NORTHWESTERN UNIVERSITY

Temporal Features of Speech in the Auditory System: Normal and Dyslexic Children and an Animal Model

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Communication Sciences and Disorders

By

Daniel Arthur Abrams

EVANSTON, ILLINOIS

June 2008

ABSTRACT

**Temporal Features of Speech in the Auditory System:
Normal and Dyslexic Children and an Animal Model**

Daniel Arthur Abrams

An amazing characteristic of the speech signal is that it contains a variety of temporal features that occur simultaneously in the signal, and each of these features provides unique and essential information for speech perception. An equally astonishing fact is that, in most cases, the human auditory system is able to efficiently extract these temporal acoustic features from the speech signal as a precursor to higher-order linguistic, cognitive and mnemonic processes associated with speech reception. A clinical population that has shown abnormal processing of rapid temporal features in speech is reading-impaired individuals (RI), and it has been proposed that auditory-temporal impairments preclude normal development of phonological systems necessary for reading acquisition. The primary goals of this work are to describe central mechanisms responsible for encoding temporal features in speech in the unimpaired human auditory system, and to examine the extent to which these mechanisms may be impaired in RI. We have pursued these goals by investigating both brainstem and cortical representations of speech-sound stimuli in unimpaired (control) and RI children using auditory evoked-potentials. Results are the first to show right-hemisphere cortical asymmetry in the representation of the speech envelope, the slow temporal cue that provides syllable pattern information in speech. This result supports the hypothesis that a neural mechanism for temporal encoding in the human auditory system is the asymmetrical routing of this acoustic information between the cerebral hemispheres. We also

provide the first neurophysiological evidence that reading-impaired (RI) individuals have impaired speech envelope representation, a finding that challenges an influential hypothesis stipulating that temporal impairments are specific to rapid features of speech. Additionally, we describe functional relationships in the ascending human auditory system: we show that temporal acuity in the human auditory brainstem predicts cerebral asymmetry for rapid acoustic processing but is not related to the slow temporal features of the speech envelope. We also examined near-field auditory responses in an animal model to explore mechanisms for auditory temporal processing in more localized neuronal populations than those afforded by the far-field potentials measured in humans. Results show that a non-primary auditory pathway may be specifically tuned to encode the slow temporal features in acoustic signals, and suggest that non-primary pathways may be important for processing the speech envelope in humans. Results from the animal model also show how ensembles of auditory cortical neurons can simultaneously represent the fundamental frequency and speech envelope in speech signal. Taken together, we have made new discoveries of how the unimpaired human auditory system encodes perceptually-important temporal features inherent to the speech signal, and the abnormal function of these mechanisms in RI. We have also proposed a role for non-primary auditory pathways in the coding of slow temporal information. These results are considered with respect to existing hypotheses addressing temporal information processing in the central nervous system, hierarchical models of speech perception and theories of dyslexia.

ACKNOWLEDGEMENTS

These studies were supported by grants from National Institutes of Health grant R01 DC01510-10 and National Organization for Hearing Research grant 340-B208. The author wishes to thank Nina Kraus, Ph.D. for her endless encouragement, enthusiasm and insight with respect to this work, the author's education, and the author's general development. Most importantly, I would like to acknowledge Dr. Kraus's patience with me as I suffered through multiple scientific mishaps, developed new methods, analysis tools and the confidence necessary to complete this work and achieve my potential as a doctoral candidate. I would like to thank committee member Steve Zecker, Ph.D., for statistical expertise, access to his vast insight on the learning-disabled population, and, most importantly, his friendship and kindness. I would like to thank committee member Mitra Hartmann, Ph.D., for providing insight on the animal components of this work and general encouragement towards my scientific development and contribution. I would also like to thank Trent Nicol for his technical help, theoretical insight, sense of humor and general assistance with respect to nearly every facet of this work. I would also like to thank Drs. Catherine Warrier and Karen Banai for their general encouragement and kindness as well as an uncountable number of lively scientific discussions. This work was enormously enhanced by the constant scrutiny, curiosity and encouragement provided by fellow members of Dr. Kraus's Auditory Neuroscience Lab: Erika Skoe, Erin Hayes, Ph.D., Gabriella Musacchia, Ph.D., Nicole Russo, Ph.D., Krista Johnson, Judy Song, Jane Hornickel, and Bradley Wible, Ph.D. I would also like to acknowledge the staff and faculty of the Department of Communication Sciences, who not only provided an excellent support network for doctoral students, but who inspired me to pursue important contributions in the field of audition.

I offer my most heartfelt thanks to my support system at home, spearheaded by my amazing wife, Heidi. Heidi made every sacrifice imaginable to enable this work to be realized, from moving across the country to be with me to watching our amazing kids, Jake and Max, on the weekends to enable me to make progress on my project. By far, her most important contribution was her constant, daily encouragement and confidence in me as a scientist and as a person in general. On the darkest days of this project, she picked me up off the ground, brushed me off, gave me a hug and told me not to give up, and for that alone I will be forever grateful. I think science would be well-served if every scientist had a Heidi at home, and I am endlessly lucky to have her by my side.

I would like to thank our wonderful little guys, Jake and Max, currently 2.5 years-old and 3 months-old, respectively, and our loving dog Naya (aka, Hoofy), for helping provide balance in my life as well as a hilarious and exhausting escape from my work. I would like to thank my parents, Susan and Alan, and all my amazing grandparents, for their encouragement with respect to my work and development. I'd like to pass on a huge thank you to my brother, Mike, for a million consults with respect to the pursuit of doctoral degree, and additional thanks to my Mom for long chats about academia. I would like to thank my amazing friends in the band Hayward -- BB, Loz, Jen Medina, Mike Simons and Kris Penn -- for all the laughs and fun music, putting up with my subpar guitar solos, and letting me sing "Vincent Black Lightning" (VBL!!) to my heart's content. I would also like to thank all of my friends for inspiring me, making me laugh and making me feel good in general – I am blessed by all of you. I'd like to make a special shout-out to: Mike Fleming, aka, Uncle Ming-a-Ding-a-Ding-Dong; Brian ("BB", "Bload"), Heiji and Cyrus Black; Chris ("Loz", "Ched"), Monisha and Anna-Banana Lozier; the Joey

Carosella-Alyssa Kleinsmith crew, including Ezzie!!!! and Mimi!!!!; and the Benjamin Fingerhut-Winnie Gundeck duo, including Guffman and Winnie the Cat. We are so lucky to have found you all; you are the best friends we could ever ask for.

