Clinical Neurophysiology 119 (2008) 2623–2635

Contents lists available at ScienceDirect

Clinical Neurophysiology

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

Brainstem encoding of voiced consonant–vowel stop syllables

Krista L. Johnson [a,b,*], Trent Nicol [a,b], Steven G. Zecker [a,b], Ann R. Bradlow [a,c], Erika Skoe [a,b], Nina Kraus [a,b,d,e,f]

[a] Auditory Neuroscience Laboratory, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA
[b] Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA
[c] Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, IL 60208, USA
[d] Institute for Neuroscience, Northwestern University, USA
[e] Neurobiology and Physiology, Northwestern University, USA
[f] Otolaryngology, Northwestern University, USA

ARTICLE INFO

Article history:
Accepted 19 July 2008
Available online 24 September 2008

Keywords:
Auditory brainstem response
Speech encoding
Stop consonant–vowel syllables

ABSTRACT

Objective: The purpose of this study is to expand our understanding of how the human auditory brainstem encodes temporal and spectral acoustic cues in voiced stop consonant–vowel syllables.
Methods: Auditory evoked potentials measuring activity from the brainstem of 22 normal learning children were recorded to the voiced stop consonant syllables [ga], [da], and [ba]. Spectrotemporal information distinguishing these voiced consonant–vowel syllables is contained within the first few milliseconds of the burst and the formant transition to the vowel. Responses were compared across stimuli with respect to their temporal and spectral content.
Results: Brainstem response latencies change in a predictable manner in response to systematic alterations in a speech syllable indicating that the distinguishing acoustic cues are represented by neural response timing (synchrony). Spectral analyses of the responses show frequency distribution differences across stimuli (some of which appear to represent acoustic characteristics created by difference tones of the stimulus formants) indicating that neural phase-locking is also important for encoding these acoustic elements.
Conclusions: Considered within the context of existing knowledge of brainstem encoding of speech–sound structure, these data are the beginning of a comprehensive delineation of how the human auditory brainstem encodes perceptually critical features of speech.
Significance: The results of this study could be used to determine how neural encoding is disrupted in the clinical populations for whom stop consonants pose particular perceptual challenges (e.g., hearing impaired individuals and poor readers).

1. Introduction

In speech, perceptual identification of vowels is determined by the frequencies of the first few formants which reflect the resonant properties of the vocal tract (Hillenbrand and Gayvert, 1993). Stop consonants are produced by a temporary obstruction of airflow through the vocal tract with three distinct phases: closing of the oral cavity, having the oral cavity remain closed while pressure builds, and finally a release of the closure allowing airflow to resume through the oral cavity (Halle et al., 1957; Johnson, 2003). A stop consonant is rich in acoustic cues that underlie the identification of its place of articulation and voicing, such as formant transitions (Blumstein et al., 1982; Delattre et al., 1955; Liberman et al., 1967), spectrum of the stop release burst (Blumstein and Stevens, 1979; Kewley-Port, 1983; Kewley-Port et al., 1983; Stevens and Blumstein, 1978), the presence or absence of post-release aspiration, and the time of the onset of voicing for the following vowel (Blumstein et al., 1982; Sinnott and Adams, 1987; Stevens and Klatt, 1974; Summerfield and Haggard, 1977).

Animal models have shown that this perceptual acoustic information is encoded across many levels of the auditory system as distinct neural events. Both peripheral and central structures such as the auditory nerve and cochlear nucleus fibers show phase-locking activity to the harmonics in a speech stimulus (Clarey et al., 2004; Sachs and Young, 1979; Sinex and McDonald, 1989; Young and Sachs, 1979). Additionally, these structures, as well as the rostral inferior colliculus, also show a marked increase in discharge rate at the onset of voicing (Chen et al., 1996). In contrast to single neurons, human brainstem evoked potentials reflect the aggregate neural response of several different types of cells, primarily those

* Corresponding author. Address: Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA. Tel.: +1 847 491 2465; fax: +1 847 491 2523.
E-mail address: kljohnson@u.northwestern.edu (K.L. Johnson).
URL: http://www.brainvolts.northwestern.edu (K.L. Johnson).

1388-2457/$34.00
doi:10.1016/j.clinph.2008.07.277

neurons in the rostral brainstem (i.e., inferior colliculus). Even at the level of the auditory cortex, the spectral content of the release of the stop burst is reflected by a latency shift in aggregate neural activity; as the spectral center-of-gravity of the stimulus' onset increases, the onset response latency decreases (McGee et al., 1996; Steinschneider et al., 1995).

The brainstem responds with a high degree of neural synchrony and is exceedingly well attuned to the temporal and spectral characteristics of sound, including speech sounds. However, the mechanisms involved in accurately encoding the many acoustic cues in speech still remain largely speculative. A large body of research has studied how the human auditory brainstem responds to the speech sound [da] (for review, see Banai et al., 2007). From this research, a framework has been proposed that suggests that separate neural mechanisms are responsible for encoding different acoustic aspects of speech sounds (Johnson et al., 2005; Kraus and Nicol, 2005). Speech sounds consist of three fundamental components: pitch, which is a source characteristic conveyed by the fundamental frequency; formants, which are filter characteristics conveyed by the selective enhancement and attenuation of harmonics; and the timing of major acoustic landmarks. All these aspects are important for speech perception. From a production standpoint, pitch arises from the "source," the rate of vibration of the vocal folds, while formants arise from the vocal tract "filtering," and timing arises from the interplay between actions of the source and filter, and from the opening and closing gestures involved in the production of sequences of segments. Although source and filter characteristics are simultaneously present in a speech signal and in a response, specific components of the brainstem response separately reflect the acoustic characteristics of pitch, formants, and segment-level timing.

The proposed framework, indeed, coincides with how the auditory system encodes spectral and periodic information. In the mature auditory system, basal regions of the cochlea are maximally responsive to high frequencies, and apical regions are maximally responsive to lower frequencies. This tonotopic organization is preserved throughout the central auditory pathways and is thought to help preserve the spectral relations in the pattern of neural activity (Langner, 1997; Langner et al., 1997; Merzenich and Reid, 1974; Rose et al., 1959). It has been demonstrated that the cortical coding of stop consonants depends on the frequency content of the component formants (Martin et al., 1997; McGee et al., 1996; Steinschneider et al., 1993). Therefore, it is thought that response latency—the time interval between the onset of the evoking sound and the response—and spectral coding are the primary analytic modes for extracting timing and formant information from filter cues. Signal periodicity, on the other hand, affords the quality of "pitch" and is thought that neural phase-locking (via the FFR) is the analytic mode of extracting this source cue. Systematic encoding of periodicity has been demonstrated, and there is some evidence for the existence of orthogonal periodic and tonotopic maps in the cochlear nucleus, inferior colliculus, and auditory cortex (Langner, 1997; Langner et al., 1997; Langner and Schreiner, 1988; Merzenich et al., 1975; Schreiner and Langner, 1988; Suga and O'Neill, 1979).

Based on the aforementioned studies, it is of interest to learn how the human auditory brainstem is actually encoding the acoustic cues in speech when the only differences among syllables are filter characteristics, as is the case in the syllables [ga], [da], and [ba] used in this study. Acoustically in natural speech, a primary (but not only) difference between these syllables is the trajectory of the second and the third formant frequencies (F_2 and F_3) during the formant transition period. However, F_2 and F_3 are beyond the phase-locking limit of the rostral brainstem (Blackburn and Sachs, 1989; Frisina, 2001; Frisina et al., 1990; Joris et al., 2004; Langner and Schreiner, 1988; Wang and Sachs, 1994), and phase-locking to these frequencies is not evident in the response. Based on the tonotopicity of the brainstem nuclei, response timing would be the most likely candidate for differentiating between these spectral cues, such that a stimulus containing higher frequencies would have earlier response latencies than a stimulus containing lower frequencies. This latency progression as a function of frequency has been demonstrated in the brainstem response to pure-tones (Gorga et al., 1988).

The purpose of this study was to determine the neural encoding of the voiced stop consonants [ga], [da], and [ba] to establish how the brainstem encodes the acoustic cues that differentiate these consonants. Syllables with an initial stop consonant were used because stop consonants are notorious for being most vulnerable to misperception in clinical populations, specifically hearing loss (Townsend and Schwartz, 1981; Van Tasell et al., 1982), poor readers (de Gelder and Vroomen, 1998; Tallal, 1980; Tallal and Stark, 1981), and people with auditory processing disorders (Tobey et al., 1979). Using synthetic speech to isolate the primary acoustic cues to the place of articulation in these consonants, we hypothesized the following: (1) because of the tonotopicity of the auditory pathway, the differing F_2 and F_3 frequencies of the formant transition period manifest themselves in terms of neural timing (synchrony) as latency shifts, with [ga], [da] and [ba] responses arising progressively later; (2) latency differences diminish over the course of the response until they vanish completely by the time the three syllables reach their shared steady-state vowel; (3) based on the phase-locking limitations of the brainstem, neural encoding of the periodic acoustic properties of speech consists largely of phase-locking to frequencies below the second formant, such that there is little difference in response spectra at frequencies where the stimuli differ; and (4) there are no between-stimulus latency or spectral differences in response to the portion of the syllable where formants have reached steady-state (steady-state portion) because the acoustic properties are identical during this period.

2. Materials and methods

2.1. Participants

Twenty-two children between the ages of 8–12 years participated in this study. All children had normal hearing as assessed by (a) pure tone air- and bone-conduction thresholds from 250 to 8000 Hz (less than 20 dB HL, no air-bone gap greater than 10 dB HL), and (b) normal wave V click auditory brainstem response (ABR) latencies. Additionally, children with language-based learning problems are known to frequently have abnormal brainstem responses to speech (Banai et al., 2005; Cunningham et al., 2001; Johnson et al., 2007; King et al., 2002; Russo et al., 2005; Wible et al., 2004a,b). Therefore, all subjects were screened and none of the children had a history or diagnosis of a language-based learning problem or attention deficit/hyperactivity disorder (ADHD). A study-specific psychoeducational test battery consisting of three literacy subtests (reading, spelling, and word attack) from the Woodcock-Johnson-III (Woodcock et al., 2001) confirmed that all children performed within or above the normal range on these measures (all scores >85). The Wechsler Abbreviated Scale of Intelligence was administered to ensure that the mental ability of all subjects was above 85. Written and oral informed assent was given by each child, and his or her parent or guardian provided written consent. The children were paid for their participation. Institutional review board approval for this study was obtained from Northwestern University.

