

Published in final edited form as:

*Hear Res.* 2011 January ; 271(1-2): 103–114. doi:10.1016/j.heares.2010.04.008.

## Enhanced Physiologic Discriminability of Stop Consonants with Prolonged Formant Transitions in Awake Monkeys Based on the Tonotopic Organization of Primary Auditory Cortex

Mitchell Steinschneider<sup>1,2</sup> and Yonatan I. Fishman<sup>1</sup>

Mitchell Steinschneider: mitchell.steinschneider@einstein.yu.edu; Yonatan I. Fishman: yonatan.fishman@einstein.yu.edu

<sup>1</sup>Department of Neurology, Rose F. Kennedy Center, Room 322, 1300 Morris Park Avenue, Albert Einstein College of Medicine, Bronx, New York, USA 10461

<sup>2</sup>Department of Neuroscience, Rose F. Kennedy Center, Room 322, 1300 Morris Park Avenue, Albert Einstein College of Medicine, Bronx, New York, USA 10461

### Abstract

Many children with specific language impairment (SLI) have difficulty perceiving stop consonant-vowel syllables (e.g., /ba/, /ga/, /da/) with rapid formant transitions, but perform normally when formant transitions are extended in time. This influential observation has helped lead to the development of the auditory temporal processing hypothesis, which posits that SLI is causally related to the processing of rapidly changing sounds in aberrantly expanded windows of temporal integration. We tested a potential physiological basis for this observation by examining whether syllables varying in their consonant place of articulation (POA) with prolonged formant transitions would evoke better differentiated patterns of activation along the tonotopic axis of A1 in awake monkeys when compared to syllables with short formant transitions, especially for more prolonged windows of temporal integration. Amplitudes of multiunit activity evoked by /ba/, /ga/, and /da/ were ranked according to predictions based on responses to tones centered at the spectral maxima of frication at syllable onset. Population responses representing consonant POA were predicted by the tone responses. Predictions were stronger for syllables with prolonged formant transitions, especially for longer windows of temporal integration. Relevance of findings to normal perception and that occurring in SLI are discussed.

### Keywords

speech; place of articulation; specific language impairment; monkey A1; multiunit activity

---

© 2010 Elsevier B.V. All rights reserved.

**Corresponding Author:** Mitchell Steinschneider, M.D., Ph.D., mitchell.steinschneider@einstein.yu.edu Tel: 718-430-4115, Fax: 718-430-8588.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Author Information:** The authors declare no competing financial interests.

## Introduction

In one of the most influential and controversial observations in phonology, Tallal and colleagues reported that many children with specific language impairment (SLI) had difficulty in the perception of stop consonant-vowel (CV) syllables (e.g., /ba/, /ga/, /da/) with rapid formant transitions (e.g., 40 ms), but performed similarly to controls when formant transitions were extended in time (e.g., 80 ms) (for reviews see Tallal et al., 1993; Tallal, 2004). This fundamental observation has been replicated (e.g., Elliott et al., 1989; Kraus et al., 1996; Stark and Heinz, 1996; for review see Burlingame et al., 2005) and expanded to include both subjects with dyslexia and with deficits in other, non-linguistic, auditory perceptions that involve processing of rapidly changing sound stimuli (e.g., Hari and Kiesilä, 1996; Wright et al., 1997; Helenius et al., 1999; Laasonen et al., 2000; King et al., 2003;). Ultimately, this evolving and frequently contradictory body of evidence led to the formulation of the auditory temporal processing hypothesis, which posits that difficulties in processing rapidly changing sound stimuli is causally related to SLI (Tallal, 2004). In its latest form, individuals with SLI exhibit deficiencies in phonological functions because they aberrantly process rapidly changing sound stimuli in expanded temporal windows of integration, “chunking” acoustic events together in time periods too large to allow the fine-grained patterns of speech to be readily discriminated. This hypothesis has become so influential that it has spawned commercially available and clinically successful remediation programs (Merzenich et al., 1996; Tallal et al., 1996, 1998; Tallal, 2004; see also Kujala et al., 2001; Stevens et al., 2008).

While the temporal processing hypothesis remains highly contentious (e.g., Studdert-Kennedy and Mody, 1995; Bishop et al., 1999; Goswami, 2003;), temporal or other auditory processing deficits are likely causal in a subset of people with SLI (Heath et al., 1999; Habib, 2000; Ramus et al., 2003). Strong evidence that temporal processing deficits are directly relevant to SLI was provided by Benasich and Tallal (2002), who found that performance in 6–9 month infants on a rapid auditory discrimination task was predictive of later language impairments at 2 and 3 years of age. While not a demonstration of causality, this result does suggest that auditory processing plays an important role in the maturation of speech functions. Furthermore, this finding is consistent with modern views of developmental disorders, according to which fundamental deficits in neural processing (e.g., auditory temporal processing), often produced by a genetic anomaly, may only be reliably identified at early developmental stages (Karmiloff-Smith, 1998, see also Goswami, 2003). At later developmental stages, these early deficits may be partially ameliorated by compensatory mechanisms and yet result in more persistent, and higher-order deficiencies (e.g., SLI).

What remains unresolved in the consideration of these larger issues is the question of why some children with SLI perceive stop CVs with longer formant transitions more accurately than those with short formant transitions. Simply stating that the cause is a deficit in rapid temporal processing does not help to identify potential neural substrates of SLI. This issue is extremely important, as illuminating neural mechanisms potentially relevant for the improvement in phonetic discrimination might provide valuable insights into the competing hypotheses regarding the etiology of SLI and encourage modifications of remediation strategies. Thus, in order to more fully examine this question at a physiological level, data and hypotheses that directly address the differential perception of stop CV syllables must be incorporated into any explanatory framework for SLI.

Almost as contentious as the temporal processing hypothesis of SLI are the competing ideas regarding the perception of stop consonants. At one extreme, their perception is thought to be accomplished through a specialized, speech-specific module that reacquires the intended

phonetic gestures of the speaker (e.g., Liberman et al., 1967; Liberman and Mattingly, 1985). Thus, speech perception is intimately tied to speech production. At the other extreme, general principles of auditory processing are thought to be sufficient to map the acoustic signal onto discrete phonetic categories (e.g., Lotto and Kluender, 1998; Stevens, 1980; Kuhl, 2004). This hypothesis is further supported by evidence that humans and non-human animals share many similarities in speech perception (e.g., Kuhl, 1986; Lotto et al., 1997; Ramus et al., 2000; Sinnott and Gilmore, 2004). These similarities justify the use of non-human animal models to examine some of the fundamental questions concerning physiological mechanisms of speech perception.

One of the most prominent acoustically-based hypotheses put forth to explain perceptual differences in stop consonants argues that the short-term acoustic spectrum within the first 20 ms of consonant onset (release) is a major determinant of their differential perception (Stevens and Blumstein, 1978; Blumstein and Stevens, 1979, 1980; Chung and Blumstein, 1981). According to this hypothesis, onset spectra that are diffuse and display maxima at higher frequencies (diffuse rising) promote the perception of /d/, those that are diffuse and display maxima at lower frequencies (diffuse falling) promote the perception of /b/, while more compact spectra with maxima at mid-frequencies promote the perception of /g/. While modifications of this basic scheme have been made that incorporate the onset spectra at consonant onset within the context of later acoustic spectra occurring during the following vowel (e.g., Lahiri et al., 1984), the preceding explanation works well when onset spectra are not ambiguous (Alexander and Kluender, 2008). Furthermore, onset spectra are sufficient for stop consonant discrimination in neonates, children, and adults when stimuli are exceedingly short and lack prolonged vowels (< 50 ms), or when formant transitions are omitted from the synthesized syllables (Bertoncini et al., 1987; Ohde et al., 1995; Ohde and Haley, 1997; Ohde and Abou-Khalil, 2001). Overall, these results support the importance of onset spectra in the perception of stop consonants varying in their place of articulation (POA).

An attractive feature of this perceptual hypothesis is that it provides a physiologically plausible means by which stop consonants can be differentially represented at early stages of auditory cortical processing. Multiple studies have shown that complex vocalizations are differentially represented in primary auditory cortex (A1) of experimental animals based on the underlying tonotopic organization and the spectral content of the sounds (e.g., Creutzfeldt et al., 1980; Wang et al., 1995; Steinschneider et al., 2003; Mesgarani et al., 2008). This relationship between stop consonant spectra and tonotopic patterns of activity in A1 implies that activation by /b/ should be maximal in low best frequency (BF) regions of A1, while /g/ and /d/ should maximally excite mid and high BF regions of A1, respectively. Physiological results in A1 support this prediction when examining onset responses evoked by stop CV syllables (Steinschneider et al., 1995; Engineer et al., 2008).

Integrating the perceptual hypothesis of Stevens and Blumstein (1978) with both the physiological characteristics of A1 noted above and the temporal processing hypothesis of SLI leads to the two predictions that are the focus of this report. First, syllables with prolonged formant transitions should evoke better differentiated patterns of activation along the A1 tonotopic axis when compared to syllables with short formant transitions. This prediction is based on the premise that one consequence of lengthening formant transitions is to prolong the duration of spectral differences between the consonants prior to their convergence to common steady-state values. Second, the enhanced response differentiation for prolonged formant transition syllables relative to syllables with rapid formant transitions should be most evident when neural activity is analyzed within longer windows of temporal integration. This work expands our previous examination of POA representation in monkey A1, which only examined neural responses within the first 35 ms after stimulus onset (short

temporal integration window) and only to syllables with 40 ms formant transitions (Steinschneider et al., 1995).

## Materials and methods

Four male macaque monkeys (*Macaca fascicularis*), weighing between 2.6 and 3.9 kg at the time of surgery, were studied following approval by the Animal Care and Use Committee of Albert Einstein College of Medicine. Experiments were conducted in accordance with institutional and federal guidelines governing the use of primates, who were housed in our AAALAC-accredited Animal Institute and were under the continuous monitoring and care of the Animal Institute veterinarians. Other protocols were performed in parallel with this experiment to minimize the overall number of monkeys used in animal research. Monkeys were trained to sit comfortably in customized primate chairs. Surgery was then performed using sterile techniques and general anesthesia (sodium pentobarbital). Methods have been detailed in previous publications (e.g., Steinschneider et al., 2008). Briefly, holes were drilled into the skull to accommodate honeycomb-shaped matrices formed from 18-gauge stainless-steel tubes that allowed access to the brain. Matrices were angled to approximate the anterior-posterior and medial-lateral tilt of the superior temporal gyrus, guiding electrode penetrations orthogonally to the surface of A1. Matrices and Plexiglas bars permitting painless head fixation were embedded in dental acrylic and secured to the skull. Peri- and post-operative anti-inflammatory, antibiotic and analgesic agents were given. Recordings began 2 weeks after surgery.