Having children while pursuing a doctoral degree is not for sissies. When this occurs, there is a need for LOTS of help with the children so that doctoral work can actually get done, and now I'd like to thank all of the people who have helped me juggle my way through this process. First, a million thanks again to my wonderful wife, Heidi, for all of the weekends, weekdays, mornings, afternoons and nights when you let me slip out the door to get more work done. The sacrifices involved here were endless, and I thank you for all of them. Second, a million thanks to Veronica Crews ("Vava"), Laran Kaplan ("Lala") and Jordan Corson ("Judu") for helping on a zillion occasions with the kids – we are all so lucky to have found you, and we couldn't have asked for better child care. Finally, I'd like to extend huge thanks to my mom ("Meme") and Heidi's mom, Pat ("MumMum"), for rushing to Chicago to help with the kids on dozens of occasions; Jake and Max won the grandma lottery, and we are all so lucky to have you in our lives.

Finally, I would like to thank the friendly baristas at various coffeehouses in Chicago, especially the cozy Starbucks on North Ave. and Sip on Grand Ave., for caffeinating me and providing a great place for me to write at all hours of the day and night.

PREFACE

The following six studies were designed to investigate temporal processing mechanisms in the mammalian auditory system. Chapter II shows that slow temporal features in speech are lateralized to the right-hemisphere auditory cortex in unimpaired children. Chapter III tests the hypothesis that reading impairments are specifically associated with rapid temporal processing deficits in acoustic signals by investigating whether slow temporal processing is also impaired in the RI auditory system. Chapter IV describes functional connectivity between the human auditory brainstem and cortex for the processing of rapid temporal features in speech. Chapter V shows that there is no apparent connectivity linking temporal acuity in the auditory brainstem and processing of slow temporal information in auditory cortex. Chapter VI investigates a neural mechanism for processing slow temporal information in localized neuronal populations in an animal model. Chapter VII shows that a general property of the mammalian auditory system is the ability to encode multiple temporal aspects of complex acoustic signals, like speech. A summary of the findings across studies is provided in the Discussion (Chapter VIII).

DEDICATION

I dedicate this dissertation to my amazing wife, Heidi. Thank you for blessing me with a more wonderful life than I ever imagined – I love you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

An amazing feature of the speech signal is that it contains multiple acoustical features that can facilitate speech understanding. Listeners constantly rely on the abundance of perceptually-salient acoustic cues present in the speech signal. For example, speech heard through a telephone is dramatically filtered by the telephone company, and has consequently been stripped of a prominent acoustical feature known as the fundamental frequency. Nevertheless, speech comprehension is excellent using telephones, a fact that can be explained by the availability of other acoustical cues in the signal that are sufficient for perception.

There has been great interest in the last 15 years in describing the role of temporal features in speech for speech perception. A framework for considering temporal features in speech was provided by Rosen (Rosen, 1992). This framework distilled the temporal features in speech to three categories, with each category providing discrete and essential acoustic information necessary for normal speech perception. A marvel of the human central auditory system is that, in most cases, it is able to simultaneously extract these various temporal features from the on-going speech signal. Nevertheless, temporal features in speech and non-speech acoustic signals are not always sufficiently resolved in the human auditory system, and a clinical population that has consistently demonstrated perceptual deficits associated with acoustic-temporal processing is reading-impaired individuals (RI; Tallal and Piercy, 1973, 1974, 1975; Tallal, 1980; Kraus et al., 1996; Ramus et al., 2003). An influential hypothesis states that temporal processing deficits in RI are specific to rapid temporal features of acoustic signals, and abnormal perception of these signals contributes to reading disorders (Tallal et al., 1998).

Motivation and Hypotheses

The temporal features in speech are extremely important for perception, and consequently there is great interest in understanding how these features are encoded in the human auditory system. Currently, there is an incomplete picture of how the unimpaired human auditory system encodes these temporal features, and a primary goal of this work is to provide a more complete description of central mechanisms associated with temporal processing. Furthermore, despite decades of research, the biological foundation of RI is still unclear, and a second goal of this work is to better understand the acoustic-temporal impairments in RI and their relations to perception and measures of academic achievement. Finally, there is great interest in understanding how highly-localized populations of auditory neurons encode temporal features in speech and non-speech signals, a question that can only be addressed in animal models of the auditory system. The third major goal of this work is describe temporal mechanisms in localized neuronal populations in an animal model.

To address these questions, six experiments were planned. This comprehensive study investigates (1) the cortical representation of slow temporal features in speech, known as the speech envelope, in the unimpaired auditory system; (2) the cortical representation of the speech envelope in RI; (3) the relationship between the representation of temporal features in speech in the auditory brainstem and cortical asymmetry for rapid temporal features; (4) the relationship between the representation of temporal features in speech in the auditory brainstem and cortical asymmetry for slow temporal features; (5) the representation of slow, non-speech temporal

features in the primary and non-primary auditory thalamus and cortex of guinea pig; (6) the simultaneous representation of the speech envelope and periodicity in cortex of guinea pig. The central hypothesis for the human work is that temporal processing of discrete elements of speech is essential for normal perception of speech and impacts phonological systems necessary for normal reading acquisition. With respect to the animal work, the hypothesis is that a fundamental property of the mammalian auditory system is that discrete neuronal populations represent specific temporal features in acoustic signals. Taken together, these studies enable a more comprehensive description of temporal processing mechanisms in the central auditory system in both unimpaired and RI children as well as an animal model.

Temporal features in speech

Historically, the acoustical description of the speech signal was considered with respect to the spectral characteristics of the signal. However, a framework for considering the temporal features in speech was proposed over 15 years ago (Rosen, 1992). This framework divided the temporal features in speech into three categories, and an essential consideration with regards to this framework is that the temporal information represented by these categorizes represents discrete and essential information associated with speech perception. One category in this temporal framework is the amplitude envelope, or "speech envelope," which is represented in the speech signal for rates between 2-50 Hz. The dominant feature of the speech envelope is that of the syllable rate of speech (3-4 Hz), and syllable patterns are essential for normal speech perception (Drullman et al., 1994a, b; Shannon et al., 1995). A second temporal category in this framework is that of periodicity cues, defined as rates between 50-500 Hz in the speech signal. Periodicity cues in speech include the representation of the fundamental frequency (F0) of

speech, which provides the "pitch" of a speaker's voice, conveys prosodic information and, in the case of tonal languages, semantic information. The third category of temporal information in speech as described in Rosen's framework is that of the temporal fine-structure, which occurs at frequencies between 600-10,000 Hz. The temporal fine-structure of speech provides information regarding the spectrum and formant structure of speech sounds, as well as dynamic frequency transitions that occur to the formant structure.

While Rosen's framework has received considerable attention in the literature, complementary categories of temporal features in speech were proposed in a more recent work (Poeppel, 2003). These categories consist of two time scales that are also highly relevant to speech perception. The first time scale refers to the syllable rate of speech, which is between ~3-6 Hz. While these particular rates are also encompassed in Rosen's definition of the speech envelope, the narrow range of frequencies proposed in the latter framework represents the most important frequencies in the speech envelope for perception (Drullman et al., 1994a). The second time-scale in this second framework corresponds to important temporal modulations in the speech signal which occur between ~25-50 Hz. This time-scale would again be encompassed by the definition of the speech envelope in Rosen's categorization of the temporal properties of speech. However, Poeppel highlights the 25-50 Hz range because it corresponds to the range of temporal information in speech relevant for encoding formant transitions in stop consonants.