2.2. Stimuli and presentation

A Klatt cascade/parallel formant synthesizer (Klatt, 1980) was used to synthesize the speech stimuli [ga], [da], and [ba] at a sampling rate of 20 kHz. The three stimuli share the following charac-

teristics. Their durations are 170 ms with voicing (100 Hz F_0) onset at 10 ms. The formant transition durations are 50 ms and comprise a linearly rising F_1 (400–720 Hz) and flat F_4 (3300 Hz), F_5 (3750 Hz), and F_6 (4900 Hz). Ten ms of initial frication are centered at frequencies around F_4 and F_5. After the 50 ms formant transition period, F_2 and F_3 remain constant at their transition end point frequencies of 1240 and 2500 Hz, respectively, for the remainder of the syllable. The stimuli differ only in the starting points of F_2 and F_3. For [ba], F_2 and F_3 rise from 900 Hz and 2,400 Hz, respectively. For [da], F_2 and F_3 fall from 1700 and 2580, respectively. And for [ga], F_2 and F_3 fall from 3000 and 3100, respectively. These synthesized stimuli have an identical and constant F_0 throughout their entire duration. Table 1 lists all stimulus parameters, and differences among the three syllables are highlighted. Fig. 1 provides a schematic representation (panel A) and the time–amplitude waveforms (panels B and C) of all three stimuli. Panel D illustrates the spectral composition of the first 60 ms of the stimuli from 0 to 4000 Hz.

Stimuli were delivered with a PC-based stimulus delivery system (Neuroscan Gentask, Compumedics USA, Charlotte, NC) that outputs the signals through a 16-bit converter. The rate of presentation was 4.35/s. Both stimulus polarities (condensation and rarefaction) were presented. The test stimuli were presented to the right ear through Etymotic ER-3 earphones (Etymotic Research, Elk Grove Village, IL) at an intensity of 83 dB SPL. The left ear was unoccluded. To ensure subject cooperation and promote stillness, all subjects watched videotaped programs such as movies or cartoons with the sound presented at a low level (<40 dB SPL). They were instructed to attend to the video rather than to the stimuli.

2.3. Recording parameters

Continuous EEG was acquired with NeuroScan 4.3 (Compumedics USA, Charlotte, NC) from Cz-to-ipsilateral earlobe, with forehead as ground, bandpass filtered from 0.05 to 3000 Hz, and digitized at 20,000 Hz. All recordings were made with silver–silver chloride electrodes (impedance < 5 kΩ). EEG was processed offline to create averages for each stimulus condition. Each continuous file was bandpass filtered from 70 to 2000 Hz to isolate the brainstem response frequencies. The EEG was then divided into 230 ms epochs (45 ms pre-stimulus onset to 185 ms post-stimulus onset). This time window allows examination of the onset response, occurring within the first 10 ms, the frequency following response (FFR) that continues for the duration of the stimulus, and the offset response occurring within 10 ms of stimulus cessation. An artifact criterion of ±35 μV was applied to reject epochs that contained myogenic artifacts. For each stimulus, the processed epochs were separately averaged (according to polarity) and then added together in order to isolate the neural response from that of the cochlear microphonic (Gorga et al., 1985). The final average waveform for each stimulus contained between 4000 and 4100 sweeps per subject.

2.4. Analysis

2.4.1. Formant transition period

2.4.1.1. Transient peak measures. The formant transition period is defined as the portion of the response corresponding to the onset and formant transition periods of the stimuli (0–50 ms). Based on our first hypothesis, latency differences in this portion of the response were expected to be present across stimuli. In order to best isolate this portion of the response and eliminate low-frequency activity that could obscure subtle latency differences, response waveforms were additionally high-pass filtered at 300 Hz. A visual analysis of the first 70 ms of the grand average waveforms showed an onset response (~9 ms) and major peaks of activity occurring approximately every 10 ms starting around 23 ms. Absolute latency was recorded for a total of 16 peaks for each subject, in each stimulus condition (see Fig. 2A). Peaks 1 and 2 are the onset response (shown as shaded circles in Fig. 2A). Peaks 3, 4, 6, 7, 9, 10, 12, and 13 are considered "major" peaks (filled circles), as they are the most robust positive and negative peaks of activity within this time period. Peaks 5, 8, 11, and 14 are considered "minor" peaks (open circles), and were defined as the next negative peak following a major negativity. Peaks 15 and 16 occur in response to the very end of the stimulus formant transition period, where the acoustic properties of the three stimuli are approaching the point of being identical, and are thus considered "end-point" peaks (shaded circles).

In order to normalize the latencies so that all 16 peaks could be depicted on the same scale, a grand mean (GM) latency for each peak across all three stimuli was computed. Then, each respective grand mean was subtracted from each individual peak latency ($Lat_{individual} - Lat_{GM}$). Thus earlier peaks (i.e. [ga]) are negative numbers, later peaks (i.e. [ba]) are positive numbers, and peaks near the grand mean (i.e. [da]) are near zero.

Statistical analyses were performed using the normalized latencies on 4 groups of peaks: onset peaks 1 and 2; major peaks 3, 4, 6, 7, 9, 10, 12, and 13; minor peaks 5, 8, 11, and 14; and end-point peaks 15 and 16. A 3 x K repeated measures ANOVA (where 3 is the number of stimulus conditions and K is the number of peaks) was conducted on each group. For those groups where the stimulus/peak interaction was significant, repeated measure ANOVA and follow-up paired *t*-tests were performed for each peak within a given group to assess between-stimulus latency differences.

2.4.1.2. Frequency-domain measures. To evaluate the spectral composition of the responses, a fast Fourier transform (FFT) analysis was performed on each response over the time period corresponding to the formant transition period. For a response time range of 18-58 ms, average response magnitudes were calculated for 50 Hz-wide bins surrounding the frequency of the stimulus F_0 and the subsequent 10 harmonics. A repeated measure ANOVA and, when appropriate, follow-up paired *t*-tests were performed for each frequency bin to test for the significant between-stimuli magnitude differences.

Table 1
Values (in Hz) of the fundamental frequency and the 6 formant frequencies of each stimulus

	F_0	F_1		F_2		F_3		F_4	F_5	F_6
	Flat	Onset	Steady-state	Onset	Steady-state	Onset	Steady-state	Flat	Flat	Flat
Stimulus formant frequencies (Hz)										
/ga/	100	400	720	**3000**	1240	**3100**	2500	3300	3750	4900
/da/	100	400	720	**1700**	1240	**2580**	2500	3300	3750	4900
/ba/	100	400	720	**900**	1240	**2400**	2500	3300	3750	4900

F_1, F_2, and F_3 have ramping frequencies during the formant transition period (first 50 ms), while F_0, F_4, F_5, and F_6 remain flat throughout the entirety of the stimulus. The differences in the starting frequencies between stimuli of F_2 and F_3 are indicated in bold.

