

## Research paper

# Inferior colliculus contributions to phase encoding of stop consonants in an animal model

Catherine M. Warrier<sup>a,d,\*</sup>, Daniel A. Abrams<sup>a</sup>, Trent G. Nicol<sup>a</sup>, Nina Kraus<sup>a,b,c,d</sup>

<sup>a</sup>Auditory Neuroscience Laboratory, Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Frances Searle Building, 2240 Campus Drive, Evanston, IL 60208-3550, USA

<sup>b</sup>Department of Neurobiology & Physiology, Northwestern University, Evanston, IL, USA

<sup>c</sup>Department of Otolaryngology, Northwestern University, Evanston, IL, USA

<sup>d</sup>Hugh Knowles Center for Clinical and Basic Science in Hearing and its Disorders, USA

## ARTICLE INFO

## Article history:

Received 7 June 2011

Received in revised form

1 September 2011

Accepted 6 September 2011

Available online 16 September 2011

## ABSTRACT

The human auditory brainstem is known to be exquisitely sensitive to fine-grained spectro-temporal differences between speech sound contrasts, and the ability of the brainstem to discriminate between these contrasts is important for speech perception. Recent work has described a novel method for translating brainstem timing differences in response to speech contrasts into frequency-specific phase differentials. Results from this method have shown that the human brainstem response is surprisingly sensitive to phase differences inherent to the stimuli across a wide extent of the spectrum. Here we use an animal model of the auditory brainstem to examine whether the stimulus-specific phase signatures measured in human brainstem responses represent an epiphenomenon associated with far-field (i.e., scalp-recorded) measurement of neural activity, or alternatively whether these specific activity patterns are also evident in auditory nuclei that contribute to the scalp-recorded response, thereby representing a more fundamental temporal processing phenomenon. Responses in anaesthetized guinea pigs to three minimally-contrasting consonant-vowel stimuli were collected simultaneously from the cortical surface vertex and directly from central nucleus of the inferior colliculus (ICc), measuring volume conducted neural activity and multiunit, near-field activity, respectively. Guinea pig surface responses were similar to human scalp-recorded responses to identical stimuli in gross morphology as well as phase characteristics. Moreover, surface-recorded potentials shared many phase characteristics with near-field ICc activity. Response phase differences were prominent during formant transition periods, reflecting spectro-temporal differences between syllables, and showed more subtle differences during the identical steady state periods. ICc encoded stimulus distinctions over a broader frequency range, with differences apparent in the highest frequency ranges analyzed, up to 3000 Hz. Based on the similarity of phase encoding across sites, and the consistency and sensitivity of response phase measured within ICc, results suggest that a general property of the auditory system is a high degree of sensitivity to fine-grained phase information inherent to complex acoustical stimuli. Furthermore, results suggest that temporal encoding in ICc contributes to temporal features measured in speech-evoked scalp-recorded responses.

© 2011 Elsevier B.V. All rights reserved.

## 1. Introduction

Brainstem encoding of complex sounds provides a unique window into the human auditory system and its function. Studies investigating brainstem responses to speech have informed our understanding of normal and impaired auditory systems (Akhoun et al., 2008; King et al., 2002; Krishnan, 2002; Song et al., 2008), the plasticity of auditory processes (Kraus and Banai, 2007), auditory system development (Johnson et al., 2008a; Vander Werff et al., 2011), and how lifelong experiences with language and music mold the auditory system (Krishnan et al., 2005;

*Abbreviations:* ABR, Auditory brainstem response; ANOVA, Analysis of variance; CF, Characteristic frequency; CV, Consonant-vowel; F0, Fundamental frequency; F1 through F6, First through sixth formant; IC, inferior colliculus; ICc, central nucleus of the inferior colliculus; FFR, frequency-following response; F2, frequency of the second formant.

\* Corresponding author. Auditory Neuroscience Laboratory, Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Frances Searle Building, 2240 Campus Drive, Evanston, IL 60208-3550, USA. Tel.: +1 847 491 2465.

E-mail address: [cwarrier@northwestern.edu](mailto:cwarrier@northwestern.edu) (C.M. Warrier).

Musacchia et al., 2007). Despite the wealth of information that these studies have provided, an inherent limitation of human studies is their inability to provide detailed information about the underlying neural mechanisms contributing to scalp-evoked potentials during the processing of speech, music, and other biologically-important acoustical signals. A method that has been successful in informing our understanding of the origins of scalp-recorded activity in humans is probing near-field auditory function in an animal model of the auditory system using identical acoustical stimuli as those used in human studies (See Cunningham et al., 2002 for a review; King et al., 1999; Kraus et al., 1985, 1988, 1992, 1994a, 1994b). For example, this approach has yielded a deeper understanding of speech processing in the presence of background noise (Cunningham et al., 2002), auditory-based asymmetries (King et al., 1999), and the cortical basis of speech discrimination (McGee et al., 1996). Here, we use this approach to better understand the exquisite timing of the human brainstem in response to speech sounds by examining the temporal dynamics of neural activity measured in the guinea pig inferior colliculus, a brain structure known to contribute to the scalp-recorded response (Chandrasekaran and Kraus, 2010; Marsh et al., 1974; Smith et al., 1975).

Recent work has focused on how the fine spectro-temporal differences distinguishing stop consonants are encoded as timing differences in the human brainstem response (Johnson et al., 2008b; Skoe et al., 2011). The submillisecond timing differences that differentiate brainstem responses to various stop consonants are clinically relevant as slight brainstem timing deficits are associated with behavioral impairments for speech and language (Banai et al., 2009; Hornickel et al., 2009b; King et al., 2002). As a means of further refining these methods for quantifying brainstem temporal processing, Skoe and colleagues recently introduced an analysis that translates these small timing differences into phase differentials between responses to various stimuli (2011). They analyzed brainstem responses to three consonant-vowel (CV) speech syllables that differed by a single formant trajectory during the consonant-vowel transition period. The “cross-phase” method revealed that responses to the CVs with a higher second formant frequency (F2) “phase-lead” responses to those with a lower F2. The phase differences were most prominent during time regions corresponding to the contrasting frequency modulations in the syllable stimuli, and were limited during the steady state portion of the response which was identical across stimuli. While it is acknowledged that phase information in this context may not contribute significantly to auditory system or behavioral differentiation of these speech sounds, the cross-phase approach reveals meaningful information regarding subtle and reliable timing differences in the brainstem’s representation of consonant-vowel stimuli.

An important question is whether the stimulus-specific phase signatures measured at the scalp represent an epiphenomenon associated with far-field (i.e., scalp-recorded) measurement of neural activity, or alternatively whether these specific activity patterns are also evident in auditory nuclei that contribute to potentials measured at the scalp, thereby representing a more fundamental temporal processing phenomenon. Classical studies investigating the frequency-following response (FFR) indirectly addressed this question using pure tone stimuli. In previous work, near-field inferior colliculus (IC) responses showed an extremely similar phase relationship with the surface-recorded FFR in response to a range of pure tone stimuli (Marsh et al., 1974; Smith et al., 1975). Given the non-linear nature of the ascending auditory system, however, phase relationships between the IC and scalp in response to more complex auditory stimuli cannot be predicted by results using simple (i.e., pure tone) stimuli.

To more thoroughly investigate the processing of spectro-temporal patterns embedded in complex stimuli in the scalp-recorded response, as well as its relation to activity in nuclei which contribute to this response, the current study evaluates speech-evoked responses to consonant-vowel stimuli /ba/, /da/, and /ga/. Near-field responses recorded directly from the central nucleus of the inferior colliculus are compared to far-field responses that serve as an analogue to the human surface-recorded response. Phase differentials are used to quantify small timing differences between responses to the three stimuli using the cross-phase method (Skoe et al., 2011). Similarities between the two recording sites will inform the extent to which patterns of phase differences recorded from the scalp reflect a more general auditory processing mechanism evident in nuclei in the ascending auditory system. We made two general predictions. First, we predicted the far-field guinea pig surface responses would strongly resemble the scalp-recorded responses in humans presented with the same stimuli. Second, assuming a contribution of the inferior colliculus to the scalp-recorded responses (Chandrasekaran and Kraus, 2010; Marsh et al., 1974; Smith et al., 1975), we expected the near-field responses to contain response patterns similar to those seen at the surface.

## 2. Material and methods

The research protocol was approved by the Animal Care and Use Committee of Northwestern University, and all US ethical guidelines for laboratory animal welfare were followed (assurance number A3283-01).