These categories for considering temporal information in speech, including contributions from both Rosen and Poeppel, will be addressed at various times in the current work.

Temporal processing in the auditory system

Given the importance of temporal features in speech for perception, there is great interest in understanding how the human auditory system is able to encode these temporal features. While the ultimate goal of our work is to understand how the human auditory system encodes temporal features in acoustic stimuli, much of what we know regarding the human auditory system has been provided by studies investigating animal models of the auditory system. Therefore, before describing properties of the human auditory system, there will be a cursory review of temporal processing properties in the ascending mammalian auditory system.

Temporal processing in animal models of the auditory system: primary pathway representations

The central auditory system consists of a highly complex network of sub-cortical and cortical nuclei characterized by an intricate pattern of connectivity between nuclei (Kaas and Hackett, 2000). Given the great complexity of this system, the question of "how the auditory system encodes temporal features in speech" is a complicated one, and the answer to this question varies considerably based on what region of the auditory system one is addressing. At the level of the auditory nerve, the most peripheral station in the central auditory system, it has been shown that auditory nerve fibers reliably encode temporal features of complex (Sachs and Young, 1979; Sachs et al., 1983; Sachs, 1984; Delgutte and Cariani, 1998) acoustic stimuli according to their temporal discharge patterns. An important aspect of temporal representation at the auditory nerve is that auditory nerve fibers can "phase-lock" to temporal features in acoustic stimuli up to ~4000 Hz. With respect to the speech signal, this means that the auditory nerve can phase-lock to nearly all of the meaningful temporal components in the signal (Rosen, 1992). The primary auditory midbrain nucleus, the inferior colliculus (IC), is an anatomically important nucleus

based on the huge convergence of auditory connections that terminate there (Kaas and Hackett, 2000), and consequently it is studied frequently with respect to its physiologic properties. Responses to acoustic stimuli have shown that single neurons in the IC can phase-lock to acoustic temporal features up to 1000 Hz (Langner and Schreiner, 1988), a considerably lower maximum frequency relative to the auditory nerve.

At the level of primary auditory cortex, it has been shown that single units show robust time-locked responses to very slow rates and cannot phase-lock to temporal features faster than ~ 40 Hz (Wang et al., 2003), a decrease in maximum following-rate two orders of magnitude slower than the auditory nerve. These time locked response to slower rates have been called an "explicit" temporal code, and are complemented by an "implicit" rate code that is thought to represent faster temporal modulations in primary auditory cortex (Lu et al., 2001). Germane to the current work, the current model for temporal processing at the level of auditory cortex does not include details on how slow rates may be differentially represented in the auditory system, a processing stage that would presumably be essential for the discrimination of these rates. Furthermore, results from studies investigating ensemble cortical representation of periodic aspects of the speech signal have shown that neuronal ensembles phase-lock to periodicities up to ~200 Hz in primary auditory cortical neurons (Steinschneider et al., 1980, 1982; Steinschneider et al., 1990; Steinschneider et al., 1994; Steinschneider et al., 1995; Steinschneider et al., 2003). Nevertheless, the essential trend in the ascending auditory system with respect to temporal information processing is that higher levels in the system are more limited in the range of frequencies they can represent with phase-locked responses relative to more peripheral stations in the system.

Temporal processing in animal models of the auditory system: non-primary pathway representations

As stated previously, the auditory system consists of a dizzying number of nuclei. The majority of research conducted in the auditory system describes properties of lemniscal, or "primary," nuclei in the ascending system. There are many other auditory-responsive nuclei that have received considerably less attention in the literature, possibly because of their less-robust response patterns (He, 2001, 2002, 2003b) and more complex acoustic preferences (Rauschecker et al., 1995). Non-primary auditory nuclei are so poorly understood that a unified hypothesis regarding the functional significance of these neurons is not evident in the literature.

A tradition of sensory neuroscience is to examine whether similar neural mechanisms exist for analogous function between sensory modalities. An intriguing hypothesis has been described in great detail regarding the role of the paralemniscal (non-primary) pathway in the somatosensory system. Results from the rat trigeminal system demonstrate that slow rates (between 2-8 Hz) are differentially coded by primary and non-primary pathways (Ahissar et al., 2000). Specifically, primary neurons in both thalamus and cortex code stimulation rate with constant latencies while non-primary neurons code stimulation frequency as systematic changes in latency. Based on its unique sensitivity for slow rates, it is suggested that the paralemniscal pathway is "optimally tuned for temporal processing of vibrissal information around the whisking frequency range (8 Hz)." An implication of the somatosensory findings is that a paralemniscal (non-primary) pathway in the auditory system may be optimally tuned to code slow rates present in acoustic signals and could serve as a neural mechanism for speech envelope coding. Furthermore, the

current models of temporal processing in auditory cortex (Wang et al., 2003) do not include details on how slow rates may be differentially represented in the auditory system, a processing stage that would presumably be essential for the discrimination of these rates. There have been no systematic investigations of non-primary representation of acoustic rate and whether it differs from the primary representation. A goal of the current work was to examine this question in an animal model of the auditory system.

Temporal processing of speech in the human auditory system

It is generally believed that temporal attributes in the ascending auditory system in animal models provides a reasonable approximation of the human auditory system. This is supported by results from the human auditory brainstem response (ABR). The ABR reflects neurophysiologic activity from synchronous neuronal ensembles in rostral and posterior brainstem structures. The ABR has emerged as an experimental tool to assess the integrity of brainstem processing of speech and other complex stimuli in normal and impaired populations (Kraus and Nicol, 2005). Speech-evoked ABRs represent temporal features of speech (Johnson et al., 2005) with great fidelity and it has been shown that brainstem responses represent temporal acoustic features up to ~1000 Hz (Johnson et al., 2005), commensurate with findings from animal models (Langner and Schreiner, 1988).

Nevertheless, animal models provide limited information regarding processing of the speech signal in the human auditory system: it has been shown that auditory systems become highly specialized to cater to the specific acoustic signals necessary for survival (Sakai and Suga, 2001). Consequently, given the acoustic complexity of the speech signal as well as the biological

importance of this signal to humans, a reasonable assumption is that specialized neuronal mechanisms may have evolved to encode the important acoustic features present in the speech signal. At the level of auditory cortex, it was recently proposed that an important neural mechanism for representing the temporal information in speech in the human auditory system is based on the component rates inherent to the speech signal (Poeppel, 2003). Specifically, a recent hypothesis, called the "asymmetric sampling in time" (AST) hypothesis, proposed that rapid temporal information in speech (20-40 Hz) is lateralized to left-hemisphere auditory cortex and slower temporal features (3-6 Hz) are lateralized to right-hemisphere auditory cortex. Importantly, the particular rates specified in this hypothesis are the rates described by Poeppel. This hypothesis is supported by findings which show that slow, non-speech acoustic stimuli (3-5 Hz) are lateralized to right-hemisphere auditory areas (Boemio et al., 2005) while rapid acoustic stimuli (20-50 Hz) are lateralized to left-hemisphere auditory areas (Zatorre et al., 2002; Zaehle et al., 2004; Schonwiesner et al., 2005). An important question addressed by the current work is that it is not known to what extent this putative mechanism applies to the slow temporal features in speech.

