brainstem processing, a finding that may not have been revealed by other methods such as cross correlation.

Although inter-trial phase coherence has been proposed as an objective method for assessing the functional integrity of the auditory brainstem (Clinard et al., 2010; Fridman et al., 1984; Picton et al., 2002), this study offers a technological and theoretical advance by being the first to link phase-based ABR measurements to the ability to listen to speech in a noisy background. By showing that the cross-phaseogram can be used to investigate the biological underpinning of an everyday listening skill, the cABR has the potential to become a valuable measure of higher-level language processing, much in the same way that inter-electrode cortical phase coherence and phase tracking are now viewed (Howard and Poeppel, 2010; Luo and Poeppel, 2007; Nagarajan et al., 1999; Weiss and Mueller, 2003).

4.2 Origins of the neural phase delays

The phase shifts that are evident in the auditory brainstem response to different stop consonants likely reflect the differences that exist between stimuli in the amplitude and phase spectra (Supplemental Figure 1) as well as frequency-dependent neural phase delays resulting from the mechanical properties of the cochlea.

Due to the tonotopic organization of the basilar membrane, traveling waves for high frequency sounds reach their maximum peak of excitation at the base of the cochlea with lower frequencies causing the greatest basilar membrane displacement apically. As such, auditory nerve (AN) fibers innervating the basal end of the cochlea are activated before those at the apex (Greenberg et al., 1998). Consequently, the timing (and accordingly the phase) of the AN response will be different depending on the fiber being stimulated, leading to greater phase delays for fibers with lower characteristic frequencies (CFs). Far field neurophysiological response timing (and consequently phase), in addition to being influenced by the transport time to the characteristic location along the basilar membrane, is impacted by a constellation of peripheral and central factors including: (a) the acoustic delay between the sound source and the oval window; (b) active cochlear filtering; (c) the synaptic

delay associated with activating afferent fibers; and (d) the conduction delay between the origin of the neural response and the scalp (John and Picton, 2000; Joris et al., 2004).

The frequency-dependent timing pattern of single AN nerve fibers is propagated to higher stages of the auditory pathway. As a result, frequency-dependent latency and phase shifts are evident in the scalp-recorded auditory brainstem response (as shown here). This effect has been previously demonstrated for high-frequency tone bursts (Gorga et al., 1988), amplitude modulated tones (John and Picton, 2000) and speech syllables (Hornickel et al., 2009; Johnson et al., 2008). An expansion of the frequency-dependent cochlear delay is also observed for cortical potentials, resulting in greater phase delays than those initially introduced peripherally. However, the extent of this additional delay, which is also seemingly dependent on the spectrotemporal complexity and predictability of the stimulus structure, is highly variable between subjects (Greenberg et al., 1998; Patel and Balaban, 2004). Based on work in an animal model, cortical feedback may also modulate human subcortical response timing modulated via the extensive network of corticofugal fibers (reviewed in (Suga et al., 2000). Thus, the individual phase-shifts differences reported here may reflect strengthened or weakened reciprocal connections between subcortical and cortical structures that modulate the auditory system's ability to boost stimulus contrasts (Hornickel et al., 2009).

4.3 Accounting for the low frequency phase delays

While the factors delineated above account for the existence of phase shifts in the cABR, they do not explain why such phase shifts emerge in the response at frequencies well below those of the second formant (900–2480 Hz), the acoustic feature that differentiated the stop consonant syllables. This transposition likely reflects (1) the use of suprathreshold levels, resulting in a spread of excitation across AN nerve fibers of differing CFs and (2) the sensitivity of the auditory system to amplitude modulations (AMs) in the speech signal. As the result of the opening and closing of the vocal cords, the amplitude of the speech formants (spectral maxima arising from the resonance properties of the vocal tract) is modulated at the rate of the fundamental frequency (in our case 100 Hz) and its harmonics (Aiken and Picton, 2008; Cebulla et al., 2006). These amplitude envelope modulations, which are linked to the perception of voice pitch, result from the time-domain interaction of different speech components and they are, as such, not represented in the energy spectrum of the speech signal (Joris et al., 2004). However, the amplitude modulation (AM) of the speech signal is relayed by the auditory system through phase locking that is evident in ANs as well as brainstem, thalamic and cortical structures (Joris et al., 2004).

Work by John and Picton offers insight into how AM phase locking can transmit higher frequency information. Using amplitude modulated sinusoids, John and Picton (2000) have modeled, in elementary form, how brainstem activity represents the interaction of high frequencies (in the range of the F2s used here) and low pitch-related amplitude modulations (John and Picton, 2000). They found that the envelope response entrained to the frequency of modulation but that the phase of the response was dictated by the carrier frequency. For example, response timing was earliest for an 80 Hz modulation frequency when the carrier signal was 6000 Hz, with the timing systematically increasing as the carrier frequency decreased to 3000, 1500 and 750 Hz. Thus, while human brainstem nuclei do not "phase lock," in the traditional sense of the word, to frequencies above ~1500 Hz (Aiken and Picton, 2008; Liu et al., 2006), brainstem activity conveys the time-varying information carried by high frequency components of the stimuli, such as the differing F2s in the syllables used in this study, through the phase of the envelope response to low frequencies. Here we demonstrate this phenomenon for the first time using spectrotemporally complex sounds.