Recordings were conducted in a sound-attenuated chamber. Monkeys maintained a relaxed, and generally alert state, facilitated by frequent human contact and delivery of juice and preferred food reinforcements. Brief periods of drowsiness were likely, however, and should be acknowledged. Animals were also monitored by video. Recordings were performed with multi-contact electrodes that contained 14 recording contacts arranged in a linear array evenly spaced at 150  $\mu\text{m}$  intervals (<10% error), permitting simultaneous recording across multiple A1 laminae. Contacts were 25  $\mu\text{m}$  stainless steel wires insulated except at the tip, and were fixed in place within the sharpened distal portion of a 30-gauge tube. Impedance of each contact was maintained at 0.1–0.4 M $\Omega$  at 1 kHz. The reference was an occipital epidural electrode. Headstage pre-amplification was followed by amplification ( $\times 5000$ ) with differential amplifiers (Grass P5, down 3 dB at 3 Hz and 3 kHz). Signals were digitized at a rate of 3400 Hz and averaged by Neuroscan software to generate auditory evoked potentials (AEPs). Previous testing in our lab has shown that spectral aliasing by very low amplitude, very high frequency signals has negligible effect on the AEPs, which are dominated by high amplitude signals of much lower frequency. Positioning of the electrodes was performed with a microdrive whose movements were guided by online inspection of AEPs and multiunit activity (MUA) evoked by 80 dB clicks. Tone bursts and speech sounds were presented when the recording contacts of the linear-array electrode straddled the inversion of the cortical AEP, and the largest evoked MUA was maximal in the middle electrode contacts. Response averages were generated from 50–75 stimulus presentations.

To derive MUA, signals were simultaneously high-pass filtered at 500 Hz, further amplified ( $\times 8$ ), full-wave rectified, and then low-pass filtered at 600 Hz (RP2 modules, Tucker Davis Technologies) prior to digitization. MUA measures the envelope of action potential activity generated by neuronal aggregates, weighted by neuronal location and size. MUA is similar to cluster activity, and has greater response stability than either cluster or single-unit activity (e.g., Nelken et al., 1994; Super and Roelfsema, 2005; Stark and Abeles, 2007). MUA patterns are sharply differentiated at 75–150  $\mu\text{m}$  spacing using the electrodes in the proposed studies, permitting resolution of net firing patterns confined to cortical sublaminae. Limitations of MUA recordings include: greater weighting of responses from larger cells

and those closer to the recording tip, an inability to define cell types, and the masking of activity contributed by single neurons.

One-dimensional current source density (CSD) analysis was used to physiologically define the laminar location of recording sites in A1. CSD was calculated from AEP laminar profiles using an algorithm that approximated the second spatial derivative of the field potentials recorded at three adjacent depths (Freeman and Nicholson, 1975). Depths of the earliest click-evoked and tone-evoked current sinks were used to locate lamina 4 and lower lamina 3 (e.g. Müller-Preuss and Mitzdorf, 1984; Steinschneider et al., 1992, 2008; Cruikshank et al., 2002). A later current sink in upper lamina 3 and a concurrent source located more superficially were almost always identified in the recordings and served as additional markers of laminar depth (e.g. Müller-Preuss and Mitzdorf, 1984; Steinschneider et al., 1992, 1994, 2003, 2008; Cruikshank et al., 2002). This physiological procedure was later checked by correlation with measured widths of A1 and its laminae at select electrode sites obtained from histological data (see below).

Isolated pure tones and synthetic CV syllables were presented. The tones were generated and delivered at a sample rate of 100 kHz by a PC-based system using RP2 modules. Tones ranged from 0.2 to 17.0 kHz and were 175 ms in duration, with linear rise/decay times of 10 ms. The syllables were constructed on the parallel branch of a KLSYN88a speech synthesizer, contained 4 formants (F1 through F4), and were also 175 ms in duration (Figure 1). Fundamental frequency ( $f_0$ ) began at 120 Hz and linearly fell to 80 Hz. All syllables shared the same F1 and F4, and only F2 and F3 changed among the syllables. Steady-state formant frequencies were 700, 1,200, 2,500, and 3,600 Hz, for F1 through F4, respectively. Onset frequencies for F2 and F3 of /ba/ were 800 and 2,000 Hz, 1,600 and 3,000 Hz for /da/, and 1,600 and 2,000 Hz for /ga/. Formant transitions for F2 and F3 were 20, 40, 60, or 80 ms in duration. F1 had an onset frequency of 200 Hz and a 30 ms formant transition, and F4 did not contain a formant transition. Thus, /ga/ had the same F2 as /da/ and the same F3 as /ba/. A 5 ms period of frication exciting F2-F4 preceded the onset of voicing. Amplitude of frication was increased by 18 dB at the start of F2 for /ba/ (800 Hz) and /ga/ (1600 Hz), and F3 for /da/ (3000 Hz). Spectral structure was such that /ba/ and /da/ had onset spectra maximal at either low or high frequencies, whereas /ga/ had an onset spectrum maximal at intermediate values. Measurements of the constructed syllables indicated that F1 in the steady-state portions of the sounds was 8.5 dB (S.D. = 1.0) louder than F2, with a resultant fall-off of 3.4 dB (S.D. = 1.4) between F2 and F3, and 4.8 dB (S.D. = 1.9) between F3 and F4. All stimuli were monaurally delivered via a dynamic headphone (MDR-7502, Sony, Inc.) coupled to a 3 inch long, 60 cc plastic tube placed adjacent to the ear contralateral to the recorded hemisphere with a stimulus onset asynchrony of 658 ms. Pure tone intensity was 60 dB SPL and syllable intensity was 80 dB measured with a Bruel and Kjaer sound level meter (type 2236) positioned at the opening of the plastic tube. The frequency response of the headphone was flattened ( $\pm 3$  dB) from 0.2 to 17.0 kHz by a graphic equalizer (GE-60, Rane, Inc.). Tones and then syllables were presented in randomized blocks that consisted of 50 artifact-free trials.

After completion of a recording series, animals were deeply anesthetized with pentobarbital and perfused through the heart with physiological saline and 10% buffered formalin. A1 was physiologically delineated by its typically large amplitude responses and by a BF map that was organized with low BFs located anterolaterally and higher BFs posteromedially (e.g. Merzenich and Brugge, 1973; Morel et al., 1993). Electrode tracks were reconstructed from coronal sections stained for Nissl and acetylcholinesterase, and A1 was anatomically identified using published criteria (e.g., Morel et al., 1993).

MUA analyzed in the present study was derived from the average of MUA recorded in three adjacent electrode channels spanning middle cortical layers. The lowest of these channels was co-located with the upper portion of the large initial current sink identified in the CSD profile (e.g., Steinschneider et al., 2008). This restriction ensured that analyses were dominated by MUA located in the lower half of lamina 3. BFs were defined as the tone frequency eliciting the largest amplitude MUA within the first 30 ms after stimulus onset (see Fishman and Steinschneider, 2009). MUA was normalized to baseline values occurring within the 25 ms prior to stimulus onset prior to subsequent analysis. Responses to the 0.8, 1.6, and 3.0 kHz tones were used to predict the responses to the speech sounds. At each recording site, the amplitudes of the responses to these three tones were ranked from largest to smallest. Amplitudes of the speech-evoked responses were then compared to the predicted ranking based on the tone responses. Responses occurring during multiple time intervals were analyzed, including 10–30, 10–50, 10–70, 10–90, and 10–110 ms post-stimulus onset. These physiologic intervals parallel expanding perceptual windows of temporal integration thought to be dysfunctional in some children with SLI.

## Results

Results are based on middle laminae MUA obtained from 47 A1 recording sites whose BFs were within the speech range (< 4 kHz). Technical issues limited the responses evoked by the 60 ms and 40 ms formant transition syllables to 46 and 45 recording sites, respectively. Thirteen sites had BFs < 1.0 kHz, 16 had BFs between 1.0 and 1.9 kHz, 9 between 2.0 and 2.9 kHz, and 9 between 3.0 and 4.0 kHz.

To ensure that there was no systematic modulation of the cortical activity based solely on the duration of the frequency-modulated component of the stimuli, responses evoked by the syllables with varying formant transition durations in the five expanding time windows of integration were subjected to a two-way ANOVA. Results are shown in Figure 2, which depicts mean response amplitudes under the four formant transition conditions within the five temporal integration windows examined. Responses to /ba/, /ga/, and /da/ were averaged together for this analysis. There was no effect of formant transition duration ( $F_{3,4} = 0.26$ ;  $p = 0.8553$ ). As expected, a large effect of temporal integration window was obtained ( $F_{3,4} = 34.56$ ;  $p < 0.0001$ ), with ever larger integration windows associated with greater excitatory activity. There was no interaction between main factors.

### Relationship between syllable-evoked activity and A1 tonotopy

Differential amplitude of A1 activity evoked by stop CV syllables can be partially predicted by the frequency selectivity of recording sites. This result is exemplified by the MUA recorded from a site with a BF of 3400 Hz (Figure 3). Figure 3A depicts the responses evoked by 0.8, 1.6, and 3.0 kHz tones, which were the center frequencies of the frication bursts occurring at the onsets of the syllables /ba/, /ga/, and /da/, respectively. Onset responses are followed by varying degrees of sustained activity and activity evoked by tone offsets. Frequency response functions based on MUA summed within two of the five integration windows (10–30 and 10–90 ms) used to analyze the data are illustrated in Figure 3B. Maximal activity in the 10–30 ms window following tone onset occurs at 3400 Hz (i.e., BF, solid line), with a pronounced fall-off in activity at higher tone frequencies and a more gradual decrease at lower frequencies.