Reading-impairments: phonological and auditory disorders

Reading impairment (RI), a disorder that affects ~5% of the general population, is defined as a specific deficit in reading that is independent of intelligence, motivation, educational opportunity and overt neurological damage. There is general consensus that many RIs suffer from a phonological deficit, defined as an impairment in the representation and processing of speech sounds (Ramus, 2003). This particular deficit is manifested when individuals with RI perform

tasks involving the manipulation of phonemes (i.e., reversing the phonemes in words) and rhyming tasks (i.e., identifying whether words end with the same sounds).

In addition to phonological deficits, a large body of evidence has accumulated that shows that many RIs are also impaired in the perception of rapid acoustic events in speech and non-speech signals (Tallal and Piercy, 1973; Tallal, 1980; Kraus et al., 1996). An influential hypothesis (Tallal et al., 1998) poses that abnormal perception of rapid acoustic events present in speech (20-40 Hz) precludes normal development of phonological systems since many phonological contrasts rely on resolving acoustic events occurring on this time scale (Phillips and Farmer, 1990).

A more recent hypothesis states that abnormal perception of slow temporal features in speech and non-speech signals additionally contribute to reading impairments (Goswami, 2002). As discussed previously, these slow acoustic features in speech, known as the speech envelope, provide syllable pattern information and segmental cues for phoneme identity and are thought to be extremely important for normal speech perception (Drullman et al., 1994a). In support of this hypothesis, it has been shown that deficits in the perception of slow temporal cues in non-speech acoustic signals accounts for significant variance in reading scores (Goswami et al., 2002; Witton et al., 2002). An important question addressed by the current work is that it is not known whether a slow temporal impairment in RI impacts cortical processing of the speech signal.

Functional relationships between auditory brainstem and cortex

As discussed previously, the central auditory system consists of a highly complex network of sub-cortical and cortical nuclei (Kaas and Hackett, 2000). An important step in understanding the dynamics of this system is to describe functional relationships between different components of this system. Addressing functional relationships between different areas of the system presents an exciting challenge to researchers since neural representations of acoustic signals change drastically in the ascending auditory system. Therefore, comparing between different stations in the auditory system generally requires the comparison of very different response forms. Previous work has shown that temporal acuity in the auditory brainstem is related to different aspects of cortical function. First, it was shown that brainstem onset responses are related to the robustness of cortical responses in the presence of background noise (Wible et al., 2005). Second, it was shown that brainstem timing was related to cortical representations that reflect fine-grained acoustic change (Banai et al., 2005a). A question that remains is whether there is a relationship between auditory brainstem timing and asymmetries for acoustic processing auditory cortex as described in the AST hypothesis (Poeppel, 2003). One of the goals of the current work was to address the extent to which temporal acuity in the auditory brainstem is related to both rapid auditory processing in cortex (20-40 Hz) and slow cortical processing of speech (3-6 Hz).

Scientific Contribution

At its most general level, this work contributes to understanding of temporal information processing in sensory systems. More specifically, results describe new mechanisms for encoding temporal information in the auditory system, including the first evidence that slow temporal information inherent to the speech signal is preferentially processed in right-hemisphere auditory

cortex. We have also shown that this mechanism shows abnormal function in reading-impaired children. We have also shown patterns of functional connectivity between the auditory brainstem and cortex with respect to important temporal processing. Finally, we have demonstrated that the non-primary pathway may be specifically tuned to code low-frequency temporal information. Each of these accomplishments represents a substantial contribution to the fields of neuroscience, communication disorders, and linguistics, and it is hoped that scientists are able to use the novel information provided by this work to further their own research. Perhaps the most important contribution of this work is that it may enable a better understanding of the reading-impaired brain and serve as a catalyst to develop strategies to remediate deficits in reading impaired individuals.

CHAPTER II: RIGHT-HEMISPHERE AUDITORY CORTEX IS DOMINANT FOR CODING SYLLABLE PATTERNS IN SPEECH

Published in *Journal of Neuroscience,* 28(15): 3958-3965

Daniel A. Abrams[1], Trent Nicol[1], Steven Zecker[1], and Nina Kraus[1,2]

[1]Auditory Neuroscience Laboratory, Department of Communication Sciences, and [2]Departments of Neurobiology and Physiology and Otolaryngology. Northwestern University, 2240 Campus Drive, Evanston, Illinois, 60208.

Abstract

Cortical analysis of speech has long been considered the domain of left-hemisphere auditory areas. A recent hypothesis poses that cortical processing of acoustic signals, including speech, is mediated bilaterally based on the component rates inherent to the speech signal. In support of this hypothesis, previous studies have shown that slow temporal features (3-5 Hz) in non-speech acoustic signals lateralize to right-hemisphere auditory areas while rapid temporal features (20-50 Hz) lateralize to the left hemisphere. These results were obtained using non-speech stimuli, and it is not known if right-hemisphere auditory cortex is dominant for coding the slow temporal features in speech known as the speech envelope. Here we show strong right-hemisphere dominance for coding the speech envelope, which represents syllable patterns and is critical for normal speech perception. Right-hemisphere auditory cortex was 100% more accurate in

following contours of the speech envelope and had 33% larger response magnitude while following the envelope compared to the left-hemisphere. Asymmetries were evident irrespective of the ear of stimulation despite dominance of contralateral connections in ascending auditory pathways. Results provide evidence that the right hemisphere plays a specific and important role in speech processing and support the hypothesis that acoustic processing of speech involves the decomposition of the signal into constituent temporal features by rate-specialized neurons in right- and left-hemisphere auditory cortex.