### 2.1. Animal preparation

The experimental materials and procedures were similar to those reported previously (Abrams et al., 2011; Cunningham et al., 2002; McGee et al., 1996). Ten pigmented guinea pigs (7 female) between 346 and 803 g (average 549 g), were used as subjects. Animals were initially anesthetized with ketamine hydrochloride (100 mg/kg) and xylazine (8 mg/kg). Smaller supplemental doses (25 mg/kg ketamine; 4 mg/kg xylazine) were administered hourly or as needed throughout the rest of the experiment. Following the induction of anesthesia, the animal was mounted in a stereotaxic device, located in a sound-treated booth (IAC), for the duration of the experiment. Body temperature was maintained at 37.5 °C by using a thermistor-controlled heating pad (Harvard) on the guinea pig’s abdomen. Prior to surgery, normal hearing sensitivity was confirmed by auditory brainstem response (ABR). ABRs were elicited by a click stimulus at 70 and 40 dB, referenced to previously-established lab-internal guinea pig click ABR threshold norms. Electromyographic needle electrodes were inserted into skin midway between ears, on snout midway between eyes and nose, and into loose skin at neck, for non-inverting, inverting, and ground, respectively. Following confirmation of normal hearing, a rostral-caudal incision was made along the scalp surface and the tissue was retracted to expose the skull. Holes were drilled in the skull under an operating microscope. The dura was removed with a cautery to prevent damage to the recording electrode, and the cortical surface was coated with mineral oil.

### 2.2. Stimuli

The stop consonants /ba/, /da/ and /ga/ were synthesized using a Klatt speech synthesizer according to previously published specifications (Klatt, 1980; Skoe et al., 2011). Briefly, stimuli were constructed to be identical except for the trajectory of the second formant (F2) during the 50 ms formant transition portion

(Liberman et al., 1954). All stimuli were 170 ms in duration with a fundamental frequency (F0) of 100 Hz, had voicing onset at 5 ms, and contained six formants (F1–F6). F4–F6 were held constant throughout the duration of the stimuli (3300, 3750, and 4900 Hz, respectively). During the 50 ms formant transition, F1 rose from 400 to 720 Hz and F3 fell from 2850 to 2500 Hz, then each stayed constant for the remaining vowel portion of the stimulus (50–170 ms). F2 began at 900, 1700 and 2480 Hz for/ba/,/da/and/ga/respectively, and all converged at 1240 Hz at 50 ms, staying constant at 1240 Hz from 50 to 170 ms. Stimulus spectrograms highlighting F2 trajectory are presented in Fig. 1 (top). Frequency content of the three stimuli are presented in Fig. 1 (bottom), illustrating differences in the formant transition but not the steady state portion of the stimuli (left and right figures, respectively).

### 2.3. Neurophysiological recording

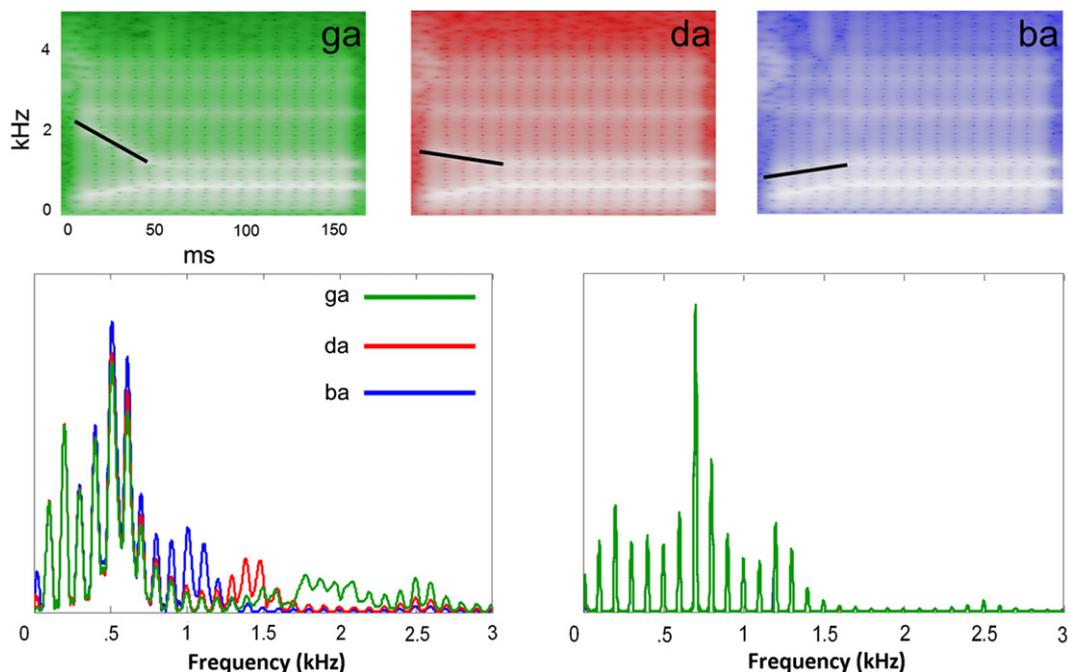
The ICc was accessed with a vertical approach using tungsten microelectrodes (Micro Probe) with impedance of approximately 2 M $\Omega$  at 1 kHz. An electrode was advanced perpendicular to the surface of cortex using a remote controlled micromanipulator (Märzhäuser-Wetzlar). For all recordings, the dorsal/ventral reference of the electrode was determined at a point slightly above cortex at the first penetration, and this coordinate was kept for the remainder of the experiment. ICc coordinates were approximately 0.3 mm caudal to the interaural line, 1.5 mm left of the sagittal suture and 4.0 mm ventral to the surface of the brain. For the surface recording, a superdural silver ball electrode was placed at the vertex 1 cm caudal to Bregma. The ground electrode was placed on the posterior scalp surface. During penetration, click stimuli (100  $\mu$ s rectangular pulses) were delivered at a rate of 3.5 Hz, and visual inspection, using a monitoring oscilloscope, of the response size and waveform morphology was considered. If the response was small in amplitude and broad in shape, electrode penetration was continued. This process was repeated until the morphology of the waveform conformed to the large amplitude, sharp onset response characteristic of recordings obtained from

the ICc, and location was further verified by comparing characteristics of responses to probe tones and noise to published response characteristics of ICc neurons (Liu et al., 2006; Rees and Palmer, 1988; Syka et al., 2000). Characteristic frequency (CF) of ICc placement was determined by presenting a series of low-intensity (30 dB HL) tones varying in frequency in third-octaves from 63 to 16000 Hz. These stimuli were 100 ms in duration with a rise-fall time of 10 ms. Fifty tones were presented at each frequency with a stimulus onset asynchrony of 110 ms. Recording sites were selected to have a CF in the range of the speech stimuli (100–6000 Hz).

For each penetration into the inferior colliculus, 150 presentations of each experimental stimulus were presented by a computer using MATLAB (Mathworks, Natick, MA), converted using a National Instruments D/A converter (National Instruments Corporation, Austin, TX) and delivered at 75 dBA using Etymotic ER2 earphones (Etymotic Research, Inc., Elk Grove Village, IL) through hollow ear bars. Stimulus onset asynchrony was 230 ms. Responses were amplified with a gain of 500 and filtered from 20 to 8000 Hz by two Grass P511 amplifiers (Grass Technologies, West Warwick, RI) and digitized by an MCC A/D board (Measurement Computing Corporation, Norton, MA) and saved to a second computer running a MATLAB acquisition program. The MCC board also received a trigger from the D/A converter marking stimulus onsets. Surface recordings comprised responses from 750 to 1500 stimulus presentations. Surface responses from one animal and two ICc penetrations were discarded from the analysis due to noise corruption, leaving 9 surface recordings and 81 ICc depth recordings.

### 2.4. Data processing

Using custom MATLAB scripts (Mathworks, Natick, MA), data from each recording (10 surface, 83 ICc) were epoched by response type from –40–190 ms, and responses to each stimulus type were averaged. Average responses were baseline corrected to the mean amplitude of the pre-stimulus period.



**Fig. 1.** Top row, left to right: Spectrograms of ga, da, and ba stimuli. F2 over the first 50 ms, the only differentiating acoustic features, are highlighted in black. Bottom row, left to right: Stimulus spectra up to 3 kHz from 0 to 50 ms and 50–170 ms. In the right panel, the peaks exactly coincide, resulting in a single uniformly-colored line.