4.4 Accounting for group differences in neural phase

What then might account for minimized stimulus contrasts in some children? Because all of the children in this study had normal audiometric thresholds and normal click-evoked ABRs, neurophysiological differences between groups cannot be explained by differences in gross cochlear processes. Instead, the diminished contrasts between cABR to different speech sounds could reflect reduced neural synchrony and/or diminished sensitivity to amplitude modulations (see (Nagarajan et al., 1999)). The diminished cross-response phase contrasts in these children may help to explain why noise imposes a greater effect on them. In addition, our results indicate that stimulus contrasts are more redundantly represented in the cABRs of children who perform above average on HINT, such that phase-shifts occur across a broader spectrum of frequencies in this group. This redundancy could enable high speech intelligibility for this group of children, especially when the speech signal is masked by noise.

Based on animal models (Suga et al., 2000; Suga et al., 2002; Xiong et al., 2009) and prevailing theories of linguistic-related brainstem plasticity in humans (Hornickel et al., 2009; Krishnan and Gandour, 2009), we argue that abnormal auditory brainstem responses to speech reflects both the malfunction of local processes in the brainstem as well as a maladaptive feedback between subcortical and cortical structures. In this theoretical framework, auditory brainstem activity can *influence* and *be influenced* by cortical processes. Consequently, diminished representation of stimulus contrasts in the auditory brainstem could result in poor SIN performance and likewise poor SIN performance could retroactively weaken how stimulus contrasts are coded in the auditory brainstem. However, the currently available cABR collection methods do not permit us to test the separate or combined influences on top-down and bottom-up process on auditory brainstem processing and their relationship to SIN perception.

While we cannot make causal inferences based on our methodology, our findings do reinforce the proposition that listening in noise depends, in part, on how distinctively the nervous system represents different speech sounds (Hornickel et al., 2010; Hornickel et al., 2009). However, no single factor can explain SIN performance. This is because listening in noise is a highly complex auditory process that depends on the interaction of multiple sensory and cognitive factors (see for example (Anderson and Kraus, 2010; Hornickel et al., 2010; Shinn-Cunningham and Best, 2008). In the case of HINT, a test that requires the listener to repeat back entire sentences, working memory (Parbery-Clark et al., 2009) and attention (Shinn-Cunningham and Best, 2008) undoubtedly impact performance. In addition, HINT performance may also depend on a child's ability to utilize semantic cues (such as semantic predictability) to overcome the degraded sensory input (Bradlow and Alexander, 2007). Future work will help to reveal extent to which phase-based measurements can uniquely predict behavioral measurements of SIN performance (see (Hornickel et al., 2010) and how these relationships might differ across populations.

4.5 Future directions

We have demonstrated the utility of the cross-phaseogram technique by employing it in our analysis of auditory brainstem responses obtained from a large cohort of children. However, the current application of the cross-phaseogram is also just one illustration of how the method could be used to study auditory processing. Based on the results reported here, we envision the successful application of the cross-phaseogram in the study of any population or individual demonstrating impaired or exceptional auditory abilities (e.g., the hearing impaired, musicians).

These phase-based methods are appropriate for analyzing subcortical responses to a great variety of other spectrotemporally complex sounds including natural speech, music, and environmental sounds. For example, given that background noise is known to induce delays in the neural response, the cross-phaseogram can complement existing techniques for comparing neural responses to speech in noisy versus quiet listening conditions, which rely heavily on the identification of response peaks (Anderson et al., 2010a; Parbery-Clark et al., 2009). Moreover, because background noise obscures the temporal features of the response, especially the small response peaks that distinguish cABRs to different stop consonants the cross-phaseogram may provide an avenue for studying the neural correlates of speech sound differentiation in noise. Traditional peak picking methods prohibit or at best hamper this kind of analysis.

4.6 Conclusions

As an objective and automated technique, the cross-phaseogram method is positioned to be a clinically viable metric in the audiological assessment and remediation of noise-induced perceptual difficulties and other language disorders. This cross-phaseogram technique, which can be used in the evaluation of a large dataset, can also provide the kind of individual-specific information that is essential for such clinical applications but is often inaccessible with other non-invasively employed neural metrics. While future efforts should focus on optimizing the cross-phaseogram for clinical applications, we have made an important first step in the translational process.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

REFERENCES

Aiken SJ, Picton TW. Envelope and spectral frequency-following responses to vowel sounds. Hear Res 2008;245:35–47. [PubMed: 18765275]

Anderson S, Kraus N. Objective neural indices of speech-in-noise perception. Trends Amplif 2010;14:73–83. [PubMed: 20724355]

Anderson S, Skoe E, Chandrasekaran B, Kraus N. Neural timing is linked to speech perception in noise. J Neurosci 2010a;30:4922–4926. [PubMed: 20371812]