Keywords: Speech syllable response; Brainstem response; Auditory brainstem response; Frequency-following response; Effects of noise

Introduction

Speech processing, defined as the neural operations responsible for transforming acoustic speech input into linguistic representations, is a well-established aspect of human cortical function. Classically, speech processing has been thought to be mediated primarily by left-hemisphere auditory areas of the cerebral cortex (Wernicke, 1874). This view continues to receive wide acceptance based on results from studies investigating the functional neuroanatomy of speech perception. *Acoustical* processing of speech involves cortical analysis of the physical features of the speech signal, and normal speech perception relies on resolving acoustic events occurring on the order of tens of milliseconds (Phillips and Farmer, 1990; Tallal et al., 1993). Since temporal processing of these rapid acoustic features has been shown to be the domain of left-hemisphere auditory cortex (Belin et al., 1998; Liegeois-Chauvel et al., 1999; Zatorre and Belin, 2001;

Zaehle et al., 2004; Meyer et al., 2005), acoustic processing of speech is thought to be predominantly mediated by left-hemisphere auditory structures (Zatorre et al., 2002). *Phonological* processing of speech, which involves mapping speech sound input to stored phonological representations, has been shown to involve a network in the superior temporal sulcus (STS) lateralized to the left-hemisphere (Scott et al., 2000; Liebenthal et al., 2005; Obleser et al., 2007). *Semantic* processing of speech, which involves retrieving the appropriate meanings of words, is thought to occur in a network localized to left inferior temporal (Rodd et al., 2005) and frontal (Wagner et al., 2001) gyri.

A recent hypothesis, called the "asymmetric sampling in time" (AST) hypothesis, has challenged the classical model by proposing that *acoustical* processing of speech occurs bilaterally in auditory cortex based on the component rates inherent to the speech signal (Poeppel, 2003). Acoustic-rate asymmetry is thought to precede language-based asymmetries (i.e., phonological and semantic asymmetries) and is supported by results which show that slow, non-speech acoustic stimuli (3-5 Hz) are lateralized to right-hemisphere auditory areas (Boemio et al., 2005) while rapid acoustic stimuli (20-50 Hz) are lateralized to left-hemisphere auditory areas (Zatorre and Belin, 2001; Zaehle et al., 2004; Schonwiesner et al., 2005).

It is not known to what extent this putative mechanism applies to the slow temporal features in speech, known as the speech envelope (Rosen, 1992). The speech envelope provides syllable pattern information and is considered both sufficient (Shannon et al., 1995) and essential (Drullman et al., 1994a) for normal speech perception. A prediction of the AST hypothesis is that slow acoustic features in speech are processed in right-hemisphere auditory areas

irrespective of left-dominant asymmetries for language processing. To examine this question, we measured cortical evoked-potentials in 12 normally-developing children in response to speech sentence stimuli and compared activation patterns measured over left and right temporal cortices.

Methods

The research protocol was approved by the Institutional Review Board of Northwestern University. Parental consent and the child's assent were obtained for all evaluation procedures and children were paid for their participation in the study.

Participants

Participants consisted of 12 children between 8-14 years old who reported no history of neurological or otological disease and were of normal intelligence (scores >85 on the Brief Cognitive Scale; Woodcock and Johnson, 1977). The reason for having children serve as subjects is that we are ultimately interested in describing auditory deficits in children with a variety of clinical disorders (Koch et al., 1999). A necessary step in describing abnormal auditory function is first describing these processes in normal children, as we have done here. Children were recruited from a database compiled in an ongoing project entitled Listening, Learning and the Brain. Children who had previously participated in this project and had indicated interest in participating in additional studies were contacted via telephone. All subjects were tested in one session.

Stimuli

Stimuli consisted of the sentence stimulus “The young boy left home” produced in three modes of speech: conversational, clear and compressed speech modes (Fig. 1). These three modes of speech have different speech envelope cues and were used as a means to elicit a variety of cortical activation patterns. Conversational speech is defined as speech produced in a natural and informal manner. Clear speech is a well-described mode of speech resulting from greater diction (Uchanski, 2005). Clear speech is naturally produced by speakers in noisy listening environments and enables greater speech intelligibility relative to conversational speech. There are many acoustic features that are thought to contribute to enhanced perception of clear speech relative to conversational speech, including greater intensity of speech, slower speaking rates and more pauses. Most importantly with respect to the current work, an established feature of clear speech is greater temporal envelope modulations at low frequencies of the speech envelope, corresponding to the syllable rate of speech (1-4 Hz) (Krause and Braida, 2004). With respect to the particular stimuli used in the current study, greater amplitude envelope modulations are evident in the clear speech relative to the conversational stimuli. For example, there is no amplitude cue between “The” and “young” (Fig. 1, 0-450 msec) evident in the broadband conversational stimulus envelope, however an amplitude cue is present in the broadband clear stimulus envelope. This phenomenon also occurs between the segments “boy” and “left” (Fig. 1, 450-900 msec). Compressed speech approximates rapidly-produced speech and is characterized by a higher-frequency speech envelope. Compressed speech is more difficult to perceive compared to conversational speech (Beasley et al., 1980) and has been used in a previous study investigating cortical phase-locking to the speech envelope (Ahissar et al., 2001b).

Conversational and clear sentences were recorded in a soundproof booth by an adult male speaker at a sampling rate of 16 kHz. Conversational and clear speech sentences were equated for overall duration to control for slower speaking rates in clear speech (Uchanski, 2005). This was achieved by compressing the clear sentence by 23% and expanding the conversational sentence by 23%. To generate the compressed sentence stimulus, we doubled the rate of the conversational sample using a signal-processing algorithm in Adobe Audition (Adobe Systems Inc.). This algorithm does not alter the pitch of the signal. The duration of the clear and conversational speech sentences was 1500 msec, and the duration of the compressed sentence was 750 msec.

Recording and data processing procedures

A PC-based stimulus delivery system (NeuroScan GenTask) was used to output the sentence stimuli through a 16-bit converter at a sampling rate of 16 kHz. Speech stimuli were presented unilaterally to the right ear through insert earphones (Etymotic Research ER-2) at 80 dB SPL. Stimulus presentation was pseudorandomly interleaved. To test ear-of-stimulation effects, 3 subjects were tested in a subsequent session using unilateral left-ear stimulation. The polarity of each stimulus was reversed for half of the stimulus presentations to avoid stimulus artifacts in the cortical responses. Polarity reversal does not affect perception of speech samples (Sakaguchi et al., 2000). An interval of 1 second separated the presentation of sentence stimuli. Subjects were tested in a sound-treated booth and were instructed to ignore the sentences. To promote subject stillness during long recording sessions as well as diminish attention to the auditory stimuli, subjects watched a videotape movie of his or her choice and listened to the soundtrack to the movie in the non-test ear with the sound level set <40 dB SPL. This paradigm for measuring

cortical evoked potentials has been used in previous studies investigating cortical asymmetry for speech sounds (Bellis et al., 2000; Abrams et al., 2006) as well as other forms of cortical speech processing (Kraus et al., 1996; Banai et al., 2005a; Wible et al., 2005). While it is acknowledged that cortical activity in response to a single stimulus presentation includes contributions from both the experimental speech stimulus and the movie soundtrack, auditory information in the movie soundtrack is highly variable throughout the recording session. Therefore, the averaging of auditory responses across 1000 stimulus presentations, which serves as an essential method for reducing the impact of noise on the desired evoked response, is thought to remove contributions from the movie soundtrack. Cortical responses to speech stimuli were recorded with 31 tin electrodes affixed to an Electrocap (Electrocap International, Inc.) brand cap (impedance <5 Kohm). Additional electrodes were placed on the earlobes and superior and outer canthus of the left eye. These act as the reference and eye blink monitor, respectively. Responses were collected at a sampling rate of 500 Hz for a total of 1000 repetitions each for clear, conversational and compressed sentences.