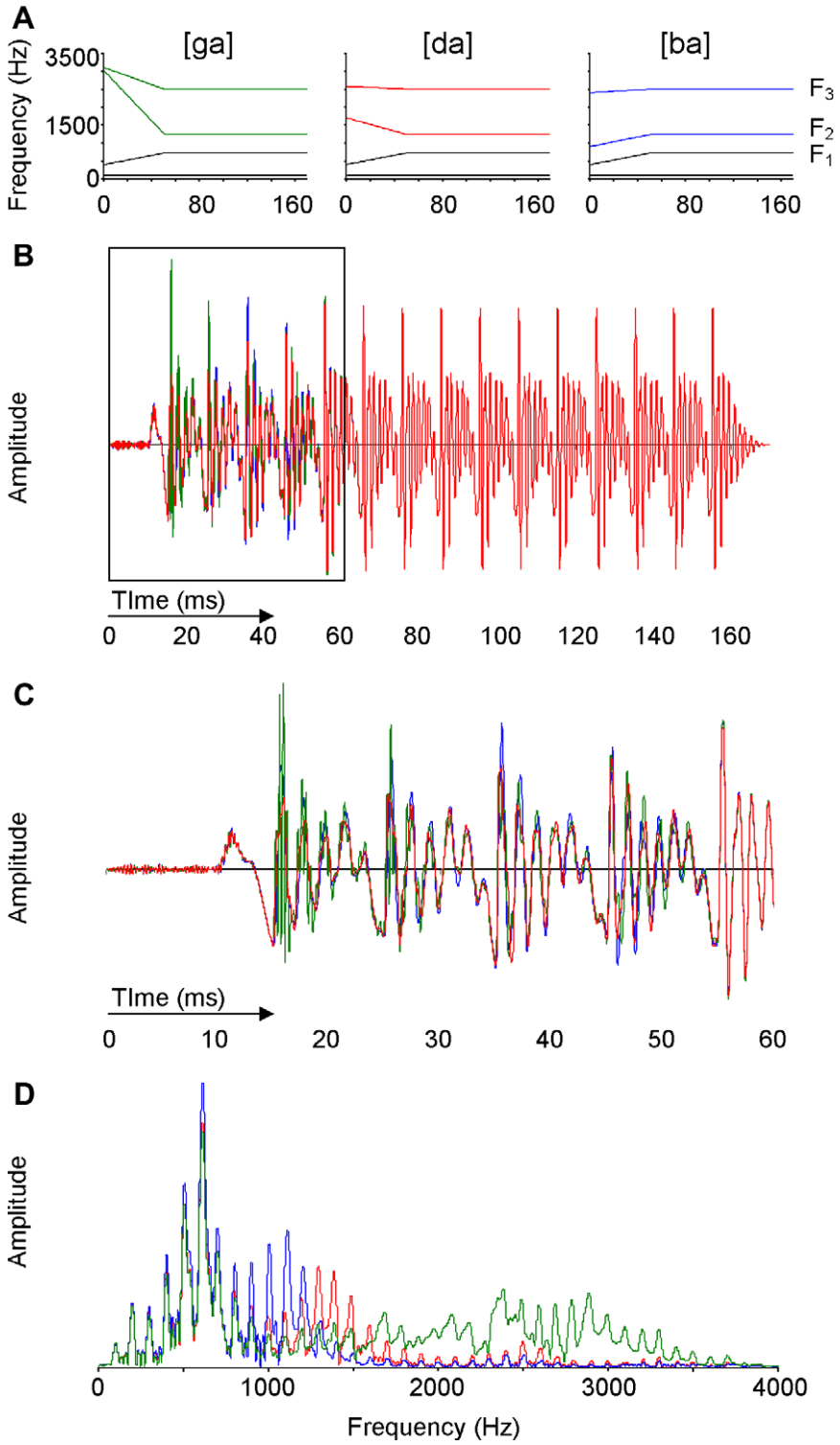

Fig. 1. (A) Schematic representations of the spectral composition of [ga] (left), [da] (center), and [ba] (right) stimuli as a function of time (in ms). The first 50 ms are the formant transition period, followed by the steady-state [a] portion. (B) Time–amplitude stimulus waveforms for [ga] (green), [da] (red), and [ba] (blue) overlaid for the entire duration of the stimulus. Where they are identical, the waveform is red. (C) Enlarged representation of the first 60 ms of the stimuli. (D) Spectral composition of the first 60 ms of the stimuli, where frequency (Hz) is along the *x*-axis and amplitude is along the *y*-axis.

2.4.2. Steady-state response

2.4.2.1. Transient peak measures. The sustained response is defined as the portion of the response corresponding to the steady-state portion, where the acoustic parameters of the stimuli are identical (51-170 ms). Peak latency values for major and minor peaks were picked for a 70 ms portion of the steady-state response (90-160 ms). Again, major peaks were defined as those peaks with the greatest amplitude departure from zero, and minor peaks were defined as the next negative peak following a major negativity. In this 70 ms time range, there were 14 major peaks and 7 minor peaks. Normalized latencies were computed for all 21 peaks, and a 3 x K repeated measures ANOVA was conducted for the major

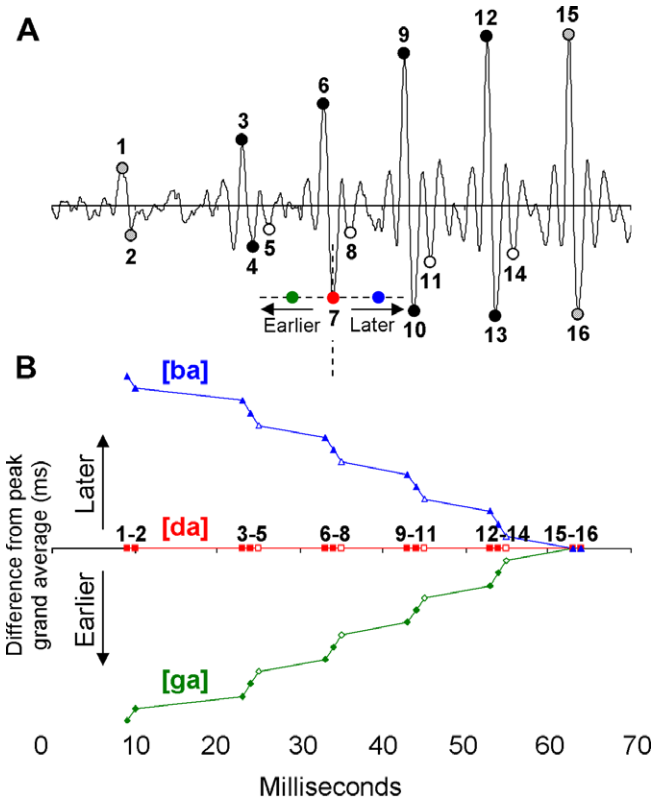

Fig. 2. (A) Representative time–amplitude waveform to illustrate the 16 peaks during the first 70 ms of the response for which latencies were identified. These were further classified as onset peaks (first two shaded circles), "major" peaks (filled circles), "minor" peaks (open circles), and end-point peaks (last two shaded circles). Based on our hypothesis, we predicted that [ga] peaks occur earliest, [ba] peaks occur latest, and [da] peaks occur in-between. This is schematically illustrated by the colored circles surrounding peak 7 ([ga] is green, [da] is red, [ba] is blue). (B) Schematic illustration of the hypothesized progression of normalized latencies ($Lat_{individual}$–Lat_{GM}, see *Analysis* section of Methods). We expected the normalized latency of [ga] to be negative (earlier), [da] to be around the mean, [ba] to be positive (later), and that latency differences would start out large and become equal at ~60 ms (end-point peaks, 15–16), where acoustic stimulus differences no longer exist.

peaks and minor peaks (where 3 is the number of stimulus conditions and K is the number of peaks).

2.4.2.2. Sustained measures. An FFT analysis of each response corresponding to a 40 ms portion of the steady-state response (110-150 ms) was performed to evaluate the spectral composition of the response. Average response magnitudes were calculated for 50 Hz-wide bins surrounding the frequency of the stimulus F_0 and the subsequent 10 harmonics. A repeated measure ANOVA was performed for each frequency bin to assess between-stimuli magnitude differences.

3. Results

3.1. Formant transition period

3.1.1. Transient measures

Table 2 shows the mean and standard deviation of the raw (non-normalized) latencies for the 16 peaks picked for each stimulus condition. It was hypothesized that during the formant transition period, the differing F_2 and F_3 frequencies would manifest themselves as latency shifts, with [ga], [da] and [ba] responses arising progressively later due to the high-to-low progression of frequency differences among the sounds. It was further hypothesized that the differences would lessen over the course of the response until they vanish completely at the time that the three syllables reached their shared steady-state portion. Fig. 2B shows a schematic representation of the hypothesized progression of the normalized latencies.

Grand average response waveforms of the first 70 ms for all three stimuli can be seen in Fig. 3A. A normalized latency plot for all 16 peaks is illustrated in Fig. 3B. There was no significant within-subject main effect of stimulus ($F_{(2,42)} = 1.717$, $p = 0.192$) or stimulus X peak interaction ($F_{(2,42)} = 0.851$, $p = 0.434$) for the onset peaks 1 and 2, or end-point peaks 15 and 16 (main effect: $F_{(2,42)} = 1.205$, $p = 0.310$; interaction: $F_{(2,42)} = 0.046$, $p = 0.955$). For the remaining peaks, the hypothesized pattern of latency shifts is visually evident and statistically significant. In addition, major peaks and minor peaks display a different degree of between-stimulus distinction, whereby minor peaks have greater between-stimulus latency differences than major peaks. Fig. 3C and D shows normalized latency plots for the major and minor peaks separately. Note that peaks 15 and 16 are plotted in both panels to illustrate the merging of the responses. Fig. 4 further highlights the response differences between the major and the minor peaks. This histogram shows the percent of subjects who displayed the hypothesized between-stimulus latency pattern ([ga] < [da] < [ba]) at each peak. It can be seen that a greater percentage of subjects show this pattern of activity for the minor peaks than that for the major peaks.

A 3 × 4 repeated measures ANOVA for the minor peaks showed a significant main effect of stimulus ($F_{(2,42)} = 353.625$, $p < 0.001$) and interaction ($F_{(6,126)} = 11.695$, $p < 0.001$). Multiple one-way repeated measure ANOVAs across stimuli were then performed on each of the four minor peaks separately. For each of the peaks, $F_{(2,42)}$ values were greater than 88.000 and p-values yielded significance of less than 0.001. Follow-up paired t-tests were performed to assess between-stimulus differences, and these results can be seen in Table 3. This table shows significant latency differences between all stimuli in all minor peaks. Individual data can be seen in Fig. 5. Fig. 5A shows the average (across all minor peaks) normalized latency for [ga], [da], and [ba], for each subject. It can be seen that 21 of 22 (~95%) display the hypothesized latency progression pattern when all minor peaks are collapsed. Furthermore, the asterisks indicate those subjects who displayed the hypothesized

Table 2
Mean and standard deviation of absolute latencies of each peak for responses to all stimuli

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Peak latency (ms)																	
/ga/	Mean	8.59	9.45	22.22	22.99	24.39	32.35	33.25	34.53	42.54	43.37	44.45	52.61	53.44	54.95	62.72	63.70
	SD	0.38	0.36	0.19	0.23	0.32	0.14	0.21	0.28	0.18	0.28	0.27	0.22	0.28	0.31	0.14	0.28
/da/	Mean	8.59	9.53	22.91	23.55	24.81	32.74	33.62	35.23	42.57	43.43	44.64	52.58	53.54	55.42	62.70	63.69
	SD	0.42	0.40	0.37	0.36	0.46	0.38	0.39	0.50	0.23	0.38	0.30	0.19	0.20	0.27	0.18	0.27
/ba/	Mean	8.67	9.54	23.15	24.04	25.98	32.95	33.98	35.95	42.73	43.80	45.75	52.74	53.74	55.83	62.68	63.68
	SD	0.45	0.26	0.28	0.36	0.49	0.28	0.28	0.37	0.19	0.26	0.26	0.17	0.24	0.29	0.15	0.23