Anderson S, Skoe E, Chandrasekaran B, Zecker S, Kraus N. Brainstem correlates of speech-in-noise perception in children. Hear Res. 2010b

Bradlow AR, Alexander JA. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. J Acoust Soc Am 2007;121:2339–2349. [PubMed: 17471746]

Cebulla M, Sturzebecher E, Elberling C. Objective detection of auditory steady-state responses: comparison of one-sample and q-sample tests. J Am Acad Audiol 2006;17:93–103. [PubMed: 16640063]

Chandrasekaran B, Kraus N. The scalp-recorded brainstem response to speech: Neural origins and plasticity. Psychophysiology 2009:236–246. [PubMed: 19824950]

Clinard C, Tremblay K, Krishnan A. Aging alters the perception and physiological representation of frequency: Evidence from human frequency-following response recordings. Hearing Research 2010;264:48–55. [PubMed: 19944140]

Gorga MP, Kaminski JR, Beauchaine KA, Jesteadt W. Auditory brainstem responses to tone bursts in normally hearing subjects. J Speech Hear Res 1988;31:87–97. [PubMed: 3352259]

Greenberg, S.; Poeppel, D.; Roberts, T. A Space-Time Theory of Pitch and Timbre Based on Cortical Expansion of the Cochlear Traveling Wave Delay. In: Palmer, QS.; Rees, A.; Meddis, R., editors. Psychophysical and Physiological Advances in Hearing. London: Whurr Publishers; 1998. p. 293-300.

Hall, JW. New Handbook of auditory evoked responses. Boston: Allyn and Bacon; 2007.

Hood, LJ. Clinical applications of the auditory brainstem response. San Diego: Singular Pub. Group; 1998.

Hornickel J, Chandrasekaran B, Zecker S, Kraus N. Auditory brainstem measures predict reading and speech-in-noise perception in school-aged children. Behav Brain Res 2010;216:597–605. [PubMed: 20826187]

Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N. Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. Proc Natl Acad Sci U S A 2009;106:13022–13027. [PubMed: 19617560]

Howard MF, Poeppel D. Discrimination of Speech Stimuli Based On Neuronal Response Phase Patterns Depends On Acoustics But Not Comprehension. J Neurophysiol. 2010

John MS, Picton TW. Human auditory steady-state responses to amplitude-modulated tones: phase and latency measurements. Hear Res 2000;141:57–79. [PubMed: 10713496]

Johnson KL, Nicol T, Zecker SG, Bradlow AR, Skoe E, Kraus N. Brainstem encoding of voiced consonant--vowel stop syllables. Clin Neurophysiol 2008;119:2623–2635. [PubMed: 18818121]

Joris PX, Schreiner CE, Rees A. Neural processing of amplitude-modulated sounds. Physiol Rev 2004;84:541–577. [PubMed: 15044682]

Kraus N, Chandrasekaran B. Music training for the development of auditory skills. Nat Rev Neurosci 2010;11:599–605. [PubMed: 20648064]

Krishnan A, Gandour JT. The role of the auditory brainstem in processing linguistically-relevant pitch patterns. Brain Lang 2009;110:135–148. [PubMed: 19366639]

Liu LF, Palmer AR, Wallace MN. Phase-locked responses to pure tones in the inferior colliculus. J Neurophysiol 2006;95:1926–1935. [PubMed: 16339005]

Luo H, Poeppel D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 2007;54:1001–1010. [PubMed: 17582338]

Nagarajan S, Mahncke H, Salz T, Tallal P, Roberts T, Merzenich MM. Cortical auditory signal processing in poor readers. Proc Natl Acad Sci U S A 1999;96:6483–6488. [PubMed: 10339614]

Nilsson M, Soli SD, Sullivan JA. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. J Acoust Soc Am 1994;95:1085–1099. [PubMed: 8132902]

Nishi K, Lewis DE, Hoover BM, Choi S, Stelmachowicz PG. Children's recognition of American English consonants in noise. The Journal of the Acoustical Society of America 2010;127:3177–3188. [PubMed: 21117766]

Parbery-Clark A, Skoe E, Kraus N. Musical experience limits the degradative effects of background noise on the neural processing of sound. J Neurosci 2009;29:14100–14107. [PubMed: 19906958]

Patel AD, Balaban E. Human auditory cortical dynamics during perception of long acoustic sequences: phase tracking of carrier frequency by the auditory steady-state response. Cereb Cortex 2004;14:35–46. [PubMed: 14654455]

Picton, TW. Human auditory evoked potentials. San Diego, CA: Plural Publishing Inc.; 2010.