Spectral tuning of the site depends upon the duration of the integration window. When the window is expanded to 10–90 ms (dotted line), the BF shifts slightly higher to 3600 Hz, responses evoked by tones near 3.0 kHz are diminished (right-sided arrow), and responses evoked by tones near 0.8 and 1.6 kHz are enhanced (left and center arrows, respectively). Integration windows of 10–50 and 10–70 ms yield intermediate frequency response

functions. In contrast, the 10–110 ms integration window displays a further shift in spectral specificity: BF is lowered to 1800 Hz, activity near 3.0 kHz continues to decrease, while responses elicited by tones between 0.8 and 2.2 kHz are further augmented such that they are larger than those elicited by the higher frequency tones (data not shown for figure clarity). This effect is due to the enhancement of sustained activity evoked by lower frequency tones and suppression of sustained responses to tones near 3.0 kHz (Figure 3A). Variations in frequency tuning over different temporal integration windows have been recently reported (Fishman and Steinschneider, 2009).

Invoking the spectral hypothesis put forth by Stevens and Blumstein (1978) to explain the differential perception of stop consonants varying in their POA, these tone-evoked response patterns were used to predict relative amplitudes of responses evoked by the syllables. For the 10–30, 10–50, and 10–70 ms integration windows, 3.0 kHz tones evoke larger responses than 1.6 kHz tones, which in turn evoke larger responses than 0.8 kHz tones. This spectral selectivity predicts that /da/, with its onset spectrum maximal near 3 kHz, will elicit the largest response when measured within these temporal windows (Rank 3), /ga/ an intermediate response because its onset spectrum is maximal near 1.6 kHz (Rank 2), and /ba/ the smallest response because its onset spectra is maximal near 0.8 kHz (Rank 1). Based on the spectral selectivity in the window of 10–90 ms, predictions change such that /ga/ should produce the largest response (Rank 3 because 1.6 kHz tones elicit a larger response in this time window than 3.0 kHz and 0.8 kHz tones), /da/ an intermediate response (Rank 2 because 3.0 kHz tones elicit a larger response than 0.8 kHz tones), and /ba/ the smallest response (Rank 1 because 0.8 kHz tones elicit a smaller response than the 1.6 and 3.0 kHz tones). Finally, for the 10–110 ms window, and using the same rationale based on tone-evoked responses, /ga/ should produce the largest response (Rank 3), /ba/ an intermediate response (Rank 2), and /da/ the smallest response (Rank 1).

These predicted rankings were then compared with the observed amplitudes of syllable-evoked responses. Figure 3C depicts the MUA evoked by the syllables segregated by formant transition duration. The MUA consists of onset responses, sustained activity with superimposed oscillations occurring at the syllable  $f_0$ , and offset responses. For purposes of illustration, amplitudes of responses evoked by syllables with 80 ms formant transitions are shown in Figure 3D. In the three shortest time windows, observed responses follow the aforementioned predictions: /da/ produces the largest response (Rank 3), /ga/ the second largest response (Rank 2), and /ba/ the smallest response (Rank 1). For longer integration windows, however, these predictions break down, as /da/ continues to evoke the largest response, whereas spectral tuning based on tone responses would have predicted an intermediate response in the 10–90 ms window and a smallest response in the 10–110 ms time window.

### Syllable-evoked activity of A1 populations can be predicted by their spectral tuning

Population-level analysis obtained by collapsing data sets across all formant transition conditions and all recording sites reveal that syllable-evoked activity can be predicted based on the relative activity evoked by the 0.8, 1.6, and 3.0 kHz tones. This result is illustrated in Figure 4, which depicts amplitudes of the syllable-evoked MUA ranked by the tone responses in all five temporal integration windows. Results of the planned series of one-way ANOVA tests are shown below each graph, while the outcomes of *post-hoc* tests (Newman-Keuls multiple comparisons) are shown above the bars in the graphs (legend at bottom right). ANOVA tests yield significant differences among ranks for all temporal integration windows. *Post hoc* analyses indicate that in the temporal windows from 10–30 ms through 10–90 ms, responses to syllables that are predicted to have the highest rank based on pure tone spectral tuning of the sites are significantly greater than those to the other syllables. This pattern partially breaks down in the longest 10–110 ms window, where responses to

syllables with predicted ranks of 3 and 2 are not significantly different. However, responses to the syllable predicted to have the lowest rank remain significantly less than the responses evoked by the other syllables.

### Enhanced discriminability of long formant transition syllables

The ability to partially predict speech-evoked activity based on tone responses is not equivalent among the syllables with varying formant transition durations. Predictions for syllables with progressively shorter formant transitions deteriorate with progressively longer temporal integration windows. This effect is shown in Figure 5, which depicts amplitudes of the syllable-evoked MUA ranked by the tone responses for each of the four syllables with different formant transition durations and for all but the longest (10–110 ms) temporal window (omitted for figure clarity). As in the previous figure, the results of planned one-way ANOVA tests are shown below each graph, and *post hoc* results are illustrated above the graph bars. When syllables contain 80 ms formant transitions, speech-evoked responses can be predicted by tone responses in the 10–30 and 10–50 ms temporal windows, and partially predicted in the later windows, including the 10–110 ms time window ( $F_{2,46} = 5.90$ ;  $p < 0.004$ ). As formant transition duration shortens progressively by 20 ms, there is a parallel decrease in the ability to predict the speech-evoked activity from the tones. This pattern progresses to the point where responses to the 20 ms formant transition syllables can be predicted only by tone-evoked responses occurring within the shortest 10–30 ms time window. This result is mirrored in *post hoc* analyses, which demonstrate the progressively reduced ability to predict the speech-evoked activity from the tone-evoked responses as formant transition duration decreases.

The above analyses indicate that tone-evoked responses can predict syllable-evoked responses in A1. This ability suggests that presenting syllables with different stop consonants having onsets dominated by low, middle, and high frequencies will produce different spatial patterns of activation across the tonotopic map of A1. Furthermore, the diminished ability of tone-evoked responses to predict responses evoked by syllables with shorter formant transitions also suggests that this differentiated A1 activation profile should deteriorate when speech-evoked responses are examined within longer temporal integration windows. To test this idea, we averaged together the frequency response functions of recording sites at which the largest response was evoked by each of the three syllables with formant transitions of 20 and 80 ms. The following two predictions were made on the basis of this analysis: First, average spectral tuning curves at sites where /ba/, /ga/, or /da/ elicit the largest response should have maxima at low, intermediate, and high BFs, respectively. Second, differences between spectral tuning curves should be more pronounced when derived from the largest responses evoked by the 80 ms formant transition syllables relative to those derived from responses to the 20 ms formant transition syllables.

Results of this analysis are shown in Figure 6, which illustrates average spectral tuning curves obtained when /ba/, /ga/, or /da/ elicited the largest syllable-evoked response. Curves are plotted separately for the different formant transition duration and temporal integration window conditions. When spectral tuning is based on responses evoked by the 80 ms formant transition syllables, well-differentiated population tuning curves are obtained for all four integration windows. Population tuning curves derived from sites at which /ba/ elicits the largest response display a peak at 0.8 kHz and fall off dramatically at higher tone frequencies. When /ga/ elicits the largest responses, curves display a plateau of near maximum activity at intermediate frequencies. Finally, when /da/ produces the largest response, curves display a peak at higher tone frequencies, and at lower frequencies, are generally below those based on responses to the other syllables. In contrast, average spectral tuning curves become markedly less differentiated for responses to syllables with 20 ms formant transitions, especially when longer temporal integration windows are analyzed.



Differentiation of spectral tuning based on formant transition duration is quantified by computing the average correlations among the three mean spectral tuning curves in the four temporal integration windows analyzed. Thus, within each temporal integration window, the correlations of tuning curves for /ba/ vs /ga/, /ba/ vs /da/, and /ga/ vs /da/ were computed. This procedure yielded 3 calculated correlation coefficients for each formant transition stimulus in each time window, and 12 values in all ( $3 \times 4$  integration windows). Results are shown in Figure 7, which depicts the average Pearson correlation coefficients for the 20 and 80 ms formant transition duration conditions. Spectral tuning functions are better differentiated (display lower correlations) when derived from activity evoked by syllables with 80 ms as compared with 20 ms formant transitions (Welsh corrected t test for unequal variances:  $t = 2.496$ ,  $df = 22$ ;  $p < 0.03$ ).

## Discussion

### 1) Summary of data

This report identifies two key features of the neural representation of stop consonants in monkey A1: 1) Spectral characteristics of the stop CV syllables /ba/, /ga/, and /da/ occurring at stimulus onset are sufficient to produce differential patterns of neural activity in A1 that are based on the underlying tonotopic organization. 2) A1 displays enhanced discrimination of stop CV syllables with prolonged formant transitions relative to those with short formant transitions when more prolonged windows of integration are used to characterize the neural activity. This differentiation occurs despite the synthetic syllables having highly similar acoustic attributes: all share common first and fourth formants, /ga/ and /da/ share common second formant transitions, and /ba/ and /ga/ share common third formant transitions. Thus, neural population responses in A1 are sensitive to subtle differences in the spectral and temporal characteristics of these complex acoustic stimuli.

### 2) Relationship to previous work examining tonotopic representation of complex sounds in A1

Present findings confirm previous observations that population responses in A1 differentiate CV syllables varying in their POA based on the relationship between the onset spectra of the syllables and the frequency selectivity of the neuronal populations (Steinschneider et al., 1995). The latter work, however, only examined responses evoked by CV syllables with 40 ms formant transitions and restricted measurements to the first 25 ms of cortical activity. Comparable relationships between single unit activity evoked by syllables varying in their consonant POA and the tonotopic organization of A1 have been reported in both rats and ferrets (Engineer et al., 2008; Mesgarani et al., 2008). Importantly, differential activation of human posterior superior temporal gyrus by speech sounds varying in their consonant POA has been shown by fMRI to be related to the spectral content of the syllables at their onset (Obleser et al., 2007). Thus, the tonotopic organization of this speech feature appears to be a general property of auditory cortex in mammals, including humans. This conclusion represents a specific example of the more general rule that spectral components of complex sounds are encoded by the spatial distribution of population activity along the tonotopic map of A1 (e.g., Creutzfeldt et al., 1980; Wang et al., 1995; Steinschneider et al., 2003).