Processing of the cortical responses consisted of the following steps. First, excessively noisy segments of the continuous file (typically associated with subject movement) were manually rejected. The continuous file was high-pass filtered at 1 Hz and removal of eye-blink artifacts was accomplished using the spatial filtering algorithm provided by NeuroScan (Compumedics, Inc). The continuous file was then low-pass filtered at 40 Hz to isolate cortical contributions and the auditory evoked potentials were then downsampled to a sampling rate of 200 Hz. All filtering was accomplished using zero phase-shift filters and downsampling was accompanied by IIR low-pass filtering to correct for aliasing (Compumedics, Inc). This goal of this filtering

scheme was to match the frequency range of the speech envelope (Rosen, 1992). Responses were artifact rejected at a +/- 75 μV criterion. Responses were then subjected to noise reduction developed by our lab that has been used in improving the signal-to-noise ratio of brainstem and cortical evoked potentials. The theoretical basis for the noise reduction is that auditory evoked potentials are largely invariant across individual stimulus repetitions while the background noise is subject to variance across stimulus repetitions. Thus, the mean evoked response is significantly diminished by the fraction of repetitions that least resembles it. If these noisy responses are removed, the signal to noise ratio of the cortical response improves considerably with virtually no change to morphology of the average waveform. The algorithm calculated the average response from all 1000 sweeps for each stimulus condition at each electrode then performed Pearson's correlations between each of the 1000 individual stimulus repetitions and the average response. The 30% of repetitions with the lowest Pearson's correlations from each stimulus condition were removed from subsequent analyses, and the remaining repetitions were averaged and re-referenced to a common reference computed across all electrodes. Therefore, following the noise reduction protocol, cortical responses from each subject represent the average of ~700 repetitions of each stimulus. Data processing resulted in an averaged response for 31 electrode sites and 3 stimulus conditions measured in all 12 subjects.

Data analysis

All data analyses were performed using software written in Matlab (The Mathworks, Inc). Broadband amplitude envelopes were determined by performing a Hilbert transform on the broadband stimulus waveforms (Drullman et al., 1994a). The unfiltered amplitude envelope was low-pass filtered at 40 Hz to isolate the speech envelope (Rosen, 1992) and match the frequency

characteristics of the cortical responses; the envelopes were then resampled to 200 Hz. Data are presented for 3 temporal electrode pairs: (1) T3-T4, (2) T5-T6 and (3) Tp7-Tp8 according to the modified International 10-20 recording system (Jasper, 1958). The modification is the addition of the Tp7-Tp8 electrode pair in which Tp7 is located midway between T3 and T5, and Tp8 is located midway between T4 and T6.

Two types of analyses were performed on the data, cross-correlation and RMS analysis. First, cross correlations between the broadband speech envelope and cortical responses at each temporal electrode for the "envelope-following period" (250-1500 msec for conversational and clear stimuli, 250-750 msec for the compressed stimulus) were performed using the "xcov" function in Matlab. The peak in the cross-correlation function was found at each electrode between 50-150 msec lags and the r-value and lag at each peak were recorded. R-values were Fisher-transformed prior to statistical analysis. RMS amplitudes at each electrode were calculated for 2 different time ranges: the "onset" period was defined by the time ranges 0-250 msec for all stimuli; the "envelope-following" period was defined as 250-1500 msec for conversational and clear stimuli and 250-750 msec for the compressed stimulus.

Statistical analysis

The statistical design used a series of 3 completely "within-subjects" RMANOVAS to assess hemispheric effects for cross-correlation and RMS measures. A primary goal of this work was to describe patterns of cortical asymmetry across speech conditions, and because 2 x 3 x 3 [hemisphere x electrode pair x stimulus condition] RMANOVAs indicated no interactions involving stimulus condition, the subsequent analysis collapsed across stimulus condition and

was performed as 2 x 3 [hemisphere x electrode pair] RMANOVAs. This enabled a matched statistical comparison of each electrode pair (i.e., T3 vs. T4; T5 vs. T6; Tp7 vs. Tp8) for each subject across stimulus conditions. A 2 x 3 x 2 [hemisphere x electrode pair x stimulation ear] RMANOVA was used to assess whether asymmetry effects seen in the cross-correlation and RMS analyses affected stimulation ear. Paired, Bonferonni-corrected t-tests (2-tailed) comparing matched electrode pairs (i.e., T3 vs. T4; T5 vs. T6; Tp7 vs. Tp8) were used for all *post-hoc* analyses. RMANOVA p-values < 0.05 and paired t-test p-values < 0.01 were considered statistically significant.

Results

Inspection of raw cortical responses measured at the 6 temporal lobe electrodes to the speech sentence stimuli revealed two discrete components in all temporal lobe electrodes: (1) a large negative onset peak and (2) a series of positive peaks that appeared to closely follow the temporal envelope of the stimulus. We called the former component the "onset" and the latter component the "envelope-following" portion of the response (see Fig. 2 for clear speech stimulus; Figs. 3 and 4 for conversational and compressed conditions, respectively). Both speech onset (Warrier et al., 2004) and envelope-following components (Ahissar et al., 2001b) have been demonstrated in previous studies of human auditory cortex; this latter study called this phenomenon speech envelope "phase-locking," and the same nomenclature will be used here. To quantify cortical phase-locking to the speech envelope, we performed cross-correlations between the broadband temporal envelope of the stimulus and individual subjects' raw cortical responses from the 6 temporal lobe electrodes for all stimulus conditions. Initially, we restricted this analysis to the envelope-following component of the response, defined as the time range 250-

1500 msec (250-750 msec for the compressed speech condition); since the onset portion of the response did not appear to closely follow the temporal envelope.

Grand average cortical responses from three matched electrode pairs (Fig. 2a-c, left column) and individual subject cross-correlograms (Fig. 2d-f, right column) indicated a number of relevant features. First, a moderate linear relationship was indicated between the broadband temporal envelope of the stimulus and raw cortical responses for all temporal lobe electrodes measured across all subjects (mean peak correlation = 0.37; SD = 0.09). Second, this peak correlation occurred in the latency range of well-established, obligatory cortical potentials measured from children of this age range (Tonnquist-Uhlen et al., 2003) (mean lag = 89.1 msec; SD = 7.42 msec). Cortical potentials in this time range, measured from temporal lobe electrodes, are associated with activity originating in secondary auditory cortex (Scherg and Von Cramon, 1986; Ponton et al., 2002). Third, and most importantly, there appeared to be qualitative differences between cortical responses from right-hemisphere electrodes compared to matched electrodes of the left-hemisphere. Specifically, right-hemisphere cortical responses appeared to conform to the contours of the stimulus envelope in greater detail than left-hemisphere responses. This was further evidenced in the correlograms, which had more consistent and sharper peaks, as well as larger overall correlations, in right-hemisphere electrodes. These particular characteristics would suggest better right-hemisphere phase-locking to the speech envelope.