## 2.5. Analysis

The cross-phase method was applied to the data as described in Skoe et al. (2011), and briefly described here. This method takes advantage of the fact that small timing differences in subcortical responses to complex stimuli are manifested as differences in the phase spectrum of the response. Phase differences between responses to two contrasting stimuli can be plotted on a time–frequency axis allowing visualization of how phase differences evolve across time at different frequencies. Phase coherence between each stimulus pair (/ga/vs./ba/(GaBa),/ga/vs./da/(GaDa), and/da/vs./ba/(DaBa)) was assessed by applying the cross-power spectral density (cpsd function in MATLAB) function in a running-window fashion (20 ms Hanning-ramped windows with 1 ms overlap, 211 total windows, 4 Hz frequency resolution). Data analysis was set up so that positive phase shift values indicate that the response to the stimulus with higher frequency content “leads” the response to the stimulus with lower frequency content. Thus, for example, when phases of the responses to /ga/and/ba/were compared, a positive number indicated that the/ga/response led the/ba/response. Phase angles with jumps greater than  $\pi$  radians were corrected to their  $2\pi$  complement (See John and Picton, 2000 for general discussion). An amplitude spectrum, via fast Fourier transform, was also computed on each of the 211 windows and compared to the amplitude spectrum of a 40 ms non-response window (i.e., the pre-stimulus period). Points where the signal to noise ratio ( $\text{amplitude}_{\text{response}}/\text{amplitude}_{\text{non-response}}$ ) was less than 1.5 were omitted from analysis and set to white in the color plots. The same procedure was performed on the stimulus waveforms for comparative purposes.

Average phase coherence of responses to the GaBa, GaDa and DaBa stimulus comparisons were plotted in cross-phaseograms, three dimensional representations of the response comparisons with time on the *x*-axis, frequency on the *y*-axis, and phase differences plotted in color. Phase coherence was extracted from 5 frequency bands of each cross-phaseogram: 70–400 Hz, 400–720 Hz, 720–1000 Hz, 1000–1500 Hz and 1500–3000 Hz. The lowest three bands correspond to the frequency bands used in the Skoe et al., 2011 analysis which relate to formant frequencies. The two highest frequency bands were chosen post-hoc based on observed response properties. Three time periods were assessed by averaging the phase data within each frequency region across the time period: Prestimulus period (–40–0 ms), Formant Transition (15–60 ms) and Steady State (60–170 ms). In order to assess the importance of CF, data from each recording obtained from the ICc was placed into one of four CF groups: Low CF (100–500 Hz), Mid CF (500–1000 Hz), High CF (1000–2000 Hz) and Higher CF (2000–6300 Hz). The ICc data were analyzed with a 4-way ANOVA, with within-recording measures of 3 stimulus comparisons (Stim) by 5 frequency regions (Freq) by 3 time periods (Time) by the across-recording measure of 4 CF ranges (CF). Surface data were analyzed by a 3-way repeated-measures ANOVA: 3 Stim by 5 Freq by 3 Time. Interactions were untangled using 1-way repeated-measures ANOVAs. Strict Bonferroni adjustments were applied to alpha levels to correct for multiple comparisons.

## 3. Results

### 3.1. Response characteristics

Representative individual recordings from surface and ICc sites are plotted in Fig. 2. To characterize the range of frequencies present in the responses of the two recording sites, we computed fast Fourier transforms on/ga/responses for each site and created grand averages. Because the responses collected directly from the

ICc are many times larger than the volume conducted responses collected at the surface (note the difference in scale on Fig. 2, top), we normalized the frequency spectra by applying a constant multiplier to the surface response spectrum that equated the magnitude of F0 (100 Hz) with that of the ICc response spectrum. These results are plotted in Fig. 2 (bottom). Note the similarity in response spectra up to approximately 3000 Hz. The highest frequency band analyzed in this paper was 1500–3000 Hz.

Cross-phaseogram plots from surface and ICc averaged across all animals and recordings are presented in Fig. 3, with stimulus comparisons shown for reference. In these and all cross-phaseogram plots, regions where responses to stimuli that have higher F2 values phase-lead are represented in red, regions where they phase-lag are represented in blue, regions of minimal difference are represented in green, and white indicates regions below the noise floor. These plots show that, consistent with the human data reported in Skoe et al. (2011), the largest phase differences of surface-recorded data are present in the formant transition period. Additionally, one can see that as in the human data, the GaBa and DaBa comparisons elicit more phase distinctions than GaDa.

Representative individual cross-phaseograms of surface recordings (Fig. 4) and ICc recordings (Fig. 5) are shown, with each row representing a separate recording. Results from the individual and average cross-phaseograms indicate that the surface-recorded data are noisier than the near-field recordings at higher frequencies, approximately 2 kHz and above. This result was not surprising due to the known attenuation of high-frequency information in volume conducted responses.

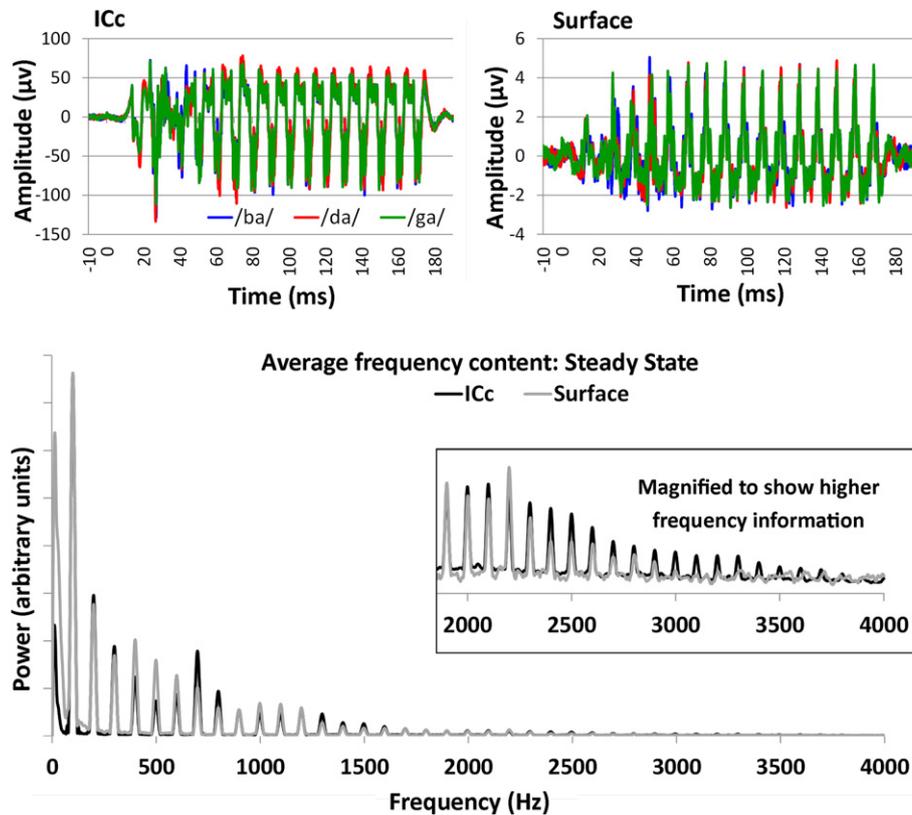
### 3.2. Surface results

Line graphs of mean phase differences extracted from each frequency range within each stimulus comparison recorded from the surface electrode are illustrated in Fig. 6, right side. For analysis, phase differences within frequency bands were averaged within formant transition and steady state time periods. All main effects and interactions of a 3-way Stim  $\times$  Freq  $\times$  Time ANOVA were significant (see Table 1). The main effect of stimulus indicated larger phase differences in the GaBa and DaBa stimulus comparisons. The main effect of frequency indicated largest phase differences in the 1–1.5 kHz range and smallest in the 70–400 Hz range. The main effect of time indicated larger phase differences during the formant transition.

To untangle the 3-way interaction, 1-way repeated-measures ANOVAs determined which frequency ranges distinguished stimuli within time periods (5 ANOVAs tested phase differences across stimulus comparisons within the formant transition, one within each frequency range, with a similar 5 tests for the steady state region; 10 tests, alpha adjusted to 0.005). Within the formant transition period, the 720–1000 Hz and 1–1.5 kHz frequency ranges showed significant phase differences across stimulus comparisons ( $F_{2,16} = 25.58$ ,  $p = 1.10 \times 10^{-5}$ ;  $F_{2,16} = 51.76$ ,  $p = 1.03 \times 10^{-7}$ , respectively), with the GaDa comparison eliciting a smaller phase difference in each case (Fig. 7, top left). None of the frequency ranges significantly differentiated stimulus comparisons in the steady state period, however the 720–1000 Hz range trended toward significance ( $F_{2,16} = 6.84$ ,  $p = 0.0071$ ; Fig. 7, top right).