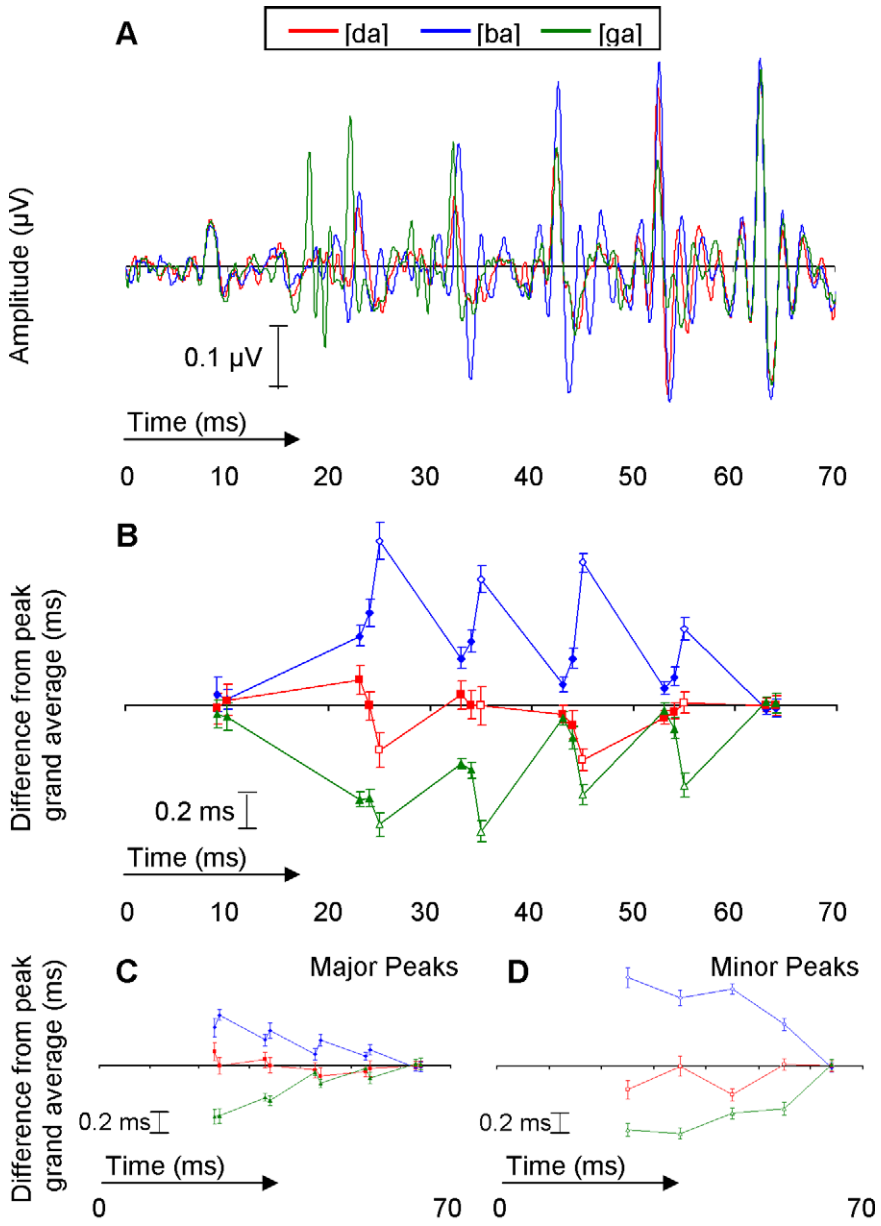

Fig. 3. (A) Time–amplitude grand averaged responses for the first 70 ms of [ga] (green), [da] (red), and [ba] (blue). (B) Normalized latency plot of all 16 peaks for [ga], [da], and [ba], where time (ms) is along the *x*-axis and time difference (in ms) from peak grand average is along the *y*-axis. Filled symbols indicate the major peaks, and open symbols indicate the minor peaks. (C) Normalized latency plots of the major peaks only. (D) Normalized latency plots for the minor peaks only. End-point peak 15 is plotted in both C and D.

latency pattern at all four minor peaks. Forty-one percent of subjects showed this pattern, while 95% of subjects showed this pattern on at least three of the four minor peaks, and 100% of subjects showed this pattern on at least two minor peaks.

A 3 × 8 repeated measures ANOVA for the major peaks showed a significant main effect of stimulus ($F_{(2,42)} = 111.593$, $p < 0.001$) and interaction ($F_{(14,294)} = 17.199$, $p < 0.001$). Multiple one-way repeated measure ANOVAs across stimuli were then performed on each of the eight major peaks separately. For each of the peaks, $F_{(2,42)}$ values were greater than 15.000 and *p*-values yielded significance of less than 0.001. Follow-up paired *t*-tests were performed to assess between-stimulus differences, and these results can be seen in Table 4. This table shows significant latency differences among all stimuli until peak 9, where there are no longer significant latency differences between [da] and [ga] (with the exception of peak 13, where this stimulus contrast just achieves significance). Fig. 5B shows individual data for the major peaks collapsed together. Although 86% of subjects displayed the hypothesized latency pattern, only 1 (~4%) subject displayed this pattern on all 8 major peaks (again, indicated by the asterisk).

From Fig. 5 it is obvious that the latency differences are more robust in the minor peaks than in the major peaks. In order to test this, the normalized latencies of [ba] were added to the absolute value of the [ga] normalized latencies. This produced a value for the degree of difference (in milliseconds) between the two most contrasted stimuli. The minor peaks (when averaged together) had an average [ba]-[ga] latency difference of 0.91 ms, and the major peaks had a 0.42 ms latency difference. A repeated measures ANOVA indicated a significant difference between the major and the minor peaks ($F_{(1,21)} = 249.364$, $p < 0.001$), indicating that the between-stimulus latency differences are significantly more robust in the minor peaks.

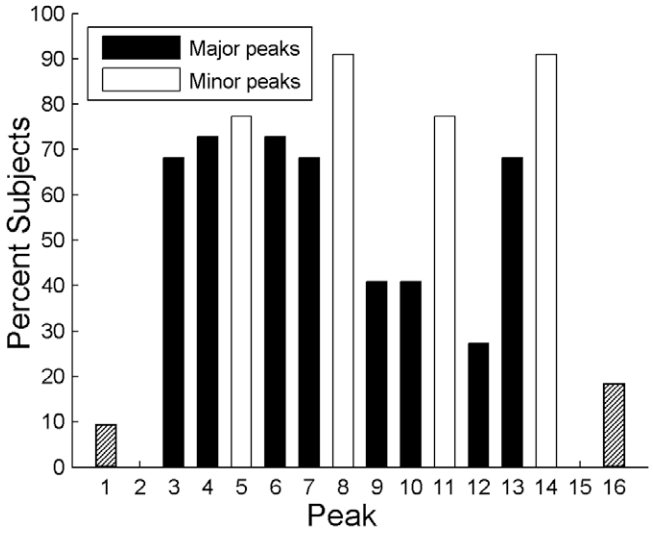

Fig. 4. Histogram of the percent of subjects displaying the hypothesized latency progression as a function of peak. The major peaks are shown as black bars, and minor peaks are shown as white bars. Onset and end-point peaks are shaded bars.

Lastly, we addressed the hypothesis that the between-stimulus latency differences lessen over the course of the formant transition period until they vanish completely at the time the three syllables reach a steady-state portion. This was implied by the interactions in the ANOVAs for both the major and the minor peaks described above, but to more parsimoniously test it, we used the overall [ba] – [ga] latency difference (discussed above), with the expectation that this latency difference would become progressively smaller at each peak. Repeated measures ANOVAs indicate that both the major peaks and the minor peaks show a significant latency difference as a function of peak ($F_{(7,147)} = 31.928$, $p < 0.001$; $F_{(3,63)} = 13.810$, $p < 0.001$, respectively). Fig. 6 illustrates the grand average progression for the major and the minor peaks separately. Note that peak 15 was added as an anchor point at which the between-stimulus differences no longer exist.

3.1.2. Frequency-domain measures

Fig. 7A shows the grand averaged response spectrum for each stimulus, for the 18–58 ms time range; for comparison, the stimulus' spectrum is shown in the inset. This time range was chosen as it encompasses the portion of the response where significant latency differences were observed among stimuli. Multiple one-way repeated measure ANOVAs across stimuli were then performed on each of the 10 frequency bins separately. For each bin, $F_{(2,42)}$ values were greater than 18.000 and p-values yielded significance of less than 0.001. Follow-up paired t-tests were performed to assess between-stimulus differences, and these results can be seen in Table 5. This table shows that the response to [ba] has significantly greater magnitudes at H_4 and H_5 than the response to either [da] or [ga]. Additionally, the [ba] response magnitude is significantly smaller than the other two responses at H_7. Lastly, the response to [ga] shows significantly greater magnitudes at H_9 and H_{10} compared to the responses of [da] and [ba]. These findings are illustrated as bar graphs in Fig. 7B.

Table 3
Paired t-test results for each stimulus contrast for the minor peaks

	5		8		11		14	
	t	p	t	p	t	p	t	p
Minor peaks								
ba vs da	10.642	**<0.001**	6.413	**<0.001**	16.706	**<0.001**	8.265	**<0.001**
ba vs ga	13.038	**<0.001**	15.598	**<0.001**	16.903	**<0.001**	12.343	**<0.001**
da vs ga	4.079	**0.001**	7.828	**<0.001**	2.459	**0.023**	6.291	**<0.001**

Significant between-stimulus normalized latency differences are indicated in bold.