### 3) Relationship to studies examining physiological windows of temporal integration

In the present study, it need not have been the case that responses to syllables with rapid formant transitions would only show a statistically reliable relationship to the spectral tuning of A1 sites for short temporal integration windows. For instance, responses encoding the onset spectra of the syllables might have persisted throughout the duration of the syllables. The finding that responses become better differentiated within progressively shorter time windows as formant transitions are shortened indicates that A1 represents in real time the

spectro-temporal features of the complex sounds. Relationships between the spectral content of the syllables and the tonotopic organization of A1 could also have broken down because amplitudes of tone-evoked responses could have significantly varied with the length of the temporal integration window. While this pattern occasionally occurred, it is unlikely that this potential confound would have significantly modified the overall results. This conclusion is supported by the observation that frequency tuning functions of responses to 60 dB SPL tones varied minimally between MUA occurring within the first 30 ms following stimulus onset and that occurring between 30 and 100 ms after onset (Fishman and Steinschneider, 2009). Only within later time windows were significant changes in the frequency tuning of neural populations observed relative to the tuning of the "onset" response.

The finding that rapidly changing spectral characteristics of syllables (e.g., those with 20 ms formant transitions) are maximally differentiated with short temporal windows of integration is in keeping with related studies that have studied the temporal windows of spike integration that afford optimal neural differentiation of complex vocalizations. In their study of A1 in barbiturate-anesthetized marmosets, Wang and colleagues (1995) found that a spike integration window of 20–40 ms was optimal for single cell discrimination of con-specific vocalizations. Importantly, these authors also observed that neural discrimination was not determined by the rapidly changing spectral components of the vocalizations, but by the introduction of discrete stimulus events. This finding suggests that the syllable-evoked responses examined here are largely determined not by formant transitions (i.e., rapidly changing spectral components), but rather by the encoding of the spectral content of the syllables within discrete time windows. Similar optimal temporal windows of spike integration ranging from 10–50 ms have been determined for rat A1 in the differential representation of human speech sounds, for ferret A1 in the representation of hetero-specific vocalizations, for field L of zebra finches in the discrimination of con-specific bird songs, and most germane to the translational relevance of the present study, for human auditory cortex in the encoding of speech (Boemio et al., 2005; Narayan et al., 2005; Schnupp et al., 2006; Engineer et al., 2008; Poeppel et al., 2008).

#### 4) Population responses in awake, but non-behaving, animals

In this study, animals were awake but not actively engaged in sound processing. It is thus reasonable to question whether the physiological responses acquired in this passive, awake state can be extrapolated to conditions in which the animals are actively performing a relevant behavioral task. Compared to responses acquired in the passive, awake state, neurons in core regions of auditory cortex did not significantly change their tuning characteristics when animals were performing a sound localization task (Scott et al., 2007). Basic temporal response properties of A1 neurons in trained ferrets were preserved in naïve animals (Schnupp et al., 2006). While spectral receptive fields in A1 of behaving ferrets were significantly modified during behavior, changes were maximal near the BF of the cells and tended to enhance this excitatory response when the target frequency overlapped the BF while decreasing responses to surrounding frequencies. There was little effect on the tuning of cells whose BFs were at a distance from the target frequency (Fritz et al., 2007). Thus, if animals were trained to discriminate /ba/, /ga/ and /da/, spectral tuning at sites whose BFs were near 800 (/ba/), 1600 (/ga/) and 3000 Hz (/da/) would likely be enhanced relative to the present findings obtained in the passive, awake state. These considerations suggest that the present findings would not have significantly changed if the monkeys were engaged in a speech sound discriminative task. Further, the awake but non-behaving state may be an effective model for examining the type of neural responses occurring in young infants when passively exposed to syllables varying in their consonant POA.

### 5) Do monkeys process consonant POA similarly to humans?

A fundamental question is whether monkeys can serve as a reasonable model for examining neural activity associated with perception of POA. Kuhl and Padden (1983) examined the differential perception of /b/, /d/, and /g/ in macaques by comparing the discrimination of syllable pairs along a 15-step continuum that varied in the starting frequency of second formants. They found enhanced discriminability of stimulus pairs when they crossed human phonetic boundaries, suggesting a similarity in perceptual mechanisms. Sinnott and Gilmore (2004) found excellent discrimination by monkeys of /b/ and /d/, especially when followed by the vowels /a/ and /u/. While discrimination was always inferior relative to human subjects, and apparent differences in the means by which monkeys discriminated /b/ from /d/ relative to adults were noted, the authors concluded that monkeys may serve as a valuable model for pre-verbal speech perception occurring in human infants.

### 6) Support for the spectral hypothesis of Stevens and Blumstein

Present data support the perceptual hypothesis that onset spectra are important cues for the differential perception of stop consonants varying in their POA, and provide a physiologically plausible means by which this discrimination occurs (Stevens and Blumstein, 1978; Blumstein and Stevens, 1979, 1980; Chung and Blumstein, 1981). In this and other animal studies, the neural discrimination of stop consonants and other complex vocalizations is at least partly based on the tonotopic organization of A1 and the distributed representation of the spectral components of these sounds (see previous citations). Human core auditory cortex shares a similar tonotopic organization (Howard et al., 1996; Formisano et al., 2003), and thus it is reasonable to predict that a similar pattern of response discrimination for stop consonants will occur in Heschl's gyrus. Further, longer formant transitions have been shown to enhance differential perception of stop consonants varying in their POA, and it was argued that this acoustic modification improved discrimination by prolonging spectral differences among the syllables (Van Wieringen and Pols, 1995). This psychoacoustic finding parallels the enhanced physiological discrimination of syllables with prolonged formant transitions observed in the present study.

While neural activity in monkey A1 supports the importance of onset spectra in the perception of POA, this cue is not the only relevant stimulus feature. Formant transitions, onset frequency of the 2<sup>nd</sup> formant, frication burst duration and amplitude, and late onset of low frequency energy relative to the burst have all been shown to be important (e.g., Kewley-Port, 1983; Ohde and Stevens, 1983; Walley and Carrell, 1983; Smits et al., 1996; Hedrick and Younger, 2007; Alexander and Kluender, 2008). We did not study these parameters, and thus cannot describe the effects they would have on A1 neural activity. However, their importance does not negate the relevance of onset spectra which was shown even in studies identifying other factors important for perception (e.g., see above citations). Further, onset spectral cues may be especially important in connected speech, noisy environments, and in people with sensorineural hearing loss (Hedrick and Younger, 2007; Alexander and Kluender, 2008, 2009).

### 7) Relevance to language learning

The importance of onset spectra for the discrimination of POA and its neural representation in A1 is carried over into early childhood when language learning is most pronounced. Neonates can discriminate CV syllables varying in their consonant POA when they are shortened to only 34–44 ms, suggesting that onset spectra are sufficient for their differential perception (Bertoncini et al., 1987). Children ranging in age from 3 to 11 years utilize burst spectra in their discrimination of POA, though other cues such as formant onset frequencies and formant transition information are also perceptually important (Walley and Carrell, 1983; Ohde et al., 1995; Ohde and Haley, 1997; Mayo and Turk, 2004).

## 8) Relevance to SLI

Current findings are compatible with the auditory temporal processing hypothesis concerning the etiology of SLI, which proposes that deficiencies in phonological functions are based on aberrant processing of rapidly changing sound stimuli due to expanded temporal windows of integration (Tallal, 2004). As discussed in the Introduction, a central finding that has helped drive the development of this hypothesis is the observation that many children with SLI have difficulty in discriminating stop CV syllables with rapid formant transitions, yet perform normally when the transitions are lengthened. Present results suggest a physiological basis for these perceptual findings. According to this scheme, stop CV syllables with rapid formant transitions (e.g., 20 and 40 ms) are not well discriminated by children with SLI because their prolonged temporal window of integration (e.g., 80 ms) obscures differences in the distribution of auditory cortical activity encoding the spectra of the syllables. This obscuration is corrected when formant transitions are prolonged (e.g., 80 ms).

Given the heterogeneity of SLI (e.g., Banai and Ahissar, 2004), the multiplicity of cues available for discriminating consonant POA, and the widely distributed neural networks engaged in speech processing (e.g., Binder et al., 2000; Obleser et al., 2007; Leff et al., 2008), it is reasonable to conclude that neither the temporal processing hypothesis nor the physiological properties of A1 can solely account for the mechanisms responsible for SLI. With regard to heterogeneity, it has recently been reported that while dysfunctional processing of rapidly changing acoustic cues is somewhat predictive of SLI, better correlations were obtained in a cohort of 7 to 11 year old children when perceptions of durational and amplitude envelope sound cues were tested (Corriveau et al., 2007). It is interesting to speculate that prolonged temporal windows of integration could also account for deficits in these perceptual tests, but further studies will be needed to examine this idea. Of greater relevance, however, is the fact that deficits in processing basic sound cues were still predictive of SLI. This suggests that fundamental mechanisms of auditory cortical physiology are implicated in the dysfunction of basic language skills. Perceptual evidence includes studies in children with dyslexia, where a subgroup had difficulty discriminating POA in consonant clusters and within vowel-consonant-vowel syllables (Adlard and Hazan, 1988). Errors in discrimination were based on acoustic similarity of the speech sounds associated with these errors, suggesting dysfunctional encoding of their spectral content. Physiological evidence includes data from adults with dyslexia, who display aberrant N1 components of the AEP elicited by both speech and non-speech sounds, indicating that deficits begin in early stages of auditory cortical processing (Nagarajan et al., 1999; Giraud et al., 2005; Moisescu-Yilflach and Pratt, 2005).

Present data are also compatible with hypotheses that SLI is due to abnormal attentional mechanisms (Hari and Renvall, 2001; Petkov et al., 2005). According to the hypothesis proposed by Hari and Renvall (2001), time is “chunked” into prolonged windows of integration due to sluggish shifting of attention. Once again, more prolonged integration windows would blur the tonotopic representations of rapidly changing formant transitions. Supportive physiological evidence for an attentional contribution to SLI includes data demonstrating that children with SLI failed to show enhancement of the N1 AEP component evoked by probe stimuli during a selective attention task (Stevens et al., 2006). This failure indicates that attentional mechanisms are dysfunctional at early auditory processing stages in SLI, and possibly limit the differential patterns of consonant-evoked activity in A1 to be identified.