Speech-envelope "phase-locking" analysis

To quantify temporal envelope phase-locking, we identified the maximum in correlograms (Fig. 2, right) for lags between 50-150 msec for all stimulus conditions. This time range was selected

since previous studies have shown that cortical synchronization to the temporal structure of brief speech sounds occurs in this range (Sharma and Dorman, 2000), and most correlograms in the current data set indicated a positive peak in this time range. An initial 2 x 3 x 3 RMANOVA [hemisphere x electrode pair x stimulus condition] indicated differences in phase-locking across stimulus conditions (main effect of stimulus condition: $F_{2,22} = 19.327$; $p < 0.0001$), which was expected given significant acoustical differences between the stimuli (see Methods), however the pattern of asymmetry for cortical phase-locking was similar for the three stimulus conditions (hemisphere x stimulus condition interaction: $F_{2,22} < 1$; $p > 0.7$). Based on this result, and our interest in describing patterns of cortical asymmetry across speech conditions, we collapsed all additional statistical analyses on correlation r-values across the 3 stimulus conditions. A 2 x 3 RMANOVA [hemisphere x electrode pair] statistical analysis on peak correlation values revealed a significant main effect of hemisphere ($F_{1,35} = 21.125$; $p < 0.0001$). All three of these electrode pairs showed this hemispheric effect (left vs. right electrode, paired t-tests: $t_{10} > 3.70$, $p \leq 0.001$ for all three pairs; Fig. 5) and there was no statistical difference in the degree of asymmetry between electrode pairs (RMANOVA hemisphere x electrode interaction: $F_{2,22} = 1.206$; $p > 0.3$). To ensure that these results were not biased by our definition of the time frame of the envelope-following component of the response, we performed identical analyses on the entire response, including the onset component, and the results were the same (0-1500 msec for conversational and clear stimuli; 0-750 msec for compressed stimulus; 2 x 3 RMANOVA [hemisphere x electrode pair]; main effect of hemisphere: $F_{1,35} = 10.658$; $p = 0.002$). These data indicate that all three temporal electrode pairs showed a significant and similar pattern of right-hemisphere asymmetry for speech envelope phase-locking.

Response magnitude analysis: onset and envelope-following period

In addition to asymmetry for phase-locking, inspection of the raw cortical data also revealed an interesting pattern of response amplitudes in the onset and envelope-following response components. At stimulus onset, response amplitudes appear to be consistently greater in left-hemisphere electrodes, particularly in T5-T6 and Tp7-Tp8 electrode pairs. Given that subjects received stimulation in their right ear, this finding was anticipated based on the relative strength of contralateral connections in the ascending auditory system (Kaas and Hackett, 2000). Surprisingly, during the envelope-following period of the response, right-hemisphere responses appeared to be larger than the left for all electrode pairs.

We quantified this phenomenon by calculating RMS amplitude over the "onset" and "envelope-following" periods for all stimulus conditions (Fig. 6). First, we performed a 2 x 3 x 3 repeated-measures ANOVA [hemisphere x electrode pair x stimulus condition] on onset RMS values which revealed that stimulus condition did not affect asymmetry for RMS onset (hemisphere x stimulus condition interaction: $F_{2,22} = 1.398$; $p > 0.25$); this result enabled us to collapse all additional statistical analyses on onset RMS across the 3 stimulus conditions. Results from 2 x 3 RMANOVA [hemisphere x electrode pair] indicated that left-hemisphere responses were significantly larger than the right over the onset period (main effect of hemisphere: $F_{1,35} = 4.686$; $p = 0.037$), and there were differences in this pattern of onset asymmetry across the 3 electrode pairs (hemisphere x electrode pair interaction: $F_{2,70} = 14.805$; $p < 0.001$). Post-hoc t-tests indicated that the main effect of hemisphere for onset RMS was driven by the posterior electrode pairs while the anterior pair, T3-T4, did not contribute to this effect (paired t-tests: T3-T4, $t_{10} = 0.924$, $p > 0.35$; T5-T6, $t_{10} = 2.892$, $p = 0.007$; Tp7-Tp8, $t_{10} = 3.348$, $p=0.002$).

For the envelope-following period, a 2 x 3 x 3 repeated-measures ANOVA [hemisphere x electrode pair x stimulus condition] was performed on envelope-following RMS values. Results again revealed that stimulus condition did not affect asymmetry (hemisphere x stimulus condition interaction: $F_{2,22} = 2.244$; $p > 0.10$), enabling us to collapse all additional statistical analyses on envelope-following RMS across the 3 stimulus conditions. Results from 2 x 3 RMANOVA [hemisphere x electrode pair] for the envelope-following RMS indicated that right-hemisphere responses were significantly larger than the left at all three electrode pairs (2 x 3 RMANOVA [hemisphere x electrode pair]; main effect of hemisphere: $F_{1,35} = 32.768$; $p < 0.00001$; paired t-tests: T3-T4, $t_{10} = 5.565$, $p < 0.00001$; T5-T6, $t_{10} = 3.385$, $p = 0.002$; Tp7-Tp8, $t_{10} = 4.767$, $p < 0.0001$). These data indicate that the right-hemisphere has significantly larger response amplitudes during the envelope-following period despite being ipsilateral to the side of acoustic stimulation.

Individual subject analysis

To quantify phase-locking and RMS amplitude asymmetries within individual subjects, we entered r-values from the cross-correlation analysis and RMS amplitudes from the envelope-following period, respectively, into the asymmetry index $(R - L) / (R + L)$ using matched electrode pairs (T3-T4; T5-T6; Tp7-Tp8). Using this index, values approaching 1 indicate a strong rightward asymmetry, values approaching -1 indicate a strong leftward asymmetry, and a value of 0 indicates symmetry. Results from this analysis indicate that greater right-hemisphere phase-locking, defined as asymmetry values greater than 0, occurred in 78% of the samples (binomial test: $z = 5.96$, $p<0.0001$) and right-hemisphere r-values were more than twice as great

as those seen for the left hemisphere (mean asymmetry index = 0.35). For RMS amplitude, 82% of the samples indicated greater envelope-following amplitude in the right-hemisphere (binomial test: $z = 6.74$, $p<0.0001$), and right-hemisphere amplitudes were ~33% greater than those seen in the left hemisphere (mean asymmetry index = 0.14) during the envelope-following period.