### 3.3. ICc results

Of the 81 recordings obtained from the ICc, 19 were recorded from the Low CF range (100–500 Hz), 18 from the Mid CF range (500–1000 Hz), 21 from the High CF range (1000–2000 Hz) and 23 from the Higher CF range (2000–6300 Hz). A 4-way Stim  $\times$  Freq  $\times$  Time  $\times$  CF ANOVA revealed that all main effects and



**Fig. 2.** Top: Representative waveforms of responses to each stimulus recorded from a single animal. ICc recordings (CF = 1250 Hz) are illustrated on the left, Surface recordings on the right. Amplitude in microvolts is indicated on the y-axis, time in milliseconds on the x-axis. (Ga = Green; Da = Red; Ba = Blue). Bottom: Frequency response to one of the stimuli, Ga, from each recording site, averaged across all animals (Black line = Surface data; Gray line = ICc data). Frequency is indicated in Hz on the x-axis. Power, on the y-axis, has been normalized to equate F0 between sites, and is therefore reported in arbitrary units. Inset represents the same data magnified to enable observation of the higher frequency content. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

interactions involving Stimulus Comparison, Frequency Region and Time Period were highly significant. There was no main effect of CF ( $F_{3,77} = 0.78$ ,  $p = 0.51$ ), and no interactions involving CF were significant. Therefore, data from all CFs were combined in subsequent analyses.

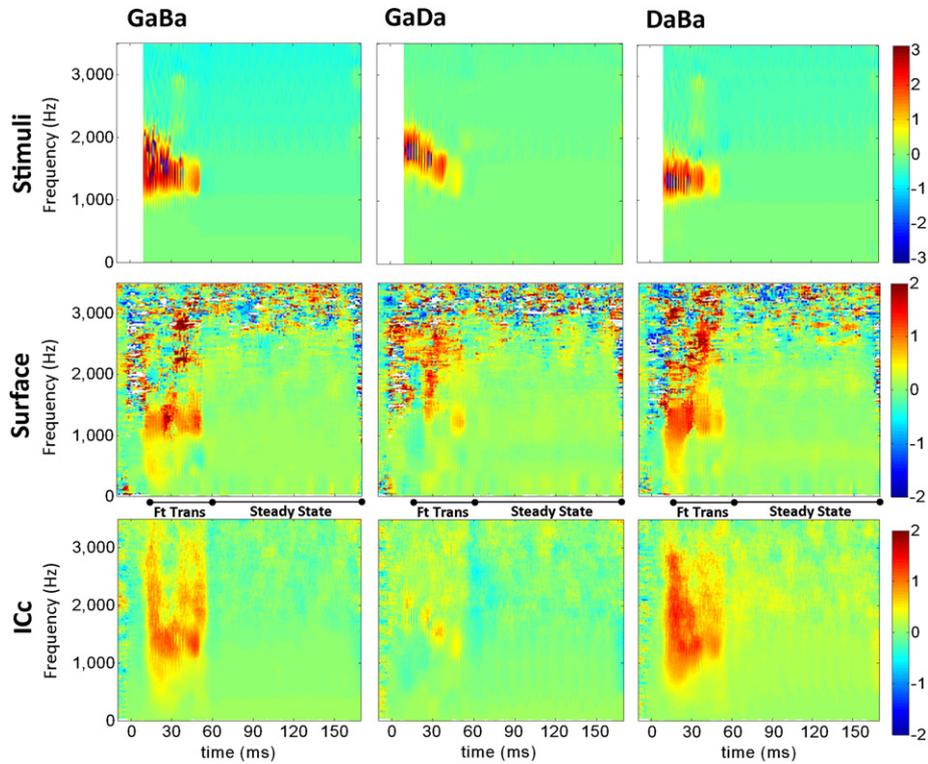
Line graphs of the mean ICc phase differences extracted from each frequency range within each stimulus comparison and averaged across all CF ranges are illustrated on the left side of Fig. 6. These plots show that, as expected, the largest phase differences exist in the formant transition period. Unexpectedly, during the steady state period when the stimuli are identical, phase differences oscillating at 100 Hz distinguish the stimulus comparisons. The GaDa and DaBa comparisons elicited the largest oscillation patterns, but were approximately 180° out of phase with each other, and the GaBa comparison evoked a much smaller oscillation.

A 3-way ANOVA without the CF term was performed (3 Stim × 5 Freq × 3 Time). All main effects and interactions were highly significant (see Table 1 for statistics). Post-hoc tests similar to the analysis of the far-field surface data were performed to untangle the 3-way interaction. 1-way repeated-measures ANOVAs determined which frequency ranges distinguished stimuli within time periods (5 ANOVAs tested phase differences across stimulus comparisons within the formant transition, one within each frequency range, with a similar 5 tests for the steady state region; 10 tests, alpha adjusted to 0.005). Within the formant transition period, all five frequency ranges showed highly significant phase differentials across stimulus comparisons, with the GaDa comparison eliciting a smaller phase difference than GaBa and DaBa in each case ( $F_{2,160}$  ranged from 61.34 (70–400 Hz) to 201.38 (1–1.5 kHz),

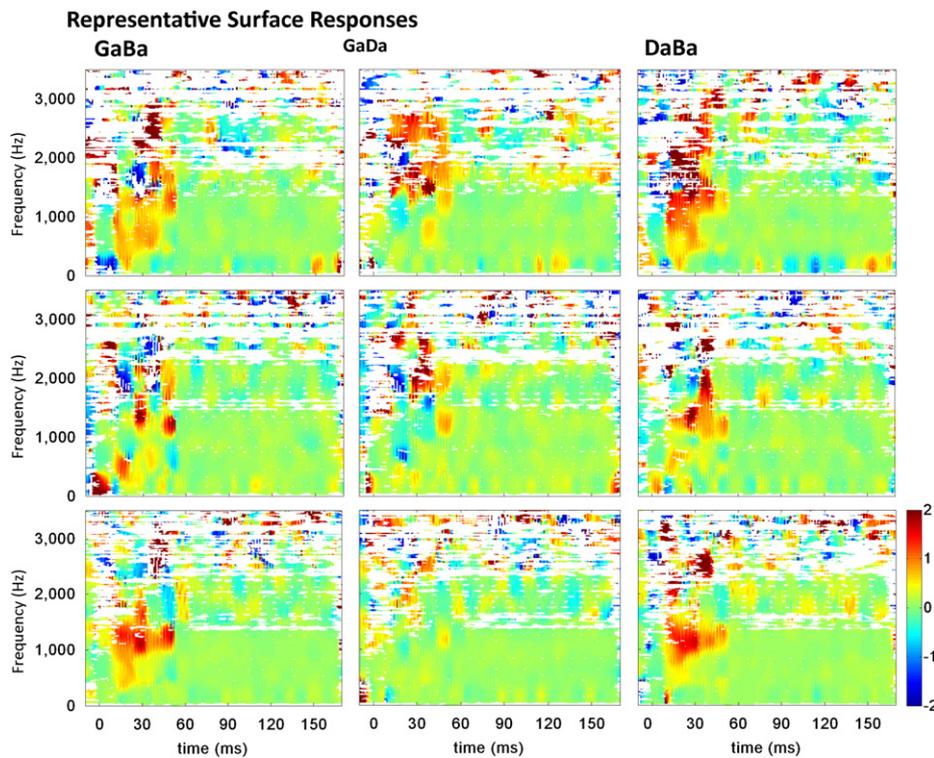
with all  $p$  values less than  $1 \times 10^{-19}$ ). Phase differences between stimuli increased from the lowest frequency range to 1–1.5 kHz, then decreased in the 1.5–3 kHz range (Fig. 7, bottom left). All frequency ranges also showed significant phase differences across stimulus comparisons in the steady state period, with the DaBa differential becoming increasingly more positive (/da/responses leading/ba/) from lower to higher frequencies as the GaBa and GaDa comparisons became increasingly negative (/ga/responses lagging/ba/and/da/) (Fig. 7, lower right). Comparisons in this time range yielded  $F_{2,160}$  values ranging from 6.62 (70–400 Hz) to 37.78 (1.5–3 kHz range), with corresponding  $p$ -values from 0.0017 to  $3.65 \times 10^{-14}$ .