While any relationship between A1 physiology in monkeys and SLI must remain speculative, it is crucial that detailed studies afforded by experimental animals be selectively applied to evaluate prevailing hypotheses concerning the development of both normal and

aberrant speech perception. In this study, we have shown that the representation of stop CV syllables across tonotopically-organized A1 supports the specific perceptual hypothesis that encoding of onset spectra is sufficient to differentially encode stop consonants in syllable-initial position. This representation, when examined within the contexts of both short and progressively prolonged temporal windows of integration, provides a plausible explanation for a key observation seen in some children with SLI. Further, the present findings are consistent with the hypothesis that prolonged windows of sensory integration, at least in a subset of affected individuals, may contribute to SLI. Finally, the present results suggest that speech remediation therapies may be effective, in part, by sharpening these extended temporal windows of integration. Future work in experimental animals will undoubtedly enhance understanding of both normal and abnormal mechanisms involved in speech perception.

## List of Abbreviations

<b>SLI</b>	specific language impairment
<b>POA</b>	place of articulation
<b>CV</b>	consonant-vowel
<b>BF</b>	best frequency
<b>AEP</b>	auditory evoked potential
<b>MUA</b>	multiunit activity
<b>CSD</b>	current source density

## Acknowledgments

The authors thank Drs. Charles E. Schroeder and David H. Reser, and Ms. Jeannie Hutagalung for their invaluable technical assistance. Supported by National Institute of Deafness and Other Communications Disorders Grant DC-00657.

## References

- Adlard A, Hazan V. Speech perception in children with specific reading difficulties (dyslexia). *Quart. J. Exp. Psychol* 1998;51A:153–177.
- Alexander JM, Kluender KR. Spectral tilt change in stop consonant perception. *J. Acoust. Soc. Am* 2008;123:386–396. [PubMed: 18177168]
- Alexander JM, Kluender KR. Spectral tilt change in stop consonant perception by listeners with hearing impairment. *J. Speech, Lang. Hear. Res* 2009;52:653–670. [PubMed: 18952854]
- Banai K, Ahissar M. Poor frequency discrimination probes dyslexics with particularly impaired working memory. *Audiol. Neurootol* 2004;9:329–340.
- Benasich AA, Tallal P. Infant discrimination of rapid auditory cues predicts later language impairment. *Behav. Brain Res* 2002;136:31–49. [PubMed: 12385788]
- Bertoncini J, Bijeljac-Babic R, Blumstein SE, Mehler J. Discrimination in neonates of very short CVs. *J. Acoust. Soc. Am* 1987;82:31–37. [PubMed: 3624638]
- Binder JR, Frost JA, Hammeke TA, Bellgowan PSF, Springer JA, Kaufman JN, Possing ET. Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex* 2000;10:512–528. [PubMed: 10847601]
- Bishop DVM, Carlyon RP, Deeks JM, Bishop SJ. Auditory temporal processing impairment: Neither necessary nor sufficient for causing language impairment in children. *J. Speech, Lang. Hear. Res* 1999;42:1295–1310.

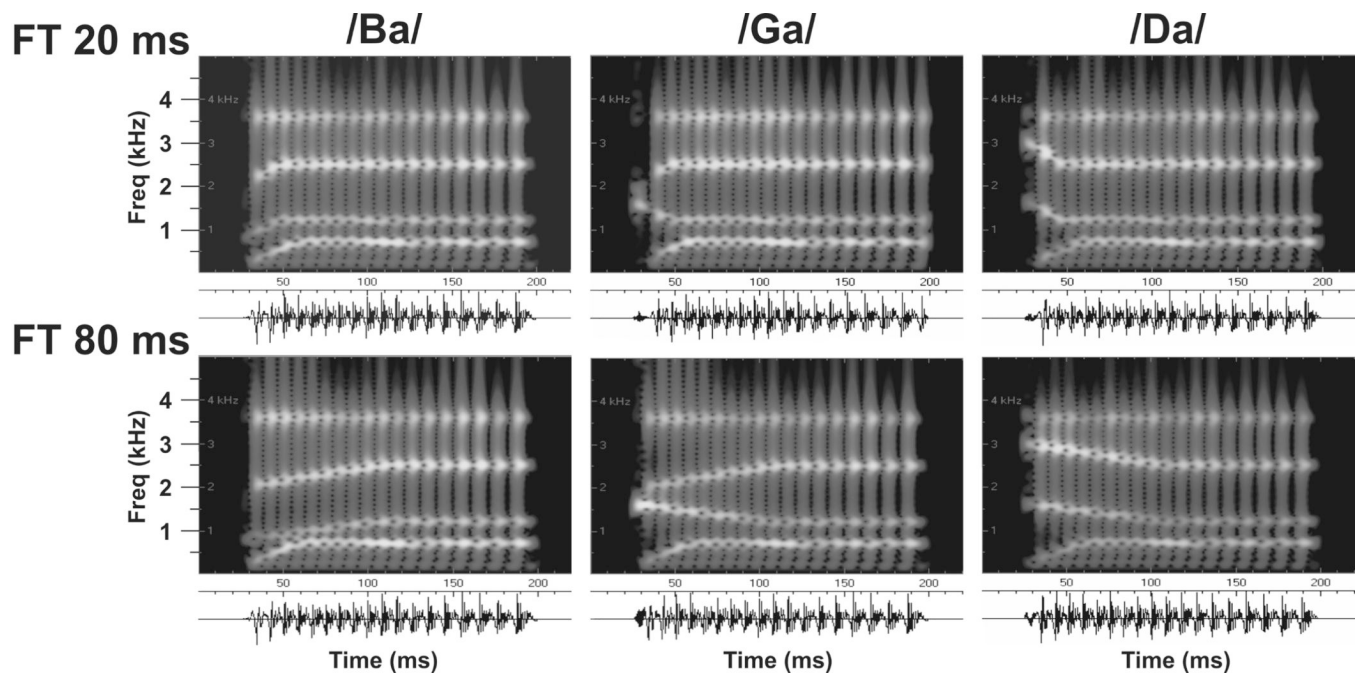
- Blumstein SE, Stevens KN. Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *J. Acoust. Soc. Am* 1979;66:1001–1017. [PubMed: 512211]
- Blumstein SE, Stevens KN. Perceptual invariance and onset spectra for stop consonant vowel environments. *J. Acoust. Soc. Am* 1980;67:648–662. [PubMed: 7358906]
- Boemio A, Fromm S, Braun A, Poeppel D. Hierarchical and asymmetrical temporal sensitivity in human auditory cortices. *Nature Neurosci* 2005;8:389–395. [PubMed: 15723061]
- Burlingame E, Sussman HM, Gillam RB, Hay JF. An investigation of speech perception in children with specific language impairment on a continuum of formant transition duration. *J. Speech, Lang. Hear. Res* 2005;48:805–816. [PubMed: 16378475]
- Chang S, Blumstein SE. The role of onsets in perception of stop consonant place of articulation: Effects of spectral and temporal discontinuity. *J. Acoust. Soc. Am* 1981;70:39–44. [PubMed: 7264070]
- Corriveau K, Pasquini E, Goswami U. Basic auditory processing skills and specific language impairment: A new look at an old hypothesis. *J. Speech, Lang. Hear. Res* 2007;50:647–666. [PubMed: 17538107]
- Creitzfeldt O, Hellweg F-C, Schreiner C. Thalamocortical transformation of responses to complex auditory stimuli. *Exp. Brain Res* 1980;39:87–104. [PubMed: 6247179]
- Cruikshank SJ, Rose HJ, Metherate R. Auditory thalamocortical synaptic transmission *in vitro*. *J. Neurophysiol* 2002;87:361–384. [PubMed: 11784756]
- Elliott LL, Hammer MA, Scholl ME. Fine-grained auditory discrimination in normal children and children with language-learning problems. *J. Speech, Lang. Hear. Res* 1989;32:112–119.
- Engineer CT, Perez CA, Chen YH, Carraway RS, Reed AC, Shetake JA, Jakkamsetti V, Chang KQ, Kilgard MP. Cortical activity patterns predict speech discrimination ability. *Nature Neurosci* 2008;11:603–608. [PubMed: 18425123]
- Fishman YI, Steinschneider M. Temporally dynamic frequency tuning of population responses in monkey primary auditory cortex. *Hearing Res* 2009;254:64–76.
- Formisano E, Kim D-S, Di Salle F, van de Moortele P-F, Ugurbil K, Goebel R. Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 2003;40:859–869. [PubMed: 14622588]
- Freeman JA, Nicholson C. Experimental optimization of current source density techniques for anuran cerebellum. *J. Neurophysiol* 1975;38:369–382. [PubMed: 165272]
- Fritz JB, Elhilali M, Shamma SA. Adaptive changes in cortical receptive fields induced by attention to complex sounds. *J. Neurophysiol* 2007;98:2337–2346. [PubMed: 17699691]
- Giraud K, Démonet JF, Habib M, Marquis P, Chauvel P, Liégeois-Chauvel C. Auditory evoked potential patterns to voiced and voiceless speech sounds in adult developmental dyslexics with persistent deficits. *Cerebral Cortex* 2005;15:1524–1534. [PubMed: 15689520]
- Goswami U. Why theories about developmental dyslexia require developmental designs. *Trends Cogn. Sci* 2003;7:534–540. [PubMed: 14643369]
- Habib M. The neurological basis of developmental dyslexia: An overview and working hypothesis. *Brain* 2000;123:2373–2399. [PubMed: 11099442]
- Hari R, Kiesilä P. Deficit of temporal processing in dyslexic adults. *Neurosci. Lett* 1996;205:138–140. [PubMed: 8907336]
- Hari R, Renvall H. Impaired processing of rapid stimulus sequences in dyslexia. *Trends Cogn. Sci* 2001;5:525–532. [PubMed: 11728910]
- Heath SM, Hogben JH, Clark CD. Auditory temporal processing in disabled readers with and without oral language delay. *J. Child Psychiatr* 1999;40:637–647.
- Hedrick MS, Younger MS. Perceptual weighing of stop consonant cues by normal and impaired listeners in reverberation versus noise. *J. Speech, Lang. Hear. Res* 2007;50:254–269. [PubMed: 17463228]
- Helenius P, Uutela K, Hari R. Auditory stream segregation in dyslexic adults. *Brain* 1999;122:907–913. [PubMed: 10355675]