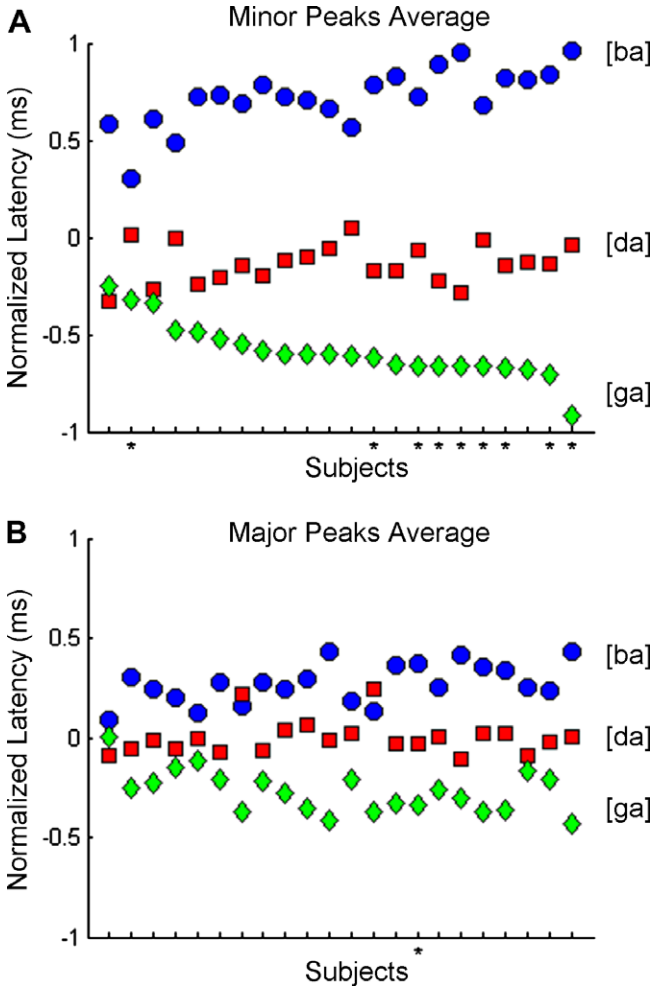

Fig. 5. Average normalized latency across all minor peaks (A) and major peaks (B) for each stimulus, for each subject (x-axis). Average [ba] latencies are blue circles, average [da] latencies are red squares, and average [ga] latencies are green diamonds. Subjects who displayed the hypothesized latency progression for all peaks within a class are indicated with an asterisk.

3.2. Steady-state portion

3.2.1. Transient response measures

Fig. 8A shows the grand averaged responses over the portion of the sustained response (90–160 ms) where no between-stimuli latency differences were hypothesized, because the acoustic properties of the stimuli are identical during this period. Fig. 8B shows the normalized latency plot for all peaks. No significant differences existed for the minor peaks (main effect: $F_{(2,42)} = 0.004$, $p = 0.996$; interaction: $F_{(12,252)} = 0.774$, $p = 0.678$) or the major peaks (main effect: $F_{(2,42)} = 3.032$, $p = 0.059$; interaction: $F_{(26,546)} = 0.531$, $p = 0.974$).

3.2.2. Sustained measures

Fig. 8C shows the grand averaged response spectra for each stimulus for the 110–150 ms time range; for comparison, the stimuli's spectra are shown in the inset. A 3×11 (3 stimuli, 11

Table 4
Paired *t*-test results for each stimulus contrast for the major peaks

	3		4		6		7		9		10		12		13	
	t	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*
Major peaks																
ba vs da	2.274	**0.034**	4.902	**<0.001**	3.233	**0.004**	4.883	<0.001	5.297	**<0.001**	5.875	**<0.001**	8.386	**<0.001**	5.591	**<0.001**
ba vs ga	12.405	**<0.001**	9.681	**<0.001**	10.594	**<0.001**	11.141	**<0.001**	5.258	**<0.001**	6.117	**<0.001**	3.817	**0.001**	4.741	**<0.001**
da vs ga	8.198	**<0.001**	5.375	**<0.001**	5.248	**<0.001**	4.034	**0.001**	0.570	0.575	0.861	0.399	−1.560	0.134	2.170	**0.042**

Significant between-stimulus normalized latency differences are indicated in bold.

frequency bins) overall repeated measures ANOVA showed no main effect of stimulus ($F_{(2,42)} = 1.203$, $p = 0.310$) and a significant stimulus × bin interaction ($F_{(20,420)} = 1.826$, $p = 0.016$). However, the origin of the interaction could not be determined through post hoc testing, as no significant differences were found when performing a multiple one-way repeated measures ANOVA on each of the 11 frequency bins separately. For each bin, $F_{(2,42)}$ values were less than 3.190 and *p*-values yielded significance of greater than 0.05. Thus, as hypothesized, there were no within-subject response magnitude differences across stimuli for the sustained responses.

4. Discussion

Ample literature addresses neural encoding of speech sounds from the 8th nerve (Delgutte, 1980; Miller and Sachs, 1983, 1984; Sachs and Young, 1980), cochlear nucleus (Caspary et al., 1977; Keilson et al., 1997; Palmer et al., 1986; Recio and Rhode, 2000; Rhode, 1998), and brainstem (Galbraith et al., 1995; Galbraith et al., 1997; Krishnan, 1999, 2002). The purpose of this study was to understand whether and how the aggregate brainstem response in humans reflects the subtle acoustic differences that exist among synthetic voiced stop consonants [ga], [da] and [ba], differing from one another only in their F_2 and F_3 frequency transitions, towards the steady-state portion. We have pursued this understanding by attempting to validate four hypotheses that encapsulate the current expectation of how speech is processed by the brainstem in the source-filter framework, expectations that have not yet been subjected to experimental confirmation in human subjects.

Because the F_2 and F_3 frequency ranges are well above the phase-locking capabilities of the brainstem, we hypothesized that these frequency differences would be manifested as latency differences among responses. Specifically, due to the tonotopicity of the auditory system, responses to [ga] would have the earliest latencies because it contains the highest F_2 and F_3 frequencies, responses to [ba] would have the latest latencies due to containing the lowest F_2 and F_3 frequencies, and responses to [da] would have latencies in-between the others. We further hypothesized that these latency differences would start off large, and progressively diminish over time as the acoustic properties of the syllables reached the steady-state. Moreover, the latencies and spectral composition of the responses over the duration of the steady-state portion were not expected to differ from one another because there are no acoustic differences in this segment of the stimuli. The results of this study support many of these hypotheses, and offer additional insight.

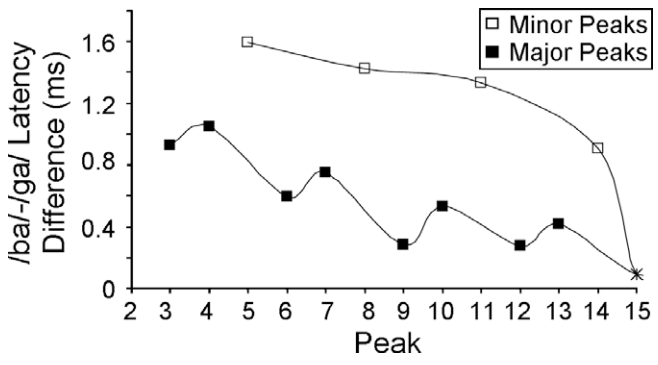

Fig. 6. Grand average normalized latency difference between [ba] and [ga] (in ms) at each peak. Minor peaks are plotted as open symbols, and major peaks are plotted as filled symbols. End-point peak 15 is plotted as an asterisk to indicate the merging of the responses.

Our results confirmed our first hypothesis that the differing F_2 and F_3 frequencies during the transition period would manifest themselves in terms of neural timing (synchrony) as latency shifts, with [ga], [da] and [ba] responses arising progressively later. Interestingly, major and minor peaks display dissimilar between-stimulus latency differences. The major peaks occur at the period of the fundamental frequency (~10 ms) and correspond (in time) to the glottal pulses in the stimuli, thereby imparting pitch or source information. The minor peaks reflect filter cues, or the formants, expressed in the time and frequency domains of the syllable.

The minor peaks revealed the most robust latency differences among stimuli. Not only are these latency differences well pronounced, but also they appear to be present on an individual level in the majority of subjects. Fig. 9A is a descriptive plot showing that all but one subject (left-most vertical line) followed the hypothesized latency progression in at least 3 of the 4 minor peaks (and 9 subjects followed it in all 4 peaks; asterisks). Furthermore, with the exception of one subject, when the hypothesized latency progression was not maintained, it was due to a peak latency reversal of [da] and [ga], such that [ga] had a later peak latency than [da]. In no case were there latency reversals of [ba] and [ga]. These findings indicate a few things. First, this supports the idea that the minor peaks are a neural representation of the filter cues contained in these stimuli, as they accurately and robustly reflect stimulus differences on an individual level. Second, it appears that [da] and [ga] have the most similar neural encoding, and thus responses to this consonant pair are the most likely to deviate from the hypothesized latency pattern. Lastly, [ba] and [ga] are the most dissimilar in terms of neural encoding. These final two points corroborate nicely what we know of the acoustic characteristics of the stimuli; [da] and [ga] are most similar in that their F_2 and F_3 are both falling, and [ga] and [ba] have the most dissimilar acoustic properties as they have the greatest frequency differences between F_2 and F_3, and the direction of their trajectories is opposite.