#### 3.4. CF results

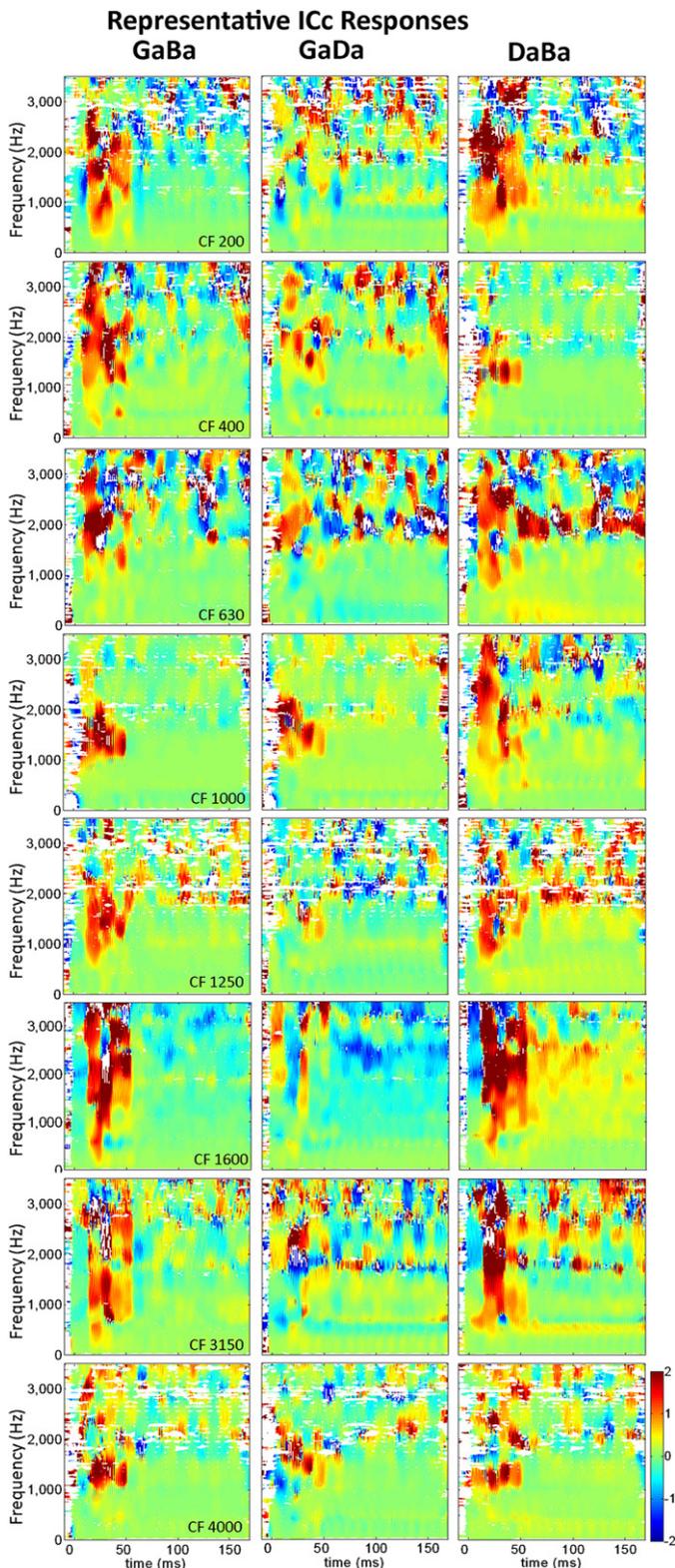
As described above, we did not find that CFs of the recording sites distinguish stimulus contrasts based on phase information (See Fig. 8 for average cross-phaseograms within each CF region). We therefore compared the amplitude of frequency responses at each CF region in response to the steady state time period of the stimuli. Fig. 9 shows waterfall plots showing the amplitude of each ICc recording across frequency. Because the amplitudes of higher frequencies are much smaller than the lower frequencies, they are plotted on a separate graph with a smaller z-axis range for better visualization. As predicted, an amplitude peak centers on F0 and its harmonics. It also appears that recordings from higher CF regions encode these frequencies more robustly than lower CF regions. To test this observation, we extracted amplitudes of 10-Hz-wide bins centered at each multiple of 100 Hz up to 2900 Hz. We then



**Fig. 3.** Cross-phaseograms comparing responses to contrasting speech sounds/ga/vs./ba/(left column), /ga/vs./da/(middle column), and /da/vs./ba/(right column). In each plot, time in ms is indicated on the x-axis, frequency in Hz is indicated on the y-axis. Colors indicate phase differences in radians, such that warm colors indicate that responses to stimuli with higher frequency content phase-lead, and cool colors indicate they phase-lag. The top row shows results from a cross-phase analysis of the stimuli as measured from the earphone output. Stimulus onset is shifted in time to line up with response onset to simplify visual comparison between stimulus and response results. The middle and bottom rows show grand-averaged data of the cross-phase analysis performed on the Surface and ICC responses, respectively. Analysis time periods for the formant transition (15–60 ms) and steady state period (60–170 ms) are indicated between the two response rows (Ft Trans = Formant Transition).



**Fig. 4.** Cross-phaseograms from surface-recorded data of three representative animals. Plot format is as described in Fig. 3 caption.



**Fig. 5.** Cross-phaseograms from near-field ICc data of eight representative recordings. See Fig. 3 caption for plot format details.

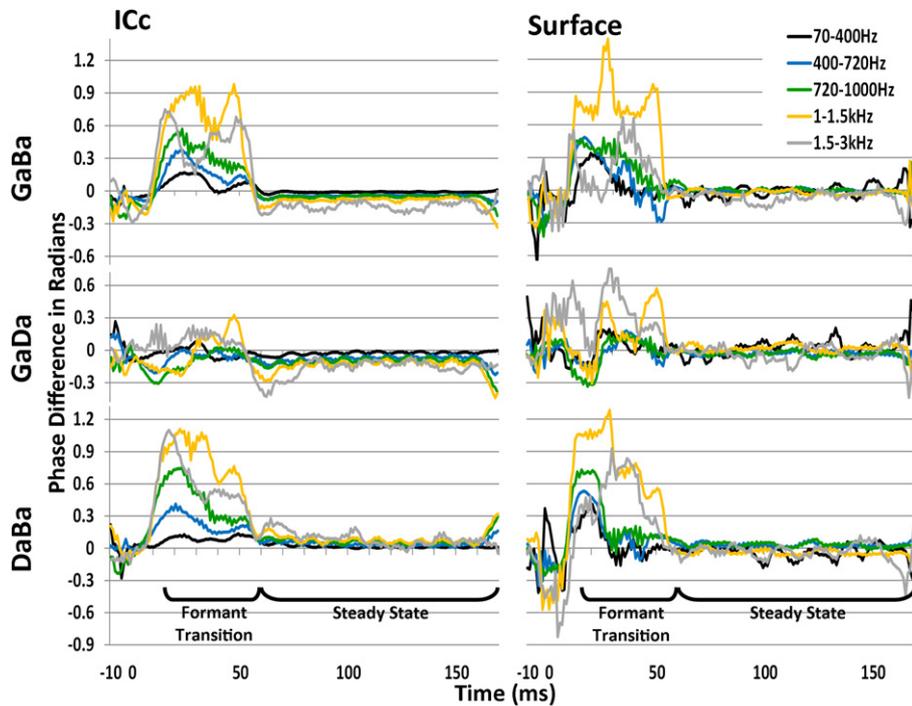
a main effect of CF ( $F_{3,78} = 8.06$ ;  $p = 9.65 \times 10^{-5}$ ) with larger amplitudes for higher-frequency CFs. We also found a significant interaction between CF and Frequency ( $F_{84,2184} = 7.85$ ;  $p = 1.87 \times 10^{-76}$ ). To understand this interaction, we correlated CF value with frequency amplitude at each frequency (29 correlations). Using a conservative alpha level,  $p < 0.01$ , significant correlations were found at F0 and several of the lower and mid-range harmonics (100 Hz–500 Hz), 800 Hz, 900 Hz, 1300 Hz, 1400 Hz and 1700 Hz. This analysis indicated that amplitudes of lower frequencies were influenced by CF while relatively higher frequencies were not correlated. Therefore, although the amplitude of frequency responses was impacted by the CF region of the recordings, the small timing differences differentiating the responses of one stimulus from another were not affected by CF.