- Howard MA, Volkov IO, Abbas PJ, Damasio H, Ollendieck MC, Granner MA. A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. *Brain Res* 1996;724:260–264. [PubMed: 8828578]
- Karmiloff-Smith A. Development itself is the key to understanding developmental disorders. *Trends Cogn. Sci* 1998;2:389–398.
- Kewley-Port D. Time-varying features as correlates of place of articulation in stop consonants. *J. Acoust. Soc. Am* 1983;73:322–335. [PubMed: 6826902]
- King WM, Lombardino LJ, Crandell CC, Leonard CM. Comorbid auditory processing disorder in developmental dyslexia. *Ear Hear* 2003;24:448–456. [PubMed: 14534414]
- Kraus N, McGee TJ, Carrell TD, Zecker SG, Nicol TG, Koch DB. Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science* 1996;273:971–973. [PubMed: 8688085]
- Kuhl PK. Theoretical contributions of tests on animals to the special-mechanisms debate in speech. *Exp. Biol* 1986;45:233–265. [PubMed: 3525222]
- Kuhl PK. Early language acquisition: Cracking the speech code. *Nature Reviews Neurosci* 2004;5:831–843.
- Kuhl PK, Padden DM. Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *J. Acoust. Soc. Am* 1983;73:1003–1010. [PubMed: 6221040]
- Kujala T, Karma K, Ceponiene R, Belitz S, Turkkila P, Tervaniemi M, Näätänen R. Plastic neural changes and reading improvement caused by audiovisual training in reading-impaired children. *Proc. Natl. Acad. Sci. USA* 2001;98:10509–10514. [PubMed: 11517333]
- Laasonen M, Tomma-Halme J, Lahti-Nuutila P, Service E, Virsu V. Rate of information segregation in developmentally dyslexic children. *Brain Lang* 75:66–81. [PubMed: 11023639]
- Lahiri A, Gewirth L, Blumstein SE. A reconsideration of acoustic information for place of articulation in diffuse stop consonants: Evidence from a cross-language study. *J. Acoust. Soc. Am* 1984;76:391–404. [PubMed: 6480990]
- Leff AP, Schofield TM, Stephan KE, Crinion JT, Friston KJ, Price CJ. The cortical dynamics of intelligible speech. *J. Neurosci* 2008;28:13209–13215. [PubMed: 19052212]
- Lieberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. *Psychol. Rev* 1967;74:431–461. [PubMed: 4170865]
- Lieberman AM, Mattingly IG. The motor theory of speech perception revised. *Cognition* 1985;21:1–36. [PubMed: 4075760]
- Lotto AJ, Kluender KR. General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perc. Psychophys* 1998;60:602–619.
- Lotto AJ, Kluender KR, Holt LL. Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *J. Acoust. Soc. Am* 1997;102:1134–1140. [PubMed: 9265760]
- Mayo C, Turk A. Adult-child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions. *J. Acoust. Soc. Am* 2004;115:3184–3194. [PubMed: 15237842]
- Mesgarani N, David SV, Fritz JB, Shamma SA. Phoneme representation and classification in primary auditory cortex. *J. Acoust. Soc. Am* 2008;123:899–909. [PubMed: 18247893]
- Merzenich MM, Brugge JF. Representation of the cochlear partition on the superior temporal plane of the macaque monkey. *Brain Res* 1973;50:275–296. [PubMed: 4196192]
- Merzenich MM, Jenkins WM, Johnston P, Schreiner C, Miller SL, Tallal P. Temporal processing deficits of language-impaired children ameliorated by training. *Science* 1996;271:77–81. [PubMed: 8539603]
- Moiescu-Yiflach T, Pratt H. Auditory event related potentials and source current density estimation in phonologic/auditory dyslexics. *Clin. Neurophysiol* 2005;116:2632–2647. [PubMed: 16221569]
- Morel A, Garraghty PE, Kaas JH. Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. *J. Comp. Neurol* 1993;335:437–459. [PubMed: 7693772]
- Müller-Preuss P, Mitzdorf U. Functional anatomy of the inferior colliculus and the auditory cortex: current source density analyses of click-evoked potentials. *Hear. Res* 1984;16:133–142. [PubMed: 6526745]

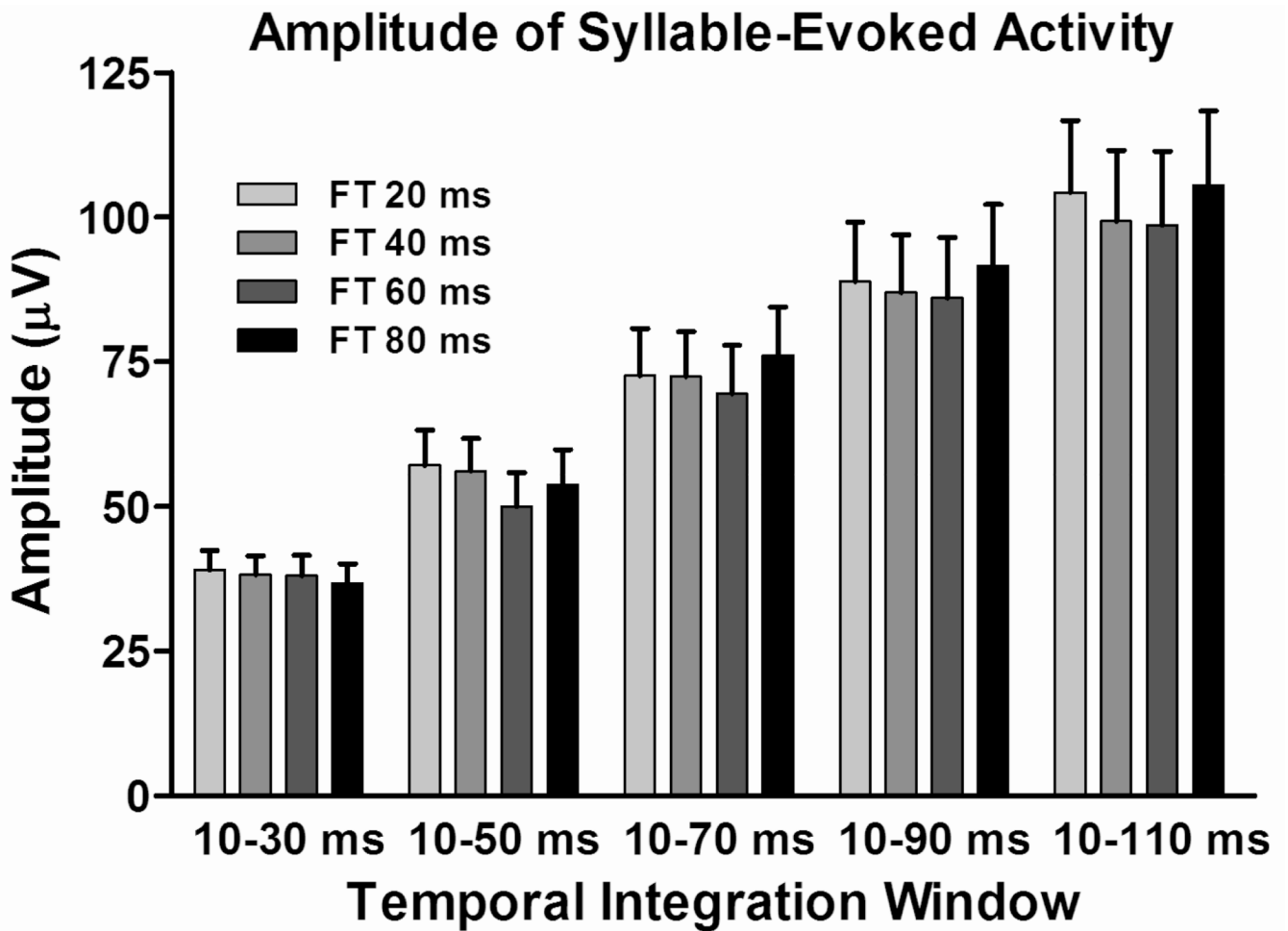
- Nagarajan S, Mahncke H, Salz T, Tallal P, Roberts T, Merzenich MM. Cortical auditory signal processing in poor readers. *Proc. Natl. Acad. Sci. USA* 1999;96:6483–6488. [PubMed: 10339614]
- Narayan R, Graña G, Sen K. Distinct time scales in cortical discrimination of natural sounds in songbirds. *J. Neurophysiol* 2005;96:252–258. [PubMed: 16571738]
- Nelken I, Prut Y, Vaadia E, Abeles M. Population responses to multifrequency sounds in the cat auditory cortex: One- and two-parameter families of sounds. *Hear. Res* 1994;72:206–222. [PubMed: 8150737]
- Obleser J, Zimmermann J, Van Meter J, Rauschecker JP. Multiple stages of auditory speech perception reflected in event-related fMRI. *Cerebral Cortex* 2007;17:2251–2257. [PubMed: 17150986]
- Ohde RN, Abou-Khalil R. Age differences for stop-consonant and vowel perception in adults. *J. Acoust. Soc. Am* 2001;110:2156–2166. [PubMed: 11681392]
- Ohde RN, Halay KL. Stop-consonant and vowel perception in 3- and 4-year-old children. *J. Acoust. Soc. Am* 1997;102:3711–3722. [PubMed: 9407663]
- Ohde RN, Halay KL, Vorperian HK, McMahan CW. A developmental study of the perception of onset spectra for stop consonants in different vowel environments. *J. Acoust. Soc. Am* 1995;97:3800–3812. [PubMed: 7790658]
- Ohde RN, Stevens KN. Effect of burst amplitude on the perception of stop consonant place of articulation. *J. Acoust. Soc. Am* 1983;74:706–714. [PubMed: 6630726]
- Petkov CI, O'Connor KN, Benmoshe G, Baynes K, Sutter ML. Auditory perceptual grouping and attention in dyslexia. *Cogn. Brain Res* 2005;24:343–354.
- Poeppel D, Idsardi WJ, van Wassenhove V. Speech perception at the interface of neurobiology and linguistics. *Phil. Trans. R. Soc. B* 2008;363:1071–1086. [PubMed: 17890189]
- Ramus F, Hauser MD, Miller C, Morris D, Mehler J. Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science* 2000;288:349–351. [PubMed: 10764650]
- Ramus F, Rosen S, Dakin SC, Day BL, Castellote JM, White S, Frith U. Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. *Brain* 2003;126:841–865. [PubMed: 12615643]
- Reed MA. Speech perception and the discrimination of brief auditory cues in reading disabled children. *J. Exp. Child Psychol* 1989;48:270–292. [PubMed: 2794857]
- Schnupp JWH, Hall TM, Kokelaar RF, Ahmed B. Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *J. Neurosci* 2006;26:4785–4795. [PubMed: 16672651]
- Scott BH, Malone BJ, Semple MN. Effect of behavioral context on representation of a spatial cue in core auditory cortex of awake macaques. *J. Neurosci* 2007;27:6489–6499. [PubMed: 17567810]
- Sinnott JM, Gilmore CS. Perception of place-of-articulation information in natural speech by monkeys versus humans. *Perc. Psychophys* 2004;66:1341–1350.
- Smits R, ten Bosch L, Collier R. Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants. I. Perception experiment. *J. Acoust. Soc. Am* 1996;100:3852–3864. [PubMed: 8969486]
- Stark E, Abeles M. Predicting movement from multiunit activity. *J. Neurosci* 2007;27:8387–8394. [PubMed: 17670985]
- Stark RE, Heinz JM. Perception of stop consonants in children with expressive and receptive-expressive language impairments. *J. Speech Hear. Res* 1996;39:676–686. [PubMed: 8844549]
- Steinschneider M, Fishman YI, Arezzo JC. Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey. *J. Acoust. Soc. Am* 2003;114:307–321. [PubMed: 12880043]
- Steinschneider M, Fishman YI, Arezzo JC. Spectrotemporal analysis of evoked and induced electroencephalographic responses in primary auditory cortex (A1) of the awake monkey. *Cerebral Cortex* 2008;18:610–625. [PubMed: 17586604]
- 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;674:147–152. [PubMed: 7773684]