The stimuli were synthesized to have an identical and constant fundamental frequency throughout the duration of the syllable, and thus any F_0 perturbations (microprosody) have been controlled for. Because the major peaks reflect the F_0 of the stimulus, they may be expected to be identical among all syllables. Yet, between-stimulus latency differences were still observed for these major peaks. However, in contrast to the minor peaks, these latency differences were seen primarily at a group level, and were not robust within a given subject. The hypothesized latency pattern is not consistently ob-

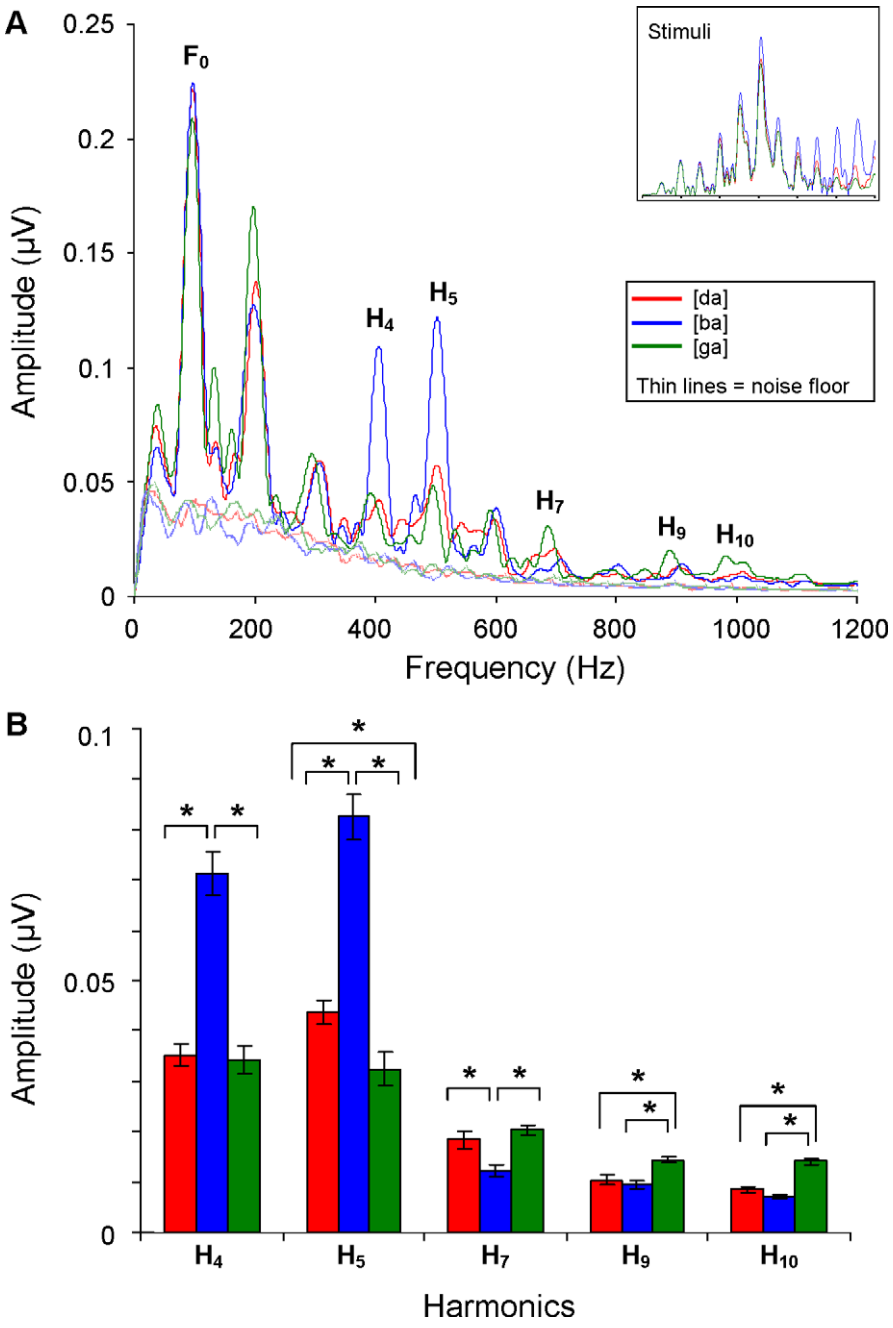

Fig. 7. (A) Grand average response FFT for [da] (red), [ba] (blue), and [ga] (green). Noise floor is indicated by thin lines. The fundamental frequency (F_0) and significantly different harmonic frequency bins are labeled. For comparison, an FFT of the onset of each stimulus (0–50 ms) from 0 to 1200 Hz is illustrated in the inset. (B) Bar graphs illustrating significant between-stimulus magnitude differences at H_4, H_5, H_7, H_9, and H_{10}. Brackets and asterisks indicate significant differences between stimuli.

Table 5
Paired *t*-test results for each stimulus contrast for harmonic frequency bins over the 18–58 ms range

	H_4		H_5		H_7		H_9		H_{10}	
	t	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*
FFT harmonics										
ba vs da	−8.356	**<0.001**	−11.290	**<0.001**	4.288	**<0.001**	1.200	0.244	2.310	0.031
ba vs ga	8.042	**<0.001**	12.188	**<0.001**	−5.814	**<0.001**	−4.024	**0.001**	−6.282	**<0.001**
da vs ga	0.310	0.760	4.897	**<0.001**	−1.551	0.136	−3.236	**0.004**	−4.898	**<0.001**

Significant between-stimulus magnitude differences are indicated in bold.

served across major peaks within a subject. For the majority of subjects, the hypothesized latency pattern is only followed at four or fewer of the eight major peaks, and the specific peaks differ across subjects (Fig. 9B). Only one subject showed the predicted latency progression at all eight major peaks (asterisk). Moreover, it does not appear that the major peaks show a distinct latency pattern of

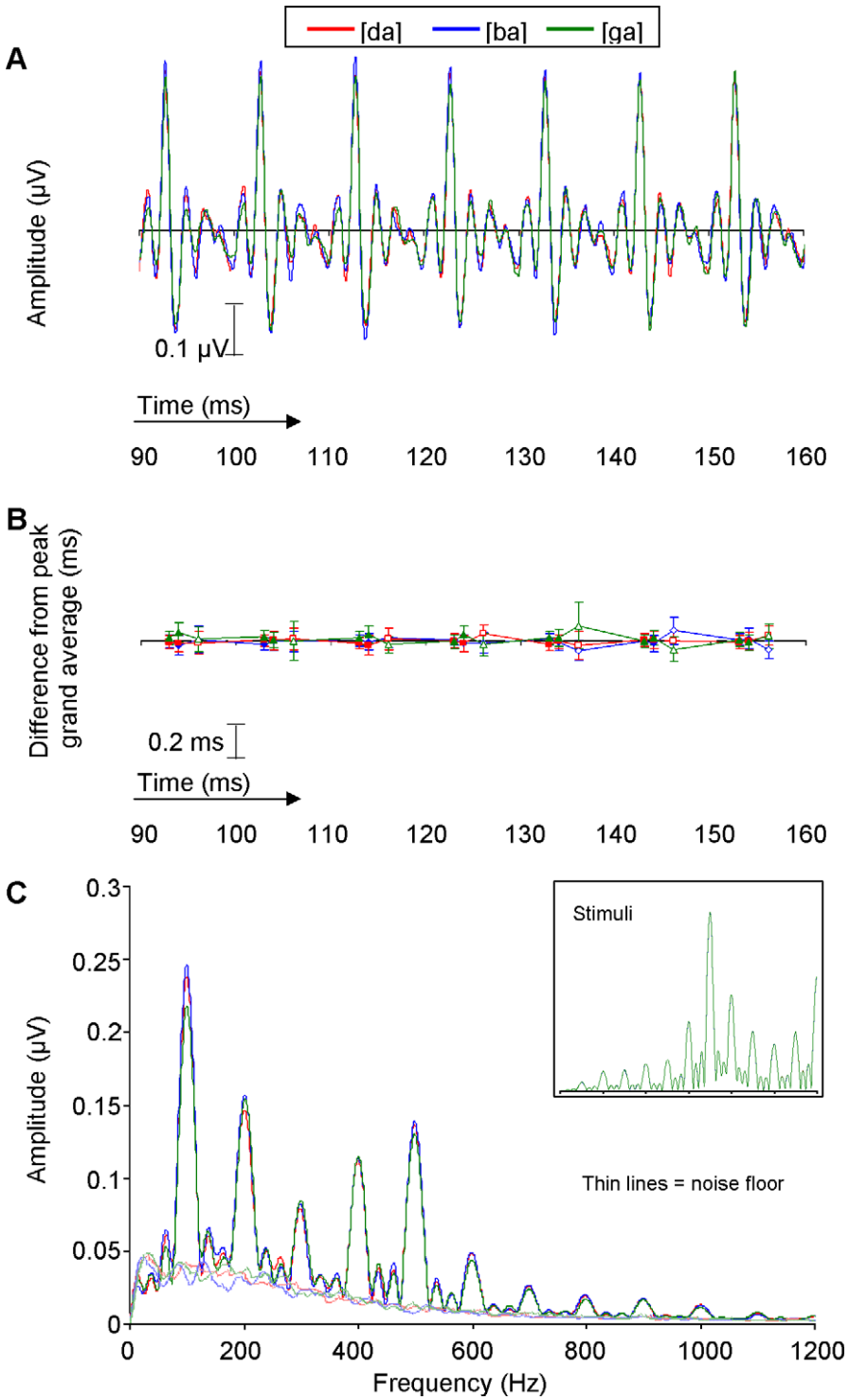

Fig. 8. (A) Time–amplitude grand averaged responses over the 90–160 ms portion of [ga] (green), [da] (red), and [ba] (blue). (B) Normalized latency plot of 21 peaks for [ga], [da], and [ba], where time (ms) is along the *x*-axis and time difference (in ms) from peak grand average is along the *y*-axis. Filled symbols indicate the major peaks, and open symbols indicate the minor peaks. (C) Grand average response spectra for [ga], [da], and [ba]. Noise floor is indicated by thin lines. For comparison, the spectrum of the stimulus steady-state portion from 0 to 1200 Hz is illustrated in the inset. Notice that only one color is seen because all stimuli are identical in this time frame.

any sort (one other than that hypothesized), as the distribution of between-stimulus latency differences appears to be varied among peak and subject (i.e., which stimulus had the earliest or latest latency at a given peak), although the [da] vs. [ga] reversal was again the most prevalent. This indicates that the acoustic dissimilarities contained within the stimuli do not necessarily translate into parallel neural disparities as they do for the minor peaks. We speculate that the latency differences observed in the major peaks are a reflection of the minor peak pattern (i.e., they are 'carried along'). On the other hand, in more natural utterances F_0 perturbations caused by articulatory movements in the vocal tract would likely be present. In this case (that is, without the enforced F_0 uniformity in the present stimuli), the systematic pattern observed in the minor peaks would be expected to be more evident in the major peaks as well.