#### 4. Discussion

To investigate frequency-specific timing information in the scalp-recorded response, as well as an auditory nucleus which contributes to the surface response, we examined phase coherence between responses to speech stimuli recorded from the inferior colliculus and surface vertex of guinea pigs. A major goal of this work was to examine whether stimulus-specific phase signatures measured at the scalp represent an epiphenomenon associated with far-field (i.e., scalp-recorded) measurement of neural activity, or alternatively whether these specific activity patterns represent a more fundamental temporal processing phenomenon as evidenced by similar activity patterns in near-field responses. Here we show that phase differences in surface-recorded responses measured in guinea pig reflect stimulus phase attributes: phase differences were prominent in the formant transition period of surface responses to both the/ga/vs/ba/and/da/vs/ba/stimulus comparisons, and these phase differences were less prominent for the/ga/vs./da/comparison. Importantly, these observations were also evident in surface recording measured in human subjects (Skoe et al., 2011), suggesting similar underlying neural mechanisms in human and animal auditory systems. Near-field responses measured from ICc showed similar results to those described for the surface recordings: the GaDa comparison elicited smaller phase differences relative to GaBa and DaBa. However, in contrast to the surface responses, significant, but relatively subtle, phase differences were also evident in the steady state portion of the near-field ICc response. Finally, we showed divergence between magnitude and phase spectra in ICc responses to these stimulus comparisons. Taken together, these results strongly suggest that the phase signatures elicited by speech sounds represent a fundamental temporal processing phenomenon which is generalized to both the surface-recorded brainstem response as well as the localized auditory nuclei which contribute to the surface response. In addition, we expect that the phase-sensitivity shown here in synthesized CV syllables would make this analysis method applicable to responses evoked by a variety of stimuli, including natural speech and non-speech sounds, or the same sound delivered in different manners such as within different maskers.

There are a number of important similarities and differences between responses measured from the surface and ICc in the current study. First, while there was a general correspondence between phase attributes in surface and ICc responses, phase differences recorded directly from the inferior colliculus encoded stimulus distinctions more broadly than those recorded at the cortical surface. These near-field responses revealed phase differences in both the formant transition and steady state portion of the responses across all assessed frequency ranges (70 Hz–3 kHz) and stimulus comparisons. The phase differences evident during the steady-state portion of the response, although significant, were

performed a repeated-measures ANOVA with a within-recording factor of Frequency (29 levels; 100 Hz, 200 Hz ... 2900 Hz) and a between-recording factor of CF region (4 levels). This analysis produced a significant main effect of Frequency ( $F_{28,2184} = 281.07$ ;  $p < 1 \times 10^{-76}$ ) with larger amplitudes at lower frequencies, and



**Fig. 6.** Average phase differences extracted from each frequency range of the response spectra. ICc data are presented on the left, surface data on the right. Stimulus comparison is indicated to the left of each row. Time in ms is indicated on the x-axis, relative to stimulus presentation at 0 ms. Phase difference in radians is indicated on the y-axis.

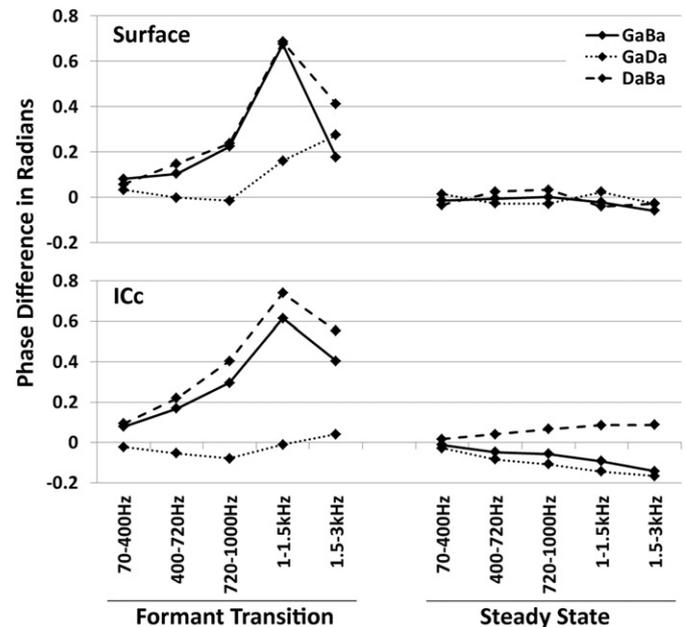
substantially smaller than those seen during the formant transition. Near-field recordings revealed a richer pool of significant points than the surface recordings, but the basic response pattern was similar across-recording sites: phase differences were most salient in the formant transition portion of the GaBa and DaBa comparisons.

Another notable difference between the surface and IC responses is that higher frequency phase differences, while present in the direct IC recordings, are not observed at the surface. A plausible explanation for the lack of sensitivity of surface responses at higher frequencies is that neural information accessible via near-field electrophysiological recordings is often reduced or absent from far-field recordings (See Wood and Allison, 1981 for a review). A contributing factor to this phenomenon is that responses from multiple sources are combined during volume-conduction to the surface of the brain. Another factor is that while both near- and far-field recordings include contributions from a population of neurons, near-field techniques record activity from a much smaller neural population. Other midbrain regions, and indeed, possibly other regions within IC, may not distinguish these stimuli by phase

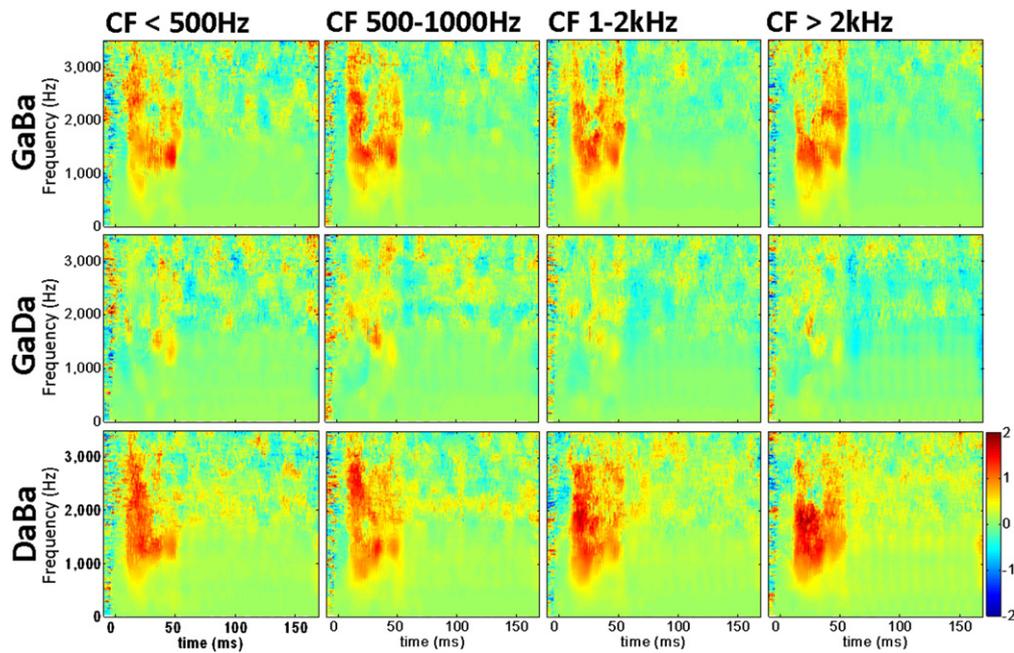
differences, and when the signal is recorded by an electrode at the cortical surface, differences detectable via near-field techniques may be obscured by other signals simultaneously transmitted to the surface.

**Table 1**  
ANOVA table.

	IC			Surface		
	F	df	p	F	df	p
<b>Main Effects</b>						
Stim	172.88	2, 160	$1.03 \times 10^{-40}$	12.94	2, 16	0.00045
Time	779.56	1, 80	$5.28 \times 10^{-43}$	121.35	1, 8	$4.10 \times 10^{-6}$
Freq	93.97	4, 320	$9.12 \times 10^{-53}$	17.03	4, 32	$1.40 \times 10^{-7}$
<b>2-Way Interactions</b>						
Stim × Time	187.97	2, 160	$1.00 \times 10^{-42}$	37.62	2, 16	$8.95 \times 10^{-7}$
Stim × Freq	35.60	8, 640	$1.15 \times 10^{-46}$	8.95	8, 64	$3.82 \times 10^{-8}$
Time × Freq	173.63	4, 320	$7.34 \times 10^{-79}$	21.62	4, 32	$1.02 \times 10^{-8}$
<b>3-Way Interaction</b>						
Stim × Time × Freq	29.00	8, 640	$1.10 \times 10^{-38}$	15.05	8, 64	$3.75 \times 10^{-12}$



**Fig. 7.** Phase differences averaged within frequency range and time period. Top: Data recorded from the surface electrode are presented in the top portion of the figure. Bottom: data recorded directly from the ICc are presented in the bottom portion. Averages within the formant transition and steady state periods are plotted on the left and right, respectively. The x-axis indicates the frequency range used to extract average phase differences. The y-axis indicates average phase difference in radians. Responses to different stimulus comparisons are plotted with separate lines.