Purcell DW, John SM, Schneider BA, Picton TW. Human temporal auditory acuity as assessed by envelope following responses. J Acoust Soc Am 2004;116:3581–3593. [PubMed: 15658709]

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science 1995;270:303–304. [PubMed: 7569981]

Shinn-Cunningham BG, Best V. Selective attention in normal and impaired hearing. Trends Amplif 2008;12:283–299. [PubMed: 18974202]

Skoe E, Kraus N. Auditory brain stem response to complex sounds: a tutorial. Ear Hear 2010;31:302–324. [PubMed: 20084007]

Song J, Skoe E, Banai K, Kraus N. Perception of Speech in Noise: Neural Correlates. J Cogn Neurosci. 2010

Suga N, Gao E, Zhang Y, Ma X, Olsen JF. The corticofugal system for hearing: recent progress. Proc Natl Acad Sci U S A 2000;97:11807–11814. [PubMed: 11050213]

Suga N, Xiao Z, Ma X, Ji W. Plasticity and corticofugal modulation for hearing in adult animals. Neuron 2002;36:9–18. [PubMed: 12367501]

Tallal P. Improving language and literacy is a matter of time. Nat Rev Neurosci 2004;5:721–728. [PubMed: 15322530]

van der Heijden M, Joris PX. Cochlear phase and amplitude retrieved from the auditory nerve at arbitrary frequencies. J Neurosci 2003;23:9194–9198. [PubMed: 14534253]

Weiss S, Mueller HM. The contribution of EEG coherence to the investigation of language. Brain Lang 2003;85:325–343. [PubMed: 12735948]

Xiong Y, Zhang Y, Yan J. The neurobiology of sound-specific auditory plasticity: a core neural circuit. Neurosci Biobehav Rev 2009;33:1178–1184. [PubMed: 19014967]

A   STIMULUS

B   NEURAL TIMING

C   NEURAL PHASE PREDICTIONS

ga *phase leads* ba   da *phase leads* ba   ga *phase leads* da
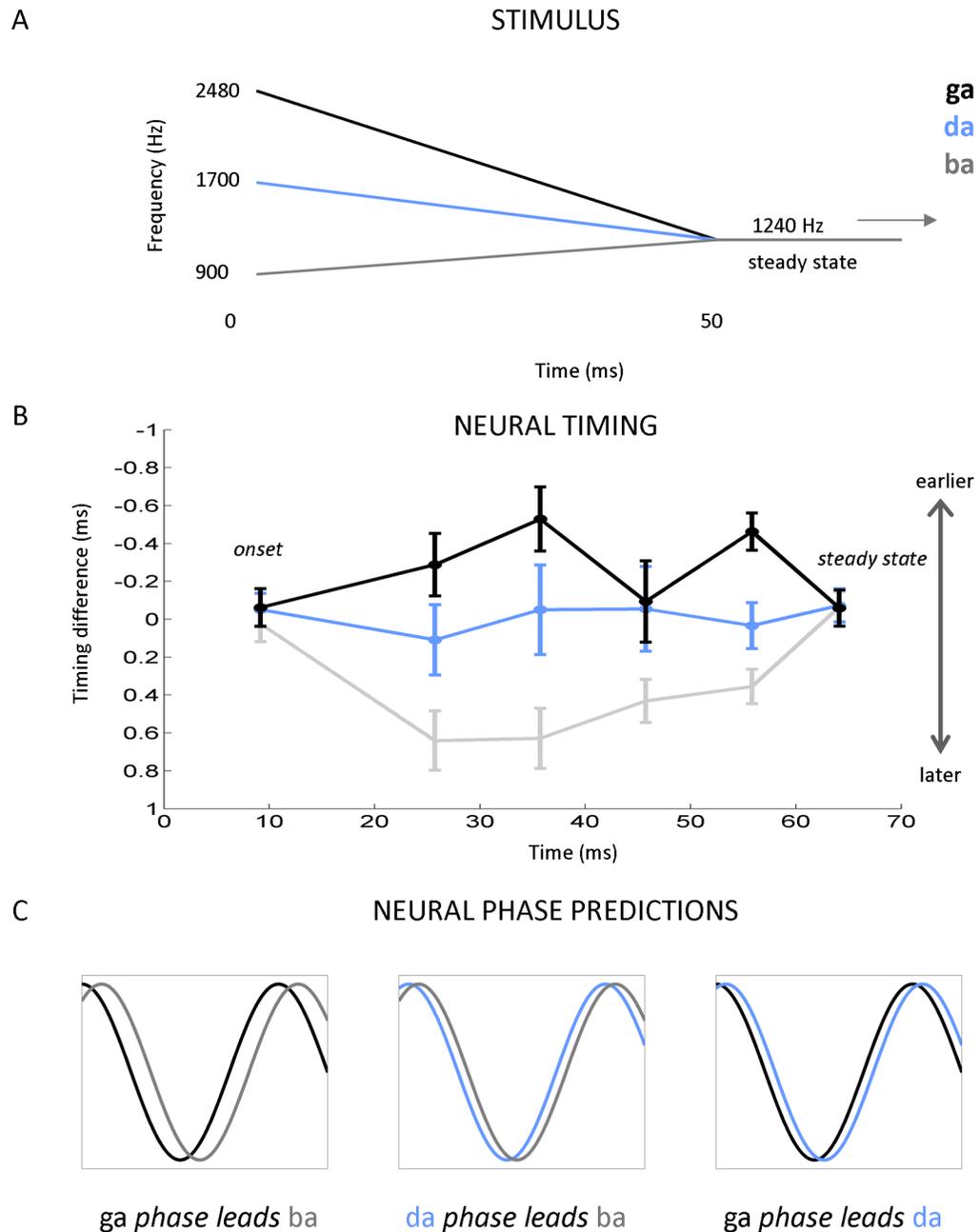
Figure 1. Current method for comparing auditory brainstem responses (ABRs) to different stop consonants syllables (A–B) and phase shift predictions (C)
The frequency differences that differentiate the stop consonants syllables [ga], [da], [ba] (A) are represented in the ABR by timing differences, with [ga] responses occurring first, followed by [da] and then [ba] (B) (i.e., higher frequencies yield earlier peak latencies than lower frequencies). This pattern is most apparent at four discrete response peaks between the onset and steady-state components of the response. However, no differences are observed during the onset response (at about 9 ms) and the response to the steady-state portions of the syllables where the stimuli are identical (beginning at 60 ms). Figure modified from Hornickel *et al.,* 2009. **(C)** Consistent with the pattern observed in the timing of the