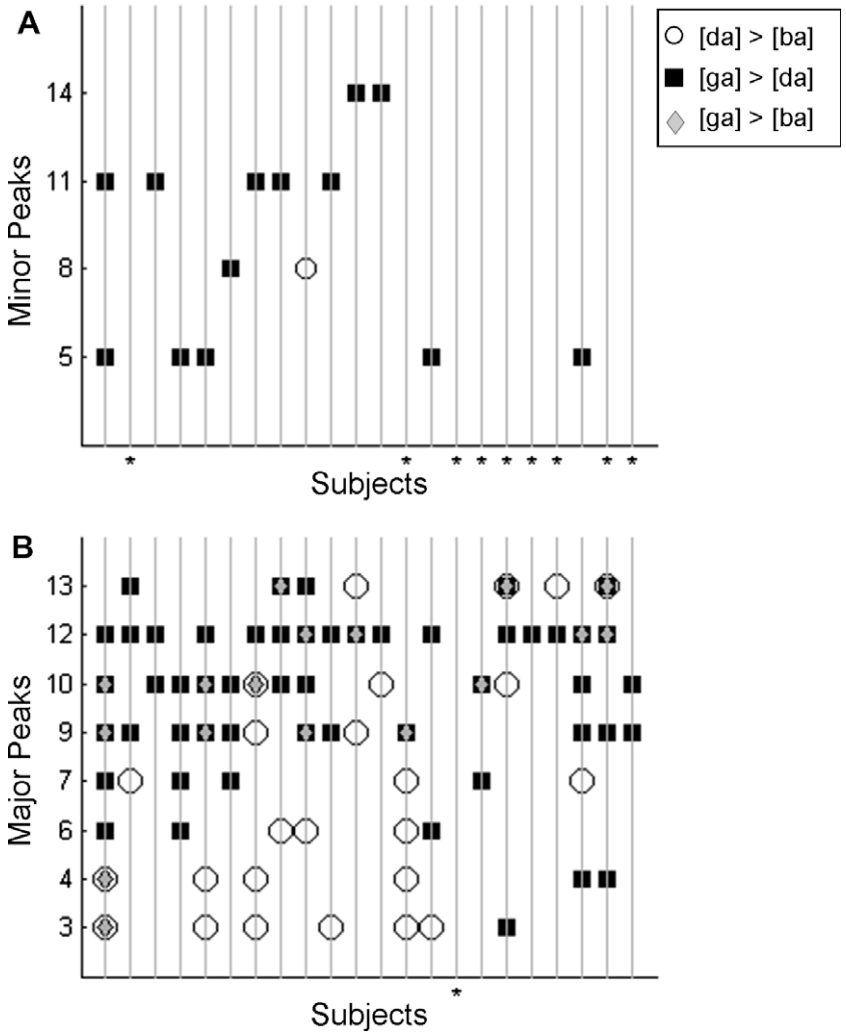

Fig. 9. The hypothesized peak latency progression across stimuli was that for a given peak, [ga] would have the earliest latency, [da] would have an intermediate latency, and [ba] would have the latest latency ([ga] < [da] < [ba]). Minor peaks (A) and major peaks (B) plotted for each subject (*x*-axis). Symbols indicate at which peak the hypothesized latency pattern was not followed, and which consonant pairs were the most likely to deviate from the hypothesized latency pattern. Subjects with no symbols along the vertical line followed the hypothesized latency pattern at all peaks (marked with an asterisk).

Our second hypothesis stated that between-stimulus latency differences diminish over the course of the response until vanishing completely by the time the three syllables reach their shared steady-state. While this pattern was observed for all major peaks and minor peaks, there were no stimulus-related latency differences in the onset response. This is likely because the onset response is triggered by the onset burst of the stimulus, and all three stimuli contained the identical initial frication of F_4 and F_5. Another interesting finding is that the stimulus by peak interaction was stronger for the major peaks, suggesting that across the stimuli the pattern of latency values across the major peaks is more different than it is for the minor peaks. This can be seen in Fig. 6, as the major peaks appear to have a more distinct merging trajectory than do the minor peaks. Furthermore, between-stimulus latency differences were significant for all minor peaks, whereas latency differences between [da] and [ga] began to disappear around peak 9 of the major peaks, a full 20 ms before the convergence at peak 15. This finding lends additional support to the idea that source and filter information are encoded by separate neural mechanisms, and that these mechanisms can be observed in the brainstem response as discrete neural events. Because the acoustic filter characteristics are consistently changing throughout the formant transition portion of the stimulus, neural responses to [ba], [da] and [ga] are also most distinct in the transition period. On the other hand, because the source is identical and constant throughout all stimuli, the neural response to this cue shows a greater propensity to be more similar, as it is phase-locking to the F_0.

The third hypothesis regarding the stimulus formant transition period stated that based on the phase-locking limitations of the brainstem, neural encoding of the periodic acoustic properties of speech would be concentrated in temporal events (phase-locking) below the second formant. As such, there would be little difference in response spectra at frequencies where the stimuli differ. An unexpected finding was seen in the spectral encoding of the [ba] harmonics at 400 and 500 Hz. At these two harmonics, the magnitude of the [ba] response was significantly greater than either the [da] or the [ga] response. Since the [ba] stimulus does not contain any more spectral energy than the other stimuli at either 400 or 500 Hz, we speculate that this magnitude enhancement is due to a distortion product. The F_1 of the [ba] stimulus ramps from 400 to 720 Hz, and the F_2 ramps from 900 to 1240 Hz. The difference between these two formants ($F_2 - F_1$) produces an almost flat 500 Hz difference tone, from which we believe the 500 Hz spectral enhancement arises. $F_2 - F_1$ distortion products have been identified in the auditory brainstem FFR when elicited by pure tones (Chertoff and Hecox, 1990; Pandya and Krishnan, 2004; Rickman

et al., 1991). However, to our knowledge, this is the first demonstration of a distortion product produced by a complex signal with a changing frequency trajectory. The explanation for the enhanced magnitude seen at 400 Hz of the [ba] stimulus is not as clear-cut, as no other formant subtraction would produce a 400 Hz difference tone. Perhaps this enhancement reflects the difference between the 500 Hz distortion product and the 100 Hz F_0. $F_2 - F_1$ distortion products were not evident in response to either the [da] or the [ga] stimuli because the formant subtraction of these two syllables does not produce a stable frequency as it does in the [ba] stimulus. The remaining magnitude differences observed at H_7, H_9, and H_{10} indicate that phase-locking cues are evident in brainstem responses up to ~1100 Hz, although, like H_5, they are not readily explainable by the frequency content of the stimuli.

The fourth, and final, hypothesis applies to the neural encoding of the steady-state portion of the stimuli. We hypothesized no between-stimulus latency or spectral differences in the response to this portion of the syllables, because the acoustic properties are identical across all syllables. This hypothesis was confirmed.

To our knowledge, this is the first study to examine how the human auditory brainstem encodes the acoustic differences between voiced consonant–vowel stop syllables. The spectral differences between [ga], [da], and [ba] were manifested primarily as systematic latency differences in the neural response, although the magnitude enhancement at 400 and 500 Hz in the [ba] response spectrum may also provide important information. Furthermore, within-subject latency differences across the three stimuli were most robust and evident for the minor peaks. From a practical standpoint, this means that it is not necessary to pick all 16 peaks in each response to evaluate normal brainstem encoding of these syllables. Rather, the information pertinent to normal speech encoding can be obtained by picking the four minor peaks in each response (12 peaks total). This is an important consideration for the clinical implications of these findings. First, numerous studies have reported abnormal neural encoding of the syllable [da] in children with language-based learning problems (Banai et al., 2005; Cunningham et al., 2001; Johnson et al., 2007; King et al., 2002; Russo et al., 2005; Wible et al., 2004a,b, 2005). Although important information pertaining to how neural encoding is disrupted in these children and how this may impact phonological processing has been gleaned from these studies, this work may help us to further specify the neural encoding challenges posed by learning difficulties in this population. For example, it may be the case that the auditory brainstem of a child with a language-based learning problem is not accurately encoding latency differences between a [ba] and a [da] syllable. This information has the potential to inform recommendations for remediation strategies, and allows for a remediation program that can be better tailored to suit a child's specific areas of weakness. Another obvious clinical population that could benefit from these results is the hearing impaired. By understanding which acoustic aspects of a syllable are being improperly encoded by the brainstem due to a hearing loss, one can envision more precise programming strategies of either hearing aids or cochlear implants.

This study has begun to provide a more comprehensive explanation of how the pitch, formants, and segmental-level timing cues in speech are encoded by the brainstem. Future work is underway that will describe the brainstem correlates of additional acoustic properties contained within speech such as voice onset and offset timing, changes in source information with constant filter information, and different vowels.

Acknowledgements

We thank Keli E. Rulf for her valuable assistance in data analysis. We would also like to thank the children and their families for participating in this study. This work was supported by NIH R01 DC01510.