**Fig. 8.** Cross-phaseograms of ICc data, average of all responses within each CF region. See Fig. 3 caption for plot format details. CF range is indicated on top, stimulus comparison is indicated to the left.

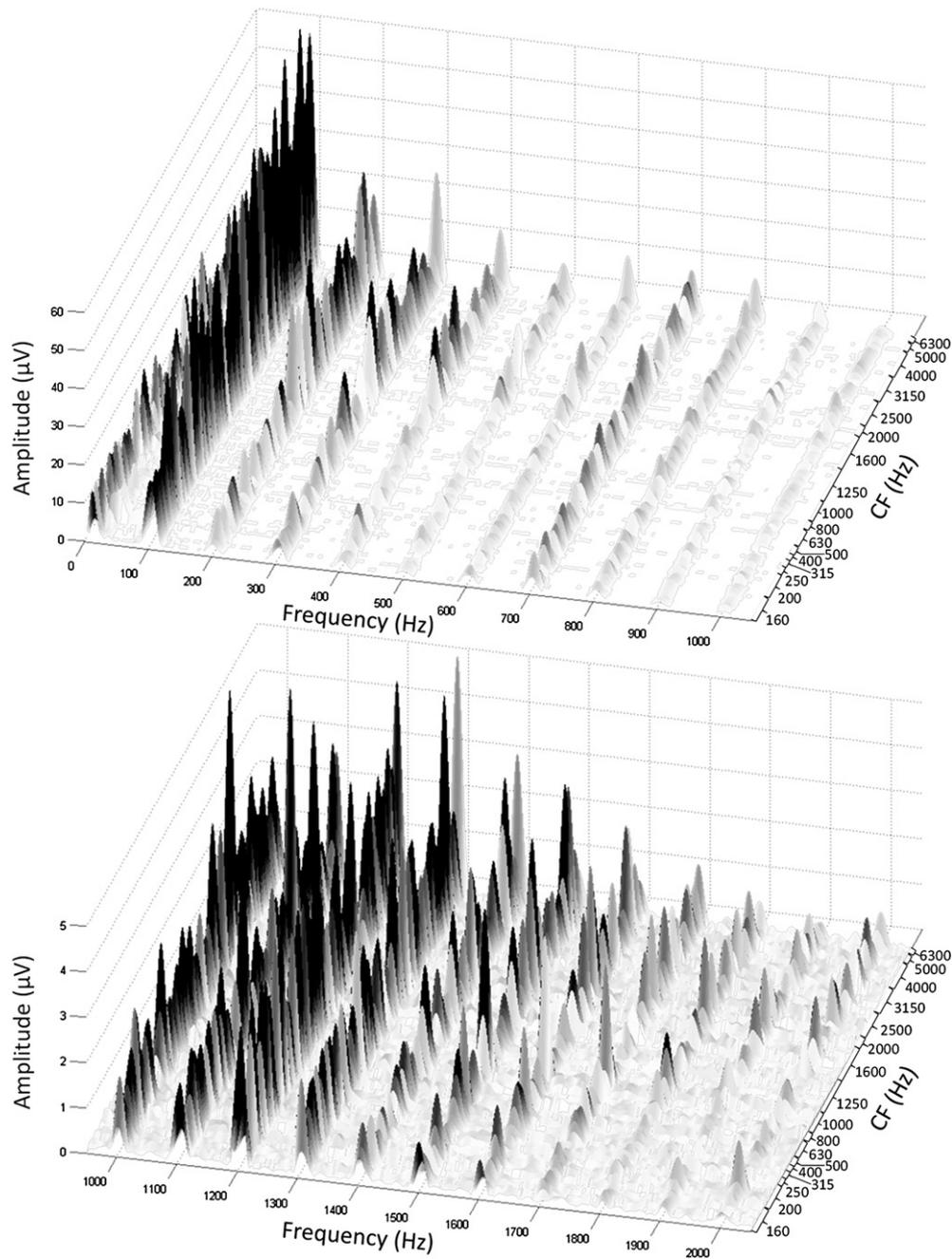
The convergence of phase-based results from surface and ICc responses strongly suggest that the phase differences recorded from the surface are generated at least in part by the ICc. This finding provides novel support for shared response features between the ICc and surface responses and adds to a literature that strongly suggests that the ICc plays an important role in shaping the surface-recorded auditory brainstem response. For example, it has been shown that the FFR, which is an important attribute of the surface-recorded auditory brainstem response to speech (Hornickel et al., 2009a; Johnson et al., 2008b), is greatly attenuated when the IC is cooled and is evident again after the IC is warmed (Smith et al., 1975). While this previous work focused on shared magnitude spectra between the surface and ICc, results from the current study add to our knowledge by showing shared phase spectra between the surface and ICc in response to speech sounds. Given the convergence of findings from both amplitude (Chandrasekaran and Kraus, 2010; Marsh et al., 1974; Smith et al., 1975) and phase spectra across the surface and ICc in animal models of the auditory system, we hypothesize that both the representation of amplitude and phase spectra in the human brainstem response reflects generalized auditory mechanisms that can be traced back to the properties of the nuclei which contribute to the surface response. It is hoped that future studies may be able to further test this hypothesis by examining perceptually-important attributes of the human brainstem response (Hornickel et al., 2011, 2009b; Johnson et al., 2008b) in animal models of the auditory system.

An important consideration of this work, as well as the previously published phase-related work in the human auditory system (Skoe et al., 2011), is how the auditory system might make use of this phase-related information in the processing of complex signals, including speech sounds. One hypothesis is that low-frequency phase-sensitivity evident at the scalp and ICc may represent an additional coding cue that may help facilitate the discrimination of speech stimuli. The logic for this hypothesis is grounded in the fact that the upper frequency range of phase-locking capability decreases in the ascending auditory pathway (See Joris et al., 2004 for a review). Therefore, the transposition of higher-frequency stimulus

differences to lower response frequencies may serve as a non-linear mechanism for encoding fast-moving frequency modulations that exceed the phase-locking capability of higher levels. This transposition may reflect the processing of amplitude modulations invoked by physical mechanisms of vocal production involving the fundamental frequency and its harmonics. John and Picton (2000) used simple amplitude modulated stimuli to illustrate how the phase of low-frequency envelope responses from the brainstem conveys information present in higher frequency regions of the stimuli. It is hoped that future studies may test this hypothesis by further examining the relationship between low-frequency phase and higher frequency components of acoustical stimuli as well as their possible link to perception.

We did not see differences in phase encoding across different CF groups in the near-field ICc data. This lack of differential encoding is not unprecedented; different CF regions of the IC have been found to respond similarly to vowels (Watanabe and Sakai, 1978). However, the magnitude of responses did vary by CF region, with higher frequency CF regions producing larger responses, particularly in the lower frequency range of the responses. Modeling work has revealed that mid- and high-frequency cochlear regions are primary contributors to low-frequency FFRs (Dau, 2003). This finding could help explain why we see greater FFR amplitudes in higher CF regions. The dissociation between magnitude and phase encoding indicates that although responses across the tonotopic map do differ, the differences in response magnitude do not impact the encoding of small timing differences that differentiate the stimuli. Response timing is encoded similarly across CF regions.

Phase differences in the steady state period of the near-field responses were not expected, as the stimuli are identical during this time period. No phase differences were evident in the steady state portion of the far-field responses. Small phase differences were seen in the human dataset collected by Skoe et al. (2011). In that study, steady state phase shifts indicated that responses to higher-frequency content phase-lagged responses to lower frequency content, the opposite direction than would be predicted if they were simply carry-over effects from earlier time periods. In



**Fig. 9.** Waterfall plots illustrating average frequency content of each ICc response to the steady state portion of the stimuli. Frequency in Hz is plotted on the x-axis. CF of individual responses are noted on the y-axis (to the right of the figure). Response magnitude at each frequency is indicated on the z-axis. Top figure shows responses from 0 to 1000 Hz. Bottom figure shows responses from 1000 to 2000 Hz. Note that the chart with higher frequencies is scaled to better illustrate response magnitudes in this frequency range.

the present study, phase oscillations at 100 Hz were evident in the near-field data, a frequency present throughout all stimuli as the fundamental frequency. As seen in Fig. 6, these oscillation patterns did not center at the zero-phase difference line; the DaBa and GaDa responses were superimposed on phase shifts of opposite direction (phase-lead and phase-lag, respectively) and were 180° out of phase. The GaBa response, which showed the least amount of steady state oscillation, ran closest to the zero-phase difference line. Therefore both oscillation pattern and static phase shift differed with stimulus comparison, suggesting that later phase encoding may provide contextual information concerning the spectro-temporal sound patterns earlier in the stimulus, in this case the formant transition. This concept is supported by data from

Watanabe and Sakai (1978) in which steady state IC responses to the vowel/a/were found to differ when the vowel was presented in isolation or preceded by connecting speech. More recently, context effects were reported in human brainstem responses to speech syllables (Chandrasekaran et al., 2009). The effects seen here may be related to the contextual effects reported in these papers.