individual peaks, we hypothesized that for a given frequency within the cABR, that the [ga] response would phase lead both the [da] and [ba] responses. We anticipated that the [da] response would phase lead [ba].
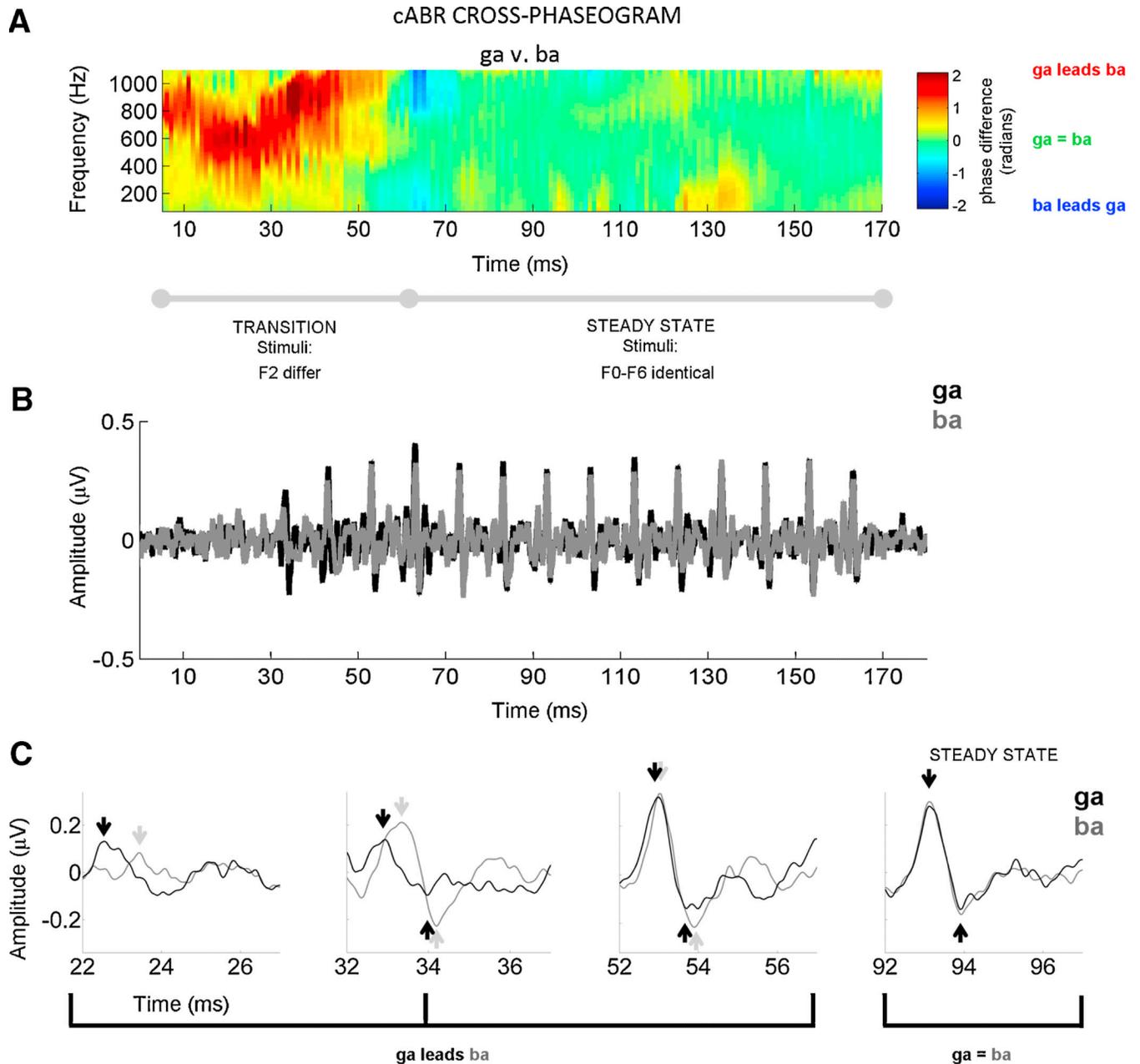
A