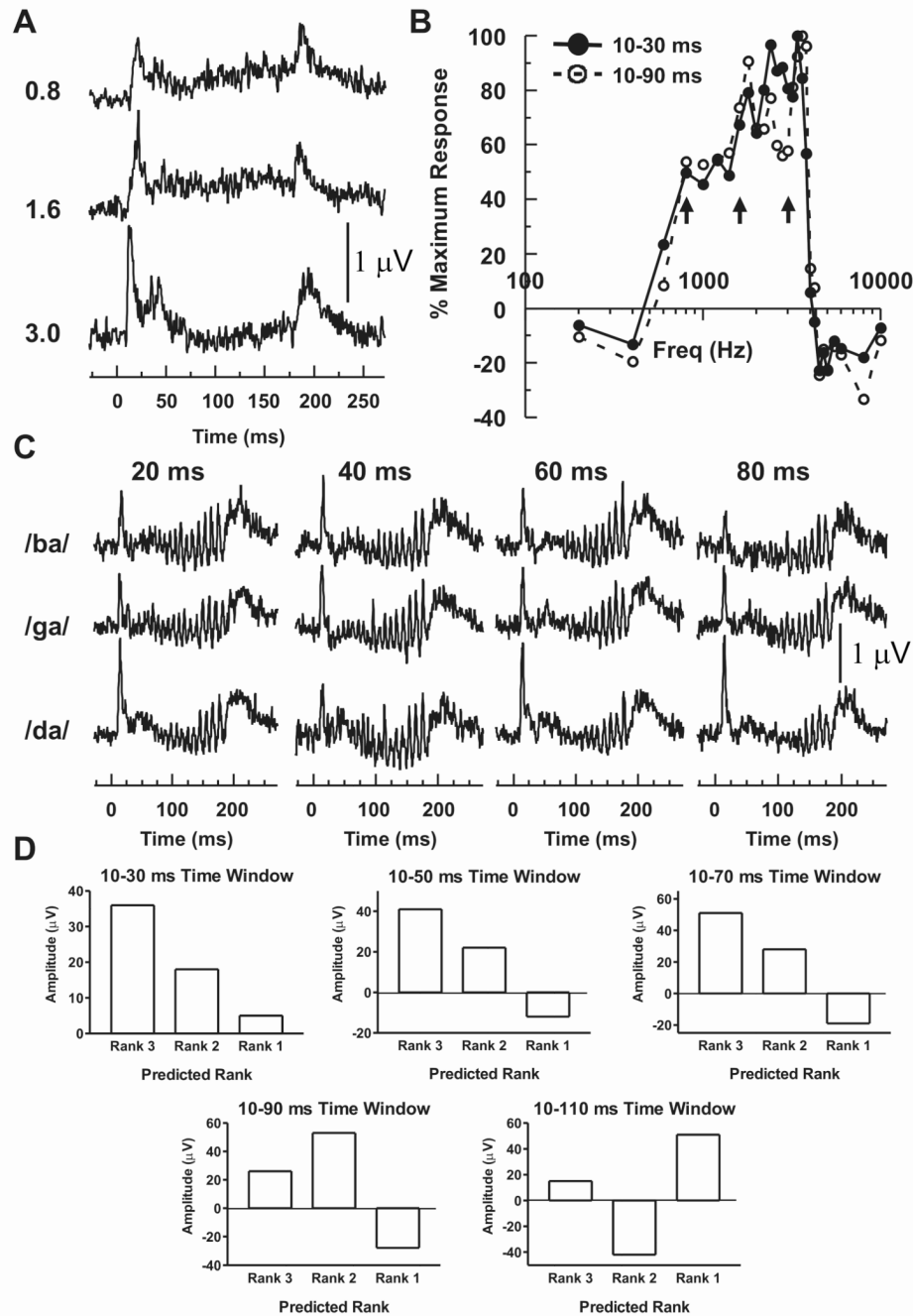
- Steinschneider M, Tenke C, Schroeder C, Javitt D, Simpson GV, Arezzo JC, Vaughan HG Jr. Cellular generators of the cortical auditory evoked potential initial component. *Electroenceph. Clin. Neurophysiol* 1992;84:196–200. [PubMed: 1372236]
- Stevens C, Fanning J, Coch D, Sanders L, Neville H. Neural mechanisms of selective auditory attention are enhanced by computerized training: Electrophysiological evidence from language-impaired and typically developing children. *Brain Res* 2008;1205:55–69. [PubMed: 18353284]
- Stevens C, Sanders L, Neville H. Neurophysiological evidence for selective auditory attention deficits in children with specific language impairment. *Brain Res* 2006;1111:143–152. [PubMed: 16904658]
- Stevens KN. Acoustic correlates of some phonetic categories. *J. Acoust. Soc. Am* 1980;68:836–842. [PubMed: 7419819]
- Stevens KN, Blumstein SE. Invariant cues for place of articulation in stop consonants. *J. Acoust. Soc. Am* 64:1358–1368. [PubMed: 744836]
- Studdert-Kennedy M, Mody M. Auditory temporal perception deficits in the reading-impaired: A critical review of the evidence. *Psychonomic. Bull. Rev* 1995;2:508–514.
- Supèr, H.; Roelfsema, PR. Chronic multiunit recordings in behaving animals: Advantages and limitations. In: vanPelt, J.; Kamermans, M.; Levelt, CN.; van Ooyen, A.; Ramakers, GJA.; Roelfsema, PR., editors. *Progress in Brain Research*. Vol. Vol. 147. Amsterdam: Elsevier Science; 2005. p. 263-282.
- Tallal P. Improving language and literacy is a matter of time. *Nature Reviews Neurosci* 2004;5:721–728.
- Tallal P, Merzenich MM, Miller S, Jenkins W. Language learning impairments: integrating basic science, technology, and remediation. *Exp. Brain Res* 1998;123:210–219. [PubMed: 9835411]
- Tallal P, Miller S, Bedi G, Byma G, Wang X, Nagarajan SS, Schreiner C, Jenkins WM, Merzenich MM. Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science* 1996;271:81–84. [PubMed: 8539604]
- Tallal P, Miller S, Fitch RH. Neurobiological basis of speech: A case for the preeminence of temporal processing. *Ann. N.Y. Acad. Sci* 1993;682:27–47. [PubMed: 7686725]
- Van Wieringen A, Pols LCW. Discrimination of single and complex consonant-vowel and vowel-consonant-like formant transitions. *J. Acoust. Soc. Am* 1995;98:1304–1312.
- Walley AC, Carrell TD. Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *J. Acoust. Soc. Am* 1983;73:1011–1022. [PubMed: 6841809]
- Wang X, Merzenich MM, Beitel R, Schreiner CE. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: Temporal and spectral characteristics. *J. Neurophysiol* 1995;74:2685–2706. [PubMed: 8747224]
- Wright BA, Lombardino LJ, King WM, Puranik CS, Leonard CM, Merzenich MM. Deficits in auditory temporal and spectral resolution in language-impaired children. *Nature* 1997;387:176–178. [PubMed: 9144287]



**Figure 1.** Sound spectrographs and waveforms of the synthetic stop consonant-vowel syllables used in the present study. Syllables with formant transitions of 20 (top) and 80 ms (bottom) are depicted. Additional syllables with formant transitions of 40 and 60 ms were also presented. See text for details.