References

Banai K, Abrams D, Kraus N. Sensory-based learning disability: insights from brainstem processing of speech sounds. Int J Audiol 2007;46:524–32.
Banai K, Nicol T, Zecker SG, Kraus N. Brainstem timing: implications for cortical processing and literacy. J Neurosci 2005;25:9850–7.
Blackburn CC, Sachs MB. Classification of unit types in the anteroventral cochlear nucleus: PST histograms and regularity analysis. J Neurophysiol 1989;62:1303–29.
Blumstein SE, Isaacs E, Mertus J. The role of the gross spectral shape as a perceptual cue to place articulation in initial stop consonants. J Acoust Soc Am 1982;72:43–50.
Blumstein SE, Stevens KN. Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants. J Acoust Soc Am 1979;66:1001–17.
Caspary DM, Rupert AL, Moushegian G. Neuronal coding of vowel sounds in the cochlear nuclei. Exp Neurol 1977;54:414–31.
Chen GD, Nuding SC, Narayn SS, Sinex DG. Responses of single neurons in the chinchilla inferior colliculus to consonant–vowel syllables differing in voice-onset time. Auditory Neurosci 1996;3:179–98.
Chertoff ME, Hecox KE. Auditory nonlinearities measured with auditory-evoked potentials. J Acoust Soc Am 1990;87:1248–54.
Clarey JC, Paolini AG, Grayden DB, Burkitt AN, Clark GM. Ventral cochlear nucleus coding of voice onset time in naturally spoken syllables. Hear Res 2004;190:37–59.
Cunningham J, Nicol TG, Zecker SG, Kraus N. Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. Clin Neurophysiol 2001;112:758–67.
de Gelder B, Vroomen J. Impaired speech perception in poor readers: evidence from hearing and speech reading. Brain Lang 1998;64:269–81.
Delattre P, Liberman AM, Cooper FS. Acoustic loci and transitional cues for consonants. J Acoust Soc Am 1955;27:769–73.
Delgutte B. Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers. J Acoust Soc Am 1980;68:843–57.
Frisina RD. Subcortical neural coding mechanisms for auditory temporal processing. Hear Res 2001;158:1–27.
Frisina RD, Smith RL, Chamberlain SC. Encoding of amplitude modulation in the gerbil cochlear nucleus: I. A hierarchy of enhancement. Hear Res 1990;44:99–122.
Galbraith GC, Arbagey PW, Branski R, Comerci N, Rector PM. Intelligible speech encoded in the human brain stem frequency-following response. Neuroreport 1995;6:2363–7.
Galbraith GC, Jhaveri SP, Kuo J. Speech-evoked brainstem frequency-following responses during verbal transformations due to word repetition. Electroencephalogr Clin Neurophysiol 1997;102:46–53.
Gorga M, Abbas P, Worthington D. Stimulus calibration in ABR measurements. In: Jacobsen J, editor. The auditory brainstem response. San Diego, Calif.: College-Hill Press; 1985. p. 49–62.
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.
Halle M, Hughes GW, Radley JPA. Acoustic properties of stop consonants. J Acoust Soc Am 1957;29:107–16.
Hillenbrand J, Gayvert RT. Vowel classification based on fundamental frequency and formant frequencies. J Speech Hear Res 1993;36:694–700.
Johnson K. Acoustic and auditory phonetics. Malden, MA: Blackwell; 2003.
Johnson KL, Nicol TG, Kraus N. The brainstem response to speech: a biological marker of auditory processing. Ear Hear 2005;26:424–34.
Johnson KL, Nicol TG, Zecker SG, Kraus N. Auditory brainstem correlates of perceptual timing deficits. J Cogn Neurosci 2007;19:376–85.
Joris PX, Schreiner CE, Rees A. Neural processing of amplitude-modulated sounds. Physiol Rev 2004;84:541–77.
Keilson SE, Richards VM, Wyman BT, Young ED. The representation of concurrent vowels in the cat anesthetized ventral cochlear nucleus: evidence for a periodicity-tagged spectral representation. J Acoust Soc Am 1997;102:1056–71.
Kewley-Port D. Time-varying features as correlates of place of articulation in stop consonants. J Acoust Soc Am 1983;73:322–35.
Kewley-Port D, Pisoni DB, Studdert-Kennedy M. Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. J Acoust Soc Am 1983;73:1779–93.
King C, Warrier CM, Hayes E, Kraus N. Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. Neurosci Lett 2002;319:111–5.
Klatt DH. Software for a cascade/parallel formant synthesizer. J Acoust Soc Am 1980;67:971–95.
Kraus N, Nicol TG. Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. Trends Neurosci 2005;28:176–81.
Krishnan A. Human frequency-following responses to two-tone approximations of steady-state vowels. Audiol Neurootol 1999;4:95–103.
Krishnan A. Human frequency-following responses: representation of steady-state synthetic vowels. Hear Res 2002;166:192–201.
Langner G. Neural processing and representation of periodicity pitch. Acta Otolaryngol Suppl 1997;532:68–76.

Langner G, Sams M, Heil P, Schulze H. Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography. J Comp Physiol A: Sensory, Neural Behav Physiol 1997;181:665–76.
Langner G, Schreiner CE. Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. J Neurophysiol 1988;60:1799–822.
Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. Psychol Rev 1967;74:431–61.
Martin BA, Sigal A, Kurtzberg D, Stapells DR. The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds/ba/ and/da/. J Acoust Soc Am 1997;101:1585–99.
McGee T, Kraus N, King C, Nicol T, Carrell TD. Acoustic elements of speechlike stimuli are reflected in surface recorded responses over the guinea pig temporal lobe. J Acoust Soc Am 1996;99:3606–14.
Merzenich MM, Knight PL, Roth GL. Representation of cochlea within primary auditory cortex in the cat. J Neurophysiol 1975;38:231–49.
Merzenich MM, Reid MD. Representation of the cochlea within the inferior colliculus of the cat. Brain Res 1974;77:397–415.
Miller MI, Sachs MB. Representation of stop consonants in the discharge patterns of auditory-nerve fibers. J Acoust Soc Am 1983;74:502–17.
Miller MI, Sachs MB. Representation of voice pitch in discharge patterns of auditory-nerve fibers. Hear Res 1984;14:257–79.
Palmer AR, Winter IM, Darwin CJ. The representation of steady-state vowel sounds in the temporal discharge patterns of the guinea pig cochlear nerve and primarylike cochlear nucleus neurons. J Acoust Soc Am 1986;79:100–13.
Pandya PK, Krishnan A. Human frequency-following response correlates of the distortion product at 2F1-F2. J Am Acad Audiol 2004;15:184–97.
Recio A, Rhode WS. Representation of vowel stimuli in the ventral cochlear nucleus of the chinchilla. Hear Res 2000;146:167–84.
Rhode WS. Neural encoding of single-formant stimuli in the ventral cochlear nucleus of the chinchilla. Hear Res 1998;117:39–56.
Rickman MD, Chertoff ME, Hecox KE. Electrophysiological evidence of nonlinear distortion products to two-tone stimuli. J Acoust Soc Am 1991;89:2818–26.
Rose JE, Galambos R, Hughes JR. Microelectrode studies of the cochlear nuclei of the cat. Bull Johns Hopkins Hosp 1959;104:211–51.
Russo N, Nicol TG, Zecker SG, Hayes EA, Kraus N. Auditory training improves neural timing in the human brainstem. Behav Brain Res 2005;156:95–103.
Sachs MB, Young ED. Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate. J Acoust Soc Am 1979;66:470–9.
Sachs MB, Young ED. Effects of nonlinearities on speech encoding in the auditory nerve. J Acoust Soc Am 1980;68:858–75.
Schreiner CE, Langner G. Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. J Neurophysiol 1988;60:1823–40.
Sinex DG, McDonald LP. Synchronized discharge rate representation of voice-onset time in the chinchilla auditory nerve. J Acoust Soc Am 1989;85:1995–2004.
Sinnott JM, Adams FS. Differences in human and monkey sensitivity to acoustic cues underlying voicing contrasts. J Acoust Soc Am 1987;82:1539–47.
Steinschneider M, Reser D, Schroeder CE, Arezzo JC. Tonotopic organization of responses reflecting stop consonant place of articulation in primary auditory cortex (A1) of the monkey. Brain Res 1995;674:147–52.
Steinschneider M, Schroeder CE, Arezzo JC, Vaughan Jr HG. Temporal encoding of phonetic features in auditory cortex. Ann NY Acad Sci 1993;682:415–7.
Stevens KN, Blumstein S. Invariant cues for place of articulation in stop consonants. J Acoust Soc Am 1978;64:1358–68.
Stevens KN, Klatt DH. Role of formant transitions in the voiced-voiceless distinction for stops. J Acoust Soc Am 1974;55:653–9.
Suga N, O'Neill WE. Neural axis representing target range in the auditory cortex of the mustache bat. Science 1979;206:351–3.
Summerfield Q, Haggard M. On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. J Acoust Soc Am 1977;62:435–48.
Tallal P. Auditory temporal perception, phonics, and reading disabilities in children. Brain Lang 1980;9:182–98.
Tallal P, Stark RE. Speech acoustic-cue discrimination abilities of normally developing and language-impaired children. J Acoust Soc Am 1981;69:568–74.
Tobey EA, Cullen Jr JK, Rampp DL. Effects of stimulus-onset asynchrony on the dichotic performance of children with auditory-processing disorders. J Speech Hear Res 1979;22:197–211.
Townsend TH, Schwartz DM. Error analysis on the California Consonant Test by manner of articulation. Ear Hear 1981;2:108–11.
Van Tasell DJ, Hagen LT, Koblas LL, Penner SG. Perception of short-term spectral cues for stop consonant place by normal and hearing-impaired subjects. J Acoust Soc Am 1982;72:1771–80.
Wang X, Sachs MB. Neural encoding of single-formant stimuli in the cat. II. Responses of anteroventral cochlear nucleus units. J Neurophysiol 1994;71:59–78.
Wible B, Nicol T, Kraus N. Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. Biol Psychol 2004a;67:299–317.
Wible B, Nicol T, Kraus N. Encoding of complex sounds in an animal model: Implications for understanding speech perception in humans. In: Heil P, König R, Budinger E, Scheich H, editors. The auditory cortex – a synthesis of human and animal research. Mahwah, NJ: Lawrence Erlbaum Associates; 2004b. p. 241–54.
Wible B, Nicol T, Kraus N. Correlation between brainstem and cortical auditory processes in normal and language-impaired children. Brain 2005;128:417–23.
Woodcock RW, McGrew KS, Mather N. Woodcock-Johnson III: tests of achievement. Itasca, IL: Riverside Publishing; 2001.
Young ED, Sachs MB. Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. J Acoust Soc Am 1979;66:1381–403.