cABR CROSS-PHASEOGRAM

ga v. ba

TRANSITION
Stimuli:
F2 differ

STEADY STATE
Stimuli:
F0-F6 identical

B

C

Figure 2. Cross-phaseogram of auditory brainstem responses
A representative subject is plotted to illustrate that individual-subject comparisons are particularly accessible, a feature that makes the cross-phaseogram method clinically useful. This subject (male, age 12) had normal audiometry (pure tone average of 1.6 dB), normal language ability, and he performed at the 87.4th percentile on the Hearing in Noise Test. **(A)** Auditory brainstem responses to the speech sounds [ga] and [ba] are compared using the cross-phaseogram, a method which calculates phase differences between responses as a function frequency and time. In the cross-phaseogram, the time displayed on the x-axis refers to the midpoint of each 20 ms time bin. The y-axis represents frequency and the color axis represents the phase difference (in radians) between the response to [ga] and the response to [ba]. When the responses [ga] and [ba] are in phase, the plot appears green. When the response to [ga] leads in phase relative to [ba], this is represented using yellows,

oranges and reds, with dark red indicating the largest differences. However, when the converse is true (i.e., [ba] response leads [ga] response), the plot is represented with shades of blue, with dark blue indicating the greatest phase differences. As can be seen in this plot, phase differences between the responses are restricted to the formant transition region (15–60 ms). As predicted from Johnson *et al.*, 2008 and Hornickel *et al.*, 2009, the [ga] response phase leads [ba] during this time region. In the response to the steady-state (60 to 170 ms), the responses are almost perfectly in phase. **(B)** To enable comparisons between the phase and peak timing measurements, the time domain versions of the response waveforms are plotted ([ga] in black, [ba] in gray). In **(C)**, the responses plotted in (B) are magnified at four time points (centered at 24, 34, 54, and 94 ms) to illustrate the timing differences between stimuli that are evident during the response to the formant transition region, but not during the response to the steady state portions of the stimuli. Thus, the timing differences between responses (B–C) manifest as continuous phase differences across a range of frequencies (A).
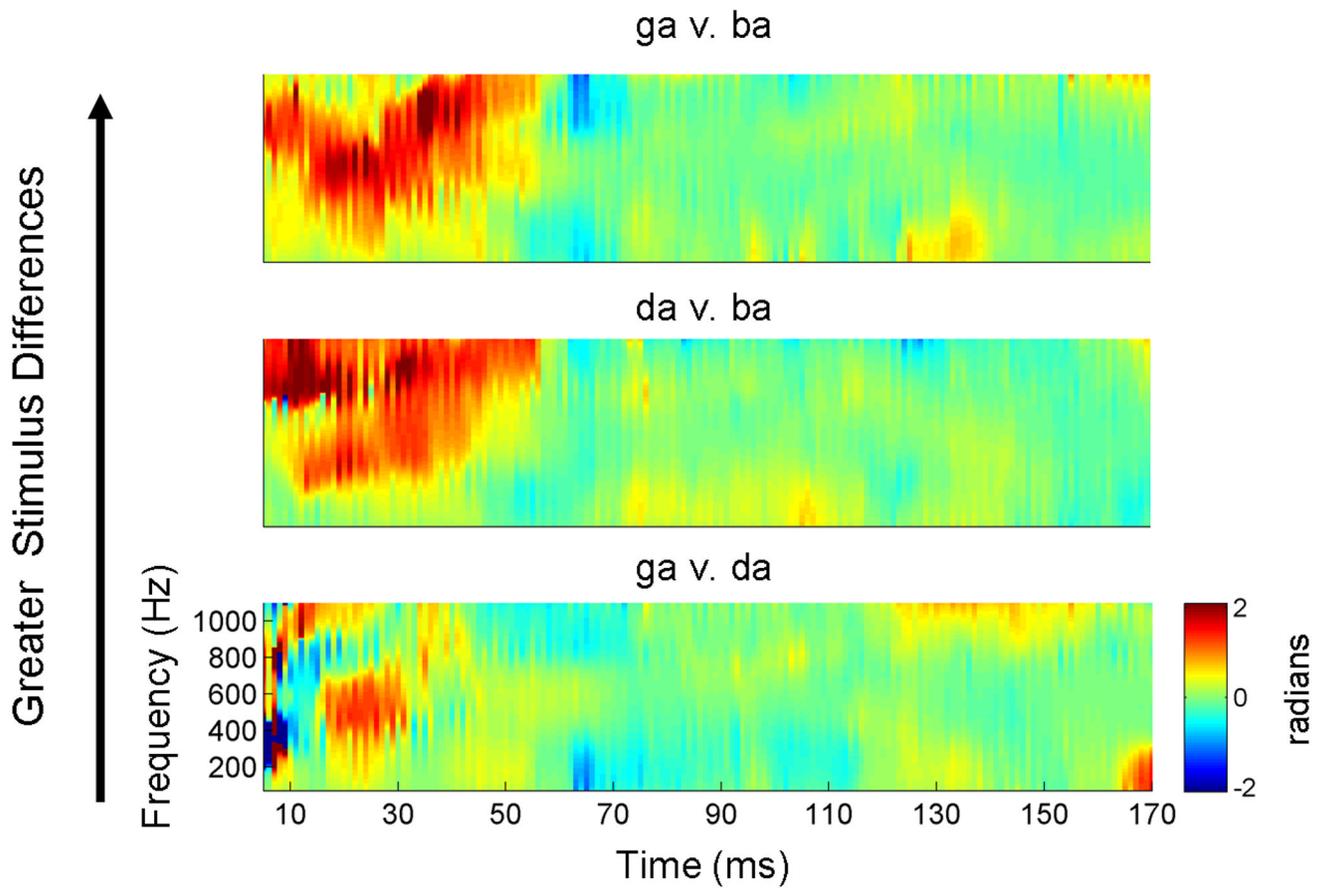
cABR CROSS-PHASEOGRAM