## 5. Conclusion

This study shows that phase-related sensitivity to speech sounds measured in surface-recorded responses in guinea pig is also reflected in near-field responses in ICc and is similar to responses measured in humans. In conjunction with previous

studies, results suggest that the representation of amplitude and phase spectra in the human brainstem response reflects generalized auditory mechanisms that can be traced back to the properties of the nuclei which contribute to the surface response. We further propose that near-field responses recorded directly from the ICC may encode phase differences present in the stimulus and transpose these differences to lower frequencies. This transposition suggests a means of maintaining crucial information at anatomical levels not equipped for high-frequency phase locking.

### Funding

This work was supported by the National Institutes of Health (NIH: R01 DC01510) and the National Organization for Hearing Research (NOHR: 340-B208).

### Author contributions

All authors contributed to study design, data interpretation and manuscript writing. In addition, CMW performed data analysis, TGN collected data and aided data analysis, DAA collected data, and NK provided resources and facilities.

### Acknowledgments

We would like to thank Erika Skoe for her careful review of this manuscript and valuable insight into the nature of the cross-phase analysis method.

### References

- Abrams, D.A., Nicol, T., Zecker, S., Kraus, N., 2011. A possible role for a paralemniscal auditory pathway in the coding of slow temporal information. *Hear Res.* 272, 125–134.
- Akhoun, I., Gallégo, S., Moulin, A., Ménard, M., Veuillet, E., Berger-Vachon, C., Collet, L., Thai-Van, H., 2008. The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clin. Neurophysiol.* 119, 922–933.
- Banai, K., Hornickel, J., Skoe, E., Nicol, T., Zecker, S., Kraus, N., 2009. Reading and subcortical auditory function. *Cereb. Cortex.*
- Chandrasekaran, B., Kraus, N., 2010. The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiol.* 47, 236–246.
- Chandrasekaran, B., Hornickel, J., Skoe, E., Nicol, T., Kraus, N., 2009. Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: implications for developmental dyslexia. *Neuron* 64, 311–319.
- Cunningham, J., Nicol, T., King, C., Zecker, S.G., Kraus, N., 2002. Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex. *Hear Res.* 169, 97–111.
- Dau, T., 2003. The importance of cochlear processing for the formation of auditory brainstem and frequency following responses. *J. Acoust. Soc. Amer.* 113, 936–950.
- Hornickel, J., Skoe, E., Kraus, N., 2009a. Subcortical laterality of speech encoding. *Audiol. Neurootol.* 14, 198–207.
- Hornickel, J., Chandrasekaran, B., Zecker, S., Kraus, N., 2011. Auditory brainstem measures predict reading and speech-in-noise perception in school-aged children. *Behav. Brain Res.* 216, 597–605.
- Hornickel, J., Skoe, E., Nicol, T., Zecker, S., Kraus, N., 2009b. Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proc. Natl. Acad. Sci. U S A* 106, 13022–13027.
- John, M.S., Picton, T.W., 2000. Human auditory steady-state responses to amplitude-modulated tones: phase and latency measurements. *Hear Res.* 141, 57–79.
- Johnson, K.L., Nicol, T., Zecker, S.G., Kraus, N., 2008a. Developmental plasticity in the human auditory brainstem. *J. Neurosci.* 28, 4000–4007.
- Johnson, K.L., Nicol, T., Zecker, S.G., Bradlow, A.R., Skoe, E., Kraus, N., 2008b. Brainstem encoding of voiced consonant–vowel stop syllables. *Clin. Neurophysiol.* 119, 2623–2635.
- Joris, P.X., Schreiner, C.E., Rees, A., 2004. Neural processing of amplitude-modulated sounds. *Physiol. Rev.* 84, 541–577.
- King, C., Nicol, T., McGee, T., Kraus, N., 1999. Thalamic asymmetry is related to acoustic signal complexity. *Neurosci. Lett.* 267, 89–92.
- King, C., Warrier, C.M., Hayes, E., Kraus, N., 2002. Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neurosci. Lett.* 319, 111–115.
- Klatt, D.H., 1980. Software for a cascade/parallel formant synthesizer. *J. Acoust. Soc. Amer.* 67, 971–995.
- Kraus, N., Banai, K., 2007. Auditory-processing malleability - focus on language and music. *Curr. Dir. Psychol.* 16, 105–110.
- Kraus, N., Smith, D.I., Grossmann, J., 1985. Cortical mapping of the auditory middle latency response in the unanesthetized guinea pig. *Electroencephalogr. Clin. Neurophysiol.* 62, 219–226.
- Kraus, N., Smith, D.I., McGee, T., 1988. Midline and temporal lobe MLRs in the guinea pig originate from different generator systems: a conceptual framework for new and existing data. *Electroencephalogr. Clin. Neurophysiol.* 70, 541–558.
- Kraus, N., McGee, T., Littman, T., Nicol, T., 1992. Reticular formation influences on primary and non-primary auditory pathways as reflected by the middle latency response. *Brain Res.* 587, 186–194.
- Kraus, N., McGee, T., Littman, T., Nicol, T., King, C., 1994a. Nonprimary auditory thalamic representation of acoustic change. *J. Neurophysiol.* 72, 1270–1277.
- Kraus, N., McGee, T., Carrell, T., King, C., Littman, T., Nicol, T., 1994b. Discrimination of speech-like contrasts in the auditory thalamus and cortex. *J. Acoust. Soc. Amer.* 96, 2758–2768.
- Krishnan, A., 2002. Human frequency-following responses: representation of steady-state synthetic vowels. *Hear Res.* 166, 192–201.
- Krishnan, A., Xu, Y., Gandour, J., Cariani, P., 2005. Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res. Cogn. Brain Res.* 25 (1), 161–168.
- Liberman, A.M., Delattre, P.C., Cooper, F.S., Gerstman, L.J., 1954. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychol. Monogr.* 68 (8) No. 379.
- Liu, L.F., Palmer, A.R., Wallace, M.N., 2006. Phase-locked responses to pure tones in the inferior colliculus. *J. Neurophysiol.* 95, 1926–1935.
- Marsh, J.T., Brown, W.S., Smith, J.C., 1974. Differential brainstem pathways for the conduction of auditory frequency-following responses. *Electroencephalogr. Clin. Neurophysiol.* 36, 415–424.
- McGee, T., Kraus, N., King, C., Nicol, T., Carrell, T.D., 1996. Acoustic elements of speechlike stimuli are reflected in surface recorded responses over the guinea pig temporal lobe. *J. Acoust. Soc. Am.* 99, 3606–3614.
- Musacchia, G., Sams, M., Skoe, E., Kraus, N., 2007. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci.* 104, 15894–15898.
- Rees, A., Palmer, A.R., 1988. Rate-intensity functions and their modification by broadband noise for neurons in the guinea pig inferior colliculus. *J. Acoust. Soc. Am.* 83, 1488–1498.
- Skoe, E., Nicol, T., Kraus, N., 2011. Cross-phaseogram: Objective neural index of speech sound differentiation. *J. Neurosci. Methods* 196, 308–317.
- Smith, J.C., Marsh, J.T., Brown, W.S., 1975. Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. *Electroencephalogr. Clin. Neurophysiol.* 39, 465–472.
- Song, J.H., Banai, K., Kraus, N., 2008. Brainstem timing deficits in children with learning impairment may result from corticofugal origins. *Audiol. Neurootol.* 13, 335–344.
- Syja, J., Popelár, J., Kvasnák, E., Astl, J., 2000. Response properties of neurons in the central nucleus and external and dorsal cortices of the inferior colliculus in guinea pig. *Exp. Brain Res.* 133, 254–266.
- Vander Werff, K.R., Burns, K.S., 2011. Brain Stem responses to speech in Younger and Older adults. *Ear Hear* 32, 168–180.
- Watanabe, T., Sakai, H., 1978. Responses of the cat's collicular auditory neuron to human speech. *J. Acoust. Soc. Am.* 64, 333–337.
- Wood, C.C., Allison, T., 1981. Interpretation of evoked potentials: a neurophysiological perspective. *Can. J. Psychol.* 35, 113–135.