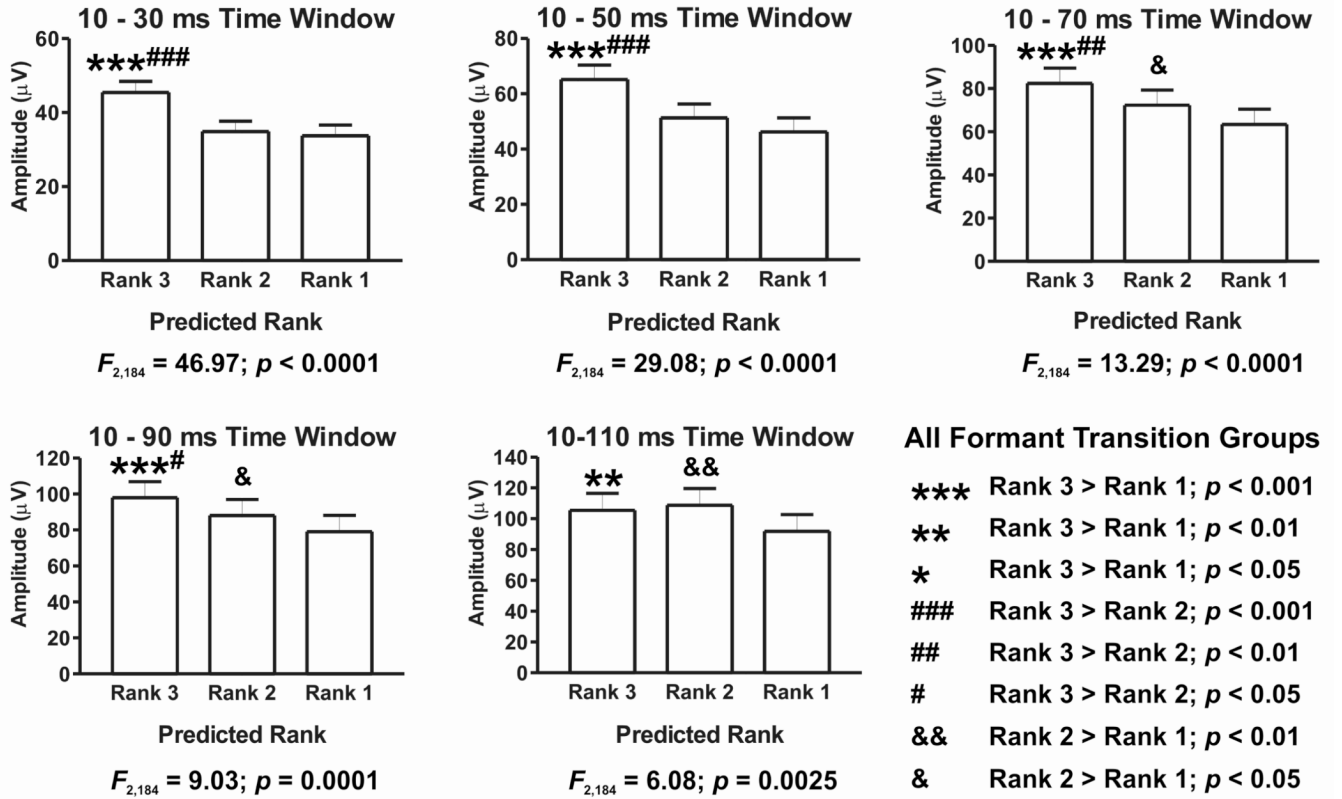
**Figure 2.** Average amplitudes of MUA evoked by the syllables /ba/, /ga/, and /da/ under the four formant transition conditions and within the five temporal integration windows examined. There is no systematic effect of formant transition duration on the amplitude of cortical MUA. As for all following figures, error bars denote the SEM.



**Figure 3.**

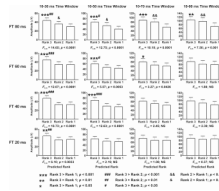
MUA recorded from a site whose BF is 3400 Hz exemplifies how activity evoked by stop CV syllables can be partially predicted by the tonotopic specificity of A1. A) MUA evoked by 0.8, 1.6, and 3.0 kHz tones, which were the center frequencies of the frication bursts occurring at the onsets of the syllables /ba/, /ga/, and /da/, respectively. B) Frequency response functions based on MUA summed within two of the five integration windows (10–30 and 10–90 ms) used in the present study. Arrows mark points along the frequency response functions at frequencies equal to the 0.8, 1.6, and 3.0 kHz tones, respectively. C) MUA evoked by the syllables under the four formant transition duration conditions. D)

Amplitudes of MUA evoked by syllables with 80 ms formant transitions ranked according to the amplitudes of responses evoked by the 0.8, 1.6, and 3.0 kHz tones. See text for details.



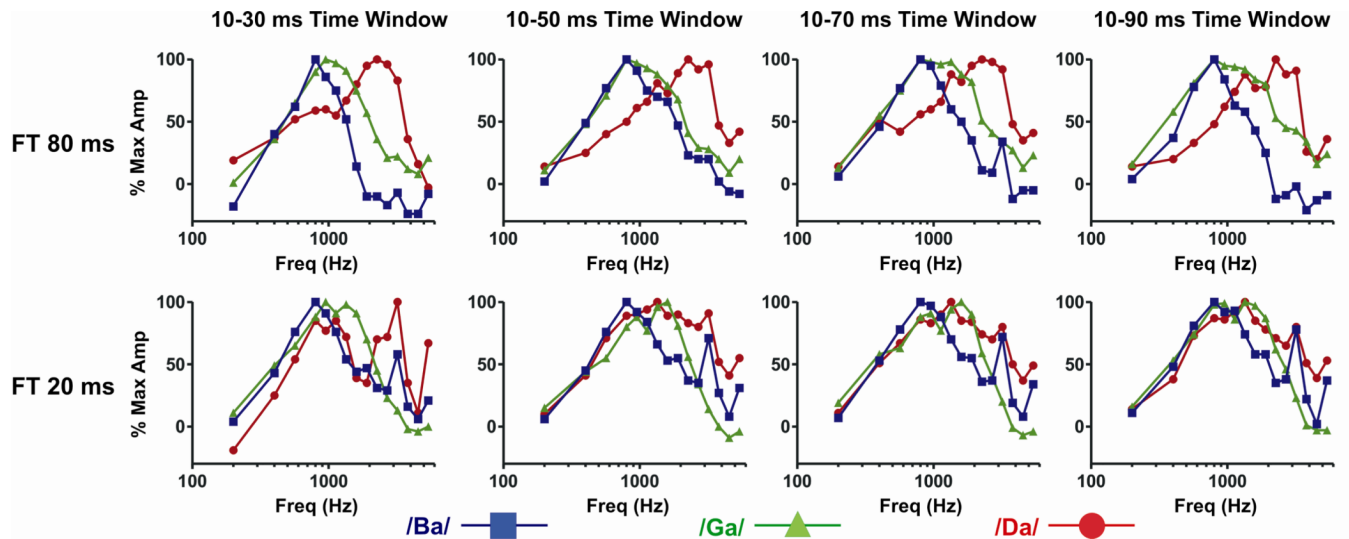
**Figure 4.**

Average amplitude of MUA evoked by syllables collapsed across all formant transition conditions ranked according to the relative activity evoked by 0.8, 1.6, and 3.0 kHz tones. Data for all five temporal integration windows are shown. Results of one-way ANOVA tests are shown below each graph, and *post-hoc* test results are indicated above the bars in the graphs (legend at bottom right). See text for details.



**Figure 5.**

Average amplitude of MUA evoked by syllables under the different formant transition duration conditions ranked according to the relative activity evoked by 0.8, 1.6, and 3.0 kHz tones. Responses in the first four of five temporal windows of integration examined are shown. Results of one-way ANOVA tests are included below each graph, and *post-hoc* test results are indicated above the bars in the graphs (legend at bottom). See text for details.

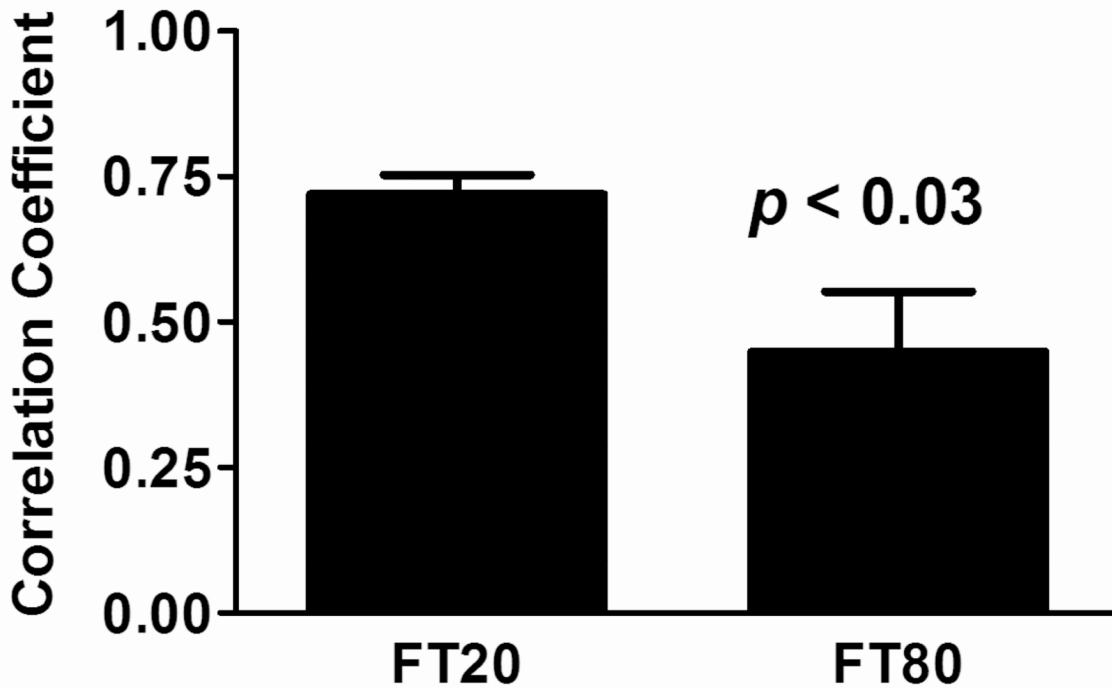


**Figure 6.**

Average spectral tuning curves of MUA evoked by tone responses at sites where /ba/, /ga/, or /da/ elicited the largest speech-evoked response within a given window of temporal integration (indicated at the top of each column of plots). Tuning curves for responses to the 80 ms and 20 ms formant transition syllables are shown in the top and bottom plots, respectively. Prior to averaging across sites, tone responses were binned in quarter octave steps above 800 Hz, half octave steps above 400 Hz, and a one octave step between 200 and 400 Hz. See text for details.



## Similarity of Tuning Curves for /ba/, /ga/, and /da/



**Figure 7.**

Average Pearson correlation coefficients based on the correlations among the three mean spectral tuning curves shown in Figure 6 ( $\pm$  SEM). The  $p$  value for the result of a Welch corrected  $t$  test for unequal variance comparing the two average correlation coefficients are shown in the figure.