Figure 3. Response cross-phaseograms for the three stimulus contrasts
For the same representative subject as in Figure 2, the top plot compares the responses to the stimuli which are the most acoustically different ([ga] vs. [ba]) during the formant transition region. In the bottom plot, phase spectra are compared for responses to the stimuli that are most acoustically similar ([ga] vs. [da]). In accordance with these stimulus differences, greater phase differences are observed in the top plot, with more minimal differences evident in the bottom plot.
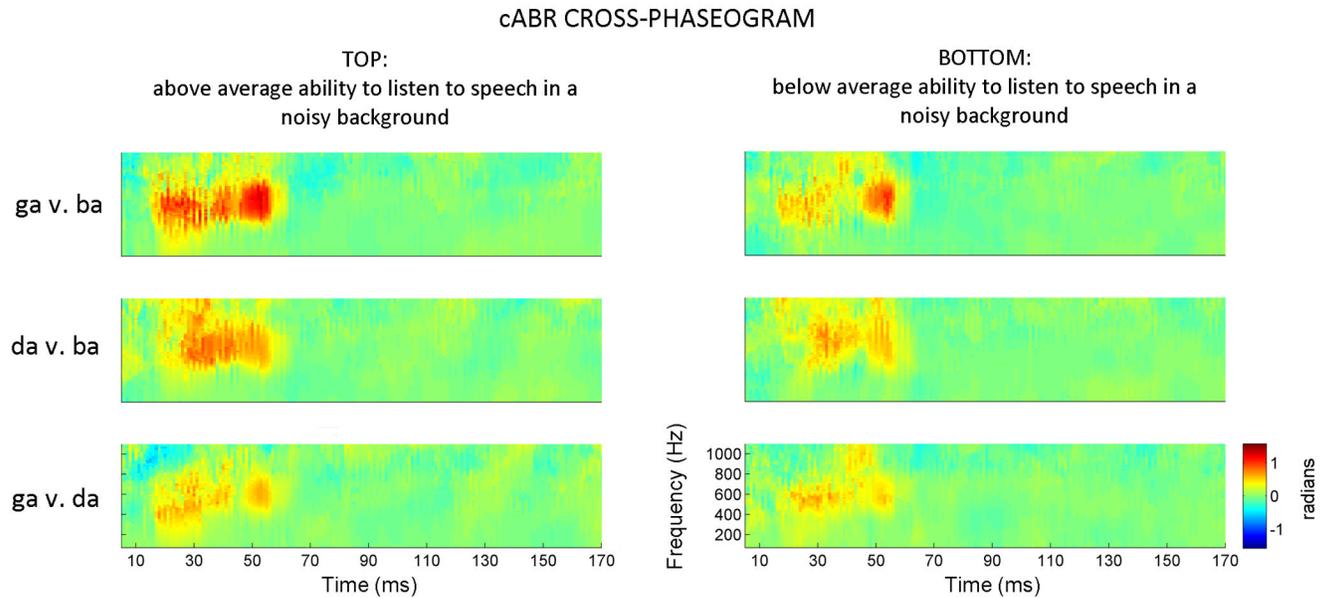
Figure 4. Applying the cross-phaseogram method to a large dataset (*n* = 90)
Children were grouped based on their performance on a standardized test that measures the ability to repeat sentences presented in a background of speech-shaped noise. Across all three stimulus comparisons, cross-response phase differences are more pronounced in the TOP performers (>50th percentile) compared to the BOTTOM performers. These plots represent average phaseograms for each group (i.e., the average of 40 and 50 individual phaseograms, respectively). Due to averaging, the phase differences are smaller in these averages compared to the individual subject plotted in Figures 2 and 3. The color axis has consequently been scaled to visually maximize phase differences. The pattern of effects (i.e., [ga] phase leading [ba], [ga] phase leading [da], and [da] phase leading [ba] during the formant transition region of the responses) is evident for both TOP and BOTTOM groups.

720-1100 Hz

400-720 Hz

70-400 Hz

Time (ms)
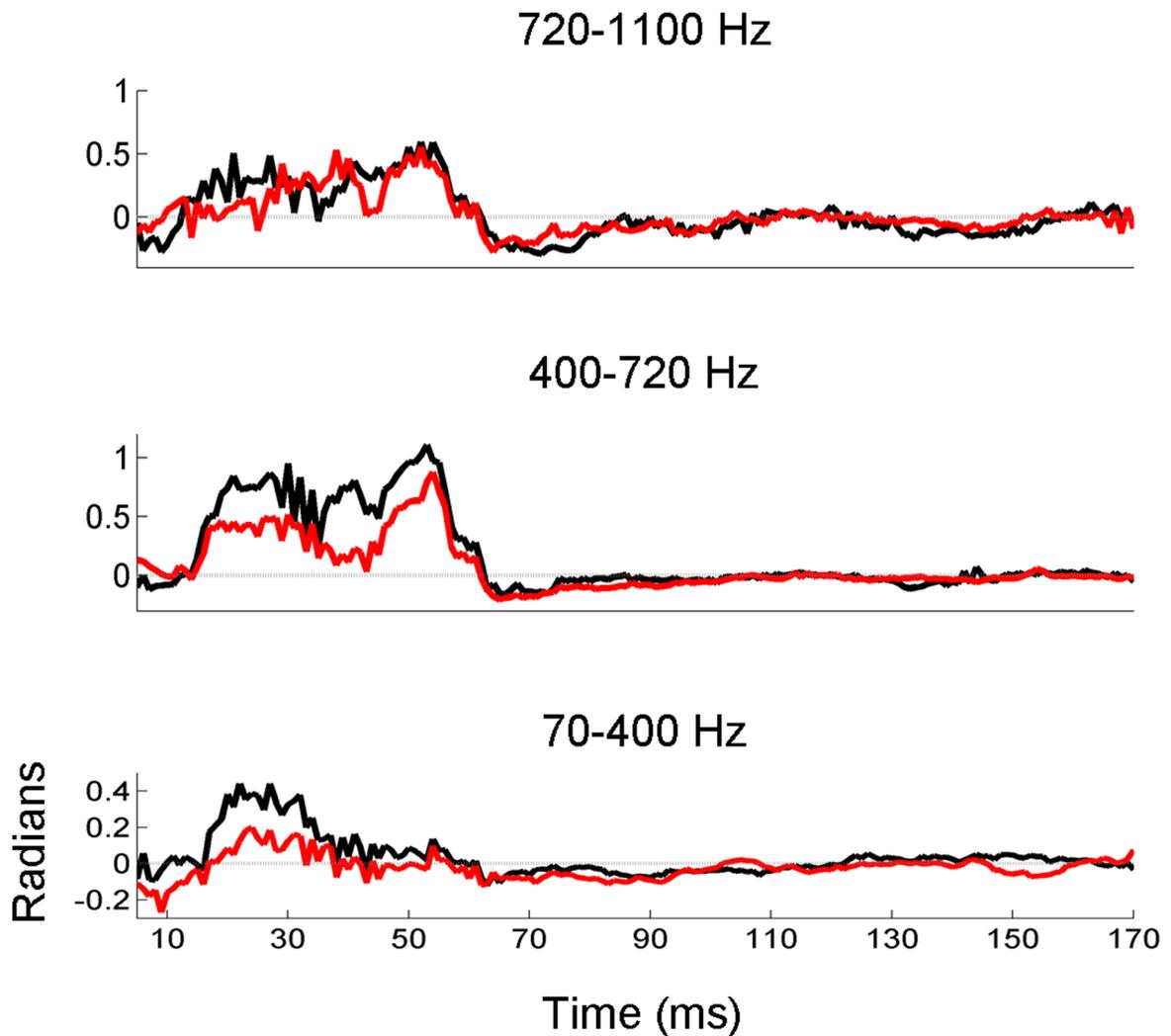
TOP: above average ability on HINT

BOTTOM: below average ability on HINT

Figure 5. Deconstructing the response cross-phaseograms into frequency bands
Average phase information is extracted over three frequency regions (A: 70–400 Hz; B: 400–720 Hz; C: 720–1100 Hz). Here the comparison between the responses to [ga] and [ba] is plotted for two groups of children, grouped according to their performance on a standardized test of speech perception in noise. Response for the children performing in the top 50th percentile are plotted in black, and the group performing below this cutoff are plotted in red. This figure illustrates greater phase separation among children who perform above average on a speech-in-noise task compared to those performing below average. For each row, the y-axis (radians) is scaled to highlight cross-response phase differences, with zero radians being demarcated by a gray line. As would be predicted from Figure 4,

responses to the steady-state portion of the consonant-vowel syllables (beginning ~60 ms) are characterized by cross-response phase differences that hover near zero radians. During the response to the formant transition region (15–60 ms), phase coherence deviates from zero, indicating that the responses differ in phase during this time region. Although the phase differences occur on different scales for each frequency range, within the formant transition region they occur in a consistent direction with the response to the higher frequency stimulus ([ga]) leading the response to the lower frequency stimulus ([ba]) in the pair.