Ear-of-stimulation analysis

To ensure that the right-hemisphere asymmetries for envelope phase-locking and RMS amplitude were not driven by the use of right-ear stimulation, we measured cortical responses to the speech sentences in 3 of the subjects using left-ear stimulation, which again enabled a completely within-subjects statistical analysis. Results indicate that when subjects were stimulated in their left ear, envelope phase-locking was again greater in the right-hemisphere (2 x 3 RMANOVA [hemisphere x electrode pair]; main effect of hemisphere: $F_{1,8} = 15.532$; $p = 0.004$). Moreover, when compared directly to responses elicited by right-ear stimulation, envelope phase-locking asymmetries were statistically similar irrespective of the ear of stimulation (Fig. 7; 2 x 3 x 2 RMANOVA [hemisphere x electrode pair x stimulation ear]; interaction [hemisphere x stimulation ear]: $F_{1,8} = .417$; $p > 0.5$). For the RMS analysis, left-ear stimulation resulted in larger onset responses in the right-hemisphere, again consistent with contralateral dominance for onsets (Fig. 8 Inset; 2 x 3 RMANOVA [hemisphere x electrode pair]; main effect of hemisphere: $F_{1,8} = 6.40$; $p = 0.035$). In addition, the asymmetry pattern for onset RMS with left-ear stimulation was statistically different from the pattern seen for right-ear stimulation (2 x 3 x 2 RMANOVA [hemisphere x electrode pair x stimulation ear]; interaction of [hemisphere x stimulation ear]: $F_{1,8} = 24.390$; $p = 0.001$). Importantly, the RMS of the envelope following period remained greater in the right-hemisphere with left-ear stimulation (Fig. 8; 2 x 3

RMANOVA [hemisphere x electrode pair]; main effect of hemisphere: $F_{1,8} = 36.028$; $p < 0.001$) and was statistically similar to the pattern of asymmetry resulting from right-ear stimulation (2 x 3 x 2 RMANOVA [hemisphere x electrode pair x stimulation ear]; interaction [hemisphere x stimulation ear]: $F_{1,8} = 0.047$; $p > 0.8$). Taken together, these data indicate that changing the ear of stimulation from right to left does not affect right-hemisphere asymmetry for envelope phase-locking or envelope RMS amplitude. On the other hand, onset RMS amplitudes are always larger in the hemisphere contralateral to the ear of stimulation.

Discussion

Biologically-significant acoustic signals contain information on a number of different time scales. The current study investigates a proposed mechanism for how the human auditory system concurrently resolves these disparate temporal components. Results indicate right-hemisphere dominance for coding the slow temporal information in speech known as the speech envelope. This form of asymmetry is thought to reflect acoustic processing of the speech signal and was evident despite well-known leftward asymmetries for processing linguistic elements of speech. Furthermore, rightward asymmetry for the speech envelope was unaffected by the ear of stimulation despite the dominance of contralateral connections in ascending auditory pathways.

Models of speech perception and the AST hypothesis

The neurobiological foundation of language has been a subject of great interest for well over a century (Wernicke, 1874). Recent studies using functional imaging techniques have enabled a detailed description of the functional neuroanatomy of spoken language. The accumulated results have yielded hierarchical models of speech perception consisting of a number of discrete

processing stages, including acoustic, phonological and semantic processing of speech (Hickok and Poeppel, 2007; Obleser et al., 2007).

It is generally accepted that each of these processing stages is dominated by left-hemisphere auditory and language areas. The acoustic basis of speech perception is typically investigated by measuring cortical activity in response to speech-like acoustic stimuli which have no linguistic value but contain acoustic features that are necessary for normal speech discrimination. Acoustic features lateralized to left-hemisphere auditory areas include rapid frequency transitions (Belin et al., 1998; Joanisse and Gati, 2003; Meyer et al., 2005) and voice-onset time (Liegeois-Chauvel et al., 1999; Zaehle et al., 2004), both of which are necessary for discriminating many phonetic categories. The cortical basis for phonological processing of speech has been investigated by measuring neural activation in response to speech phoneme (Obleser et al., 2007), syllable (Liebenthal et al., 2005), word (Binder et al., 2000) and sentence (Scott et al., 2000; Narain et al., 2003) stimuli while carefully controlling for the spectrotemporal acoustic characteristics of the speech signal. Results from these studies have consistently demonstrated that a region of the left-hemisphere STS underlies phonological processing of speech. Studies of cortical processing of semantic aspects of speech have measured brain activation while the subject performed a task in which semantic retrieval demands were varied. Results from these studies have shown that activation of inferior temporal (Rodd et al., 2005) and frontal (Wagner et al., 2001) gyri, again biased to the left hemisphere, underlie semantic processing. It should be noted that right-hemisphere areas are also activated in studies of acoustical, phonological and semantic speech processing, however left-hemisphere cortical structures have typically shown dominant activation patterns across studies.

Results from the current study are among the first to show that the right-hemisphere of cerebral cortex is dominant during speech processing. These data contradict the conventional thinking that language processing consists of neural operations largely confined to the left-hemisphere of the cerebral cortex. Moreover, results from the current study show right-dominant asymmetry for the speech envelope despite these other well-established forms of leftward asymmetry.

Results add to the literature describing hierarchical models of speech processing by providing important details about the initial stage of cortical speech processing: pre-linguistic, acoustic processing of speech input. Results support the notion that the anatomical basis of speech perception is initially governed by the component rates present in the speech signal. This statement raises a number of interesting question regarding hierarchical models of speech perception. What is the next stage of processing for syllable pattern information in right-hemisphere auditory areas? Does slow temporal information in speech follow a parallel processing route relative to phonological processing? It is hoped that these questions will receive additional consideration and investigation.

Right-hemisphere dominance for slow temporal features in speech supports the AST hypothesis, which states that slow temporal features in acoustic signals lateralize to right-hemisphere auditory areas while rapid temporal features lateralize to the left (Poeppel, 2003). Results extend the AST hypothesis by providing a new layer of detail regarding the nature of this asymmetric processing. Beyond showing asymmetry for the magnitude of neural activation (RMS amplitude results; Fig. 6), which might have been predicted from previous studies, our

results show that right-hemisphere auditory neurons follow the contours of the speech envelope with greater precision compared to the left-hemisphere (Fig. 5). This is an important consideration, as this characteristic of right-hemisphere neurons had not been proposed in previous work and could represent an important cortical mechanism for speech envelope coding.

An influential hypothesis that predates AST states that there is a relative trade-off in auditory cortex for representing spectral and temporal information in complex acoustic signals such as speech and music (Zatorre et al., 2002). It is proposed that temporal resolution is superior in left-hemisphere auditory cortex at the expense of fine-grained spectral processing whereas the right-hemisphere's superior spectral resolution is accompanied by reduced temporal resolution. The current results suggest that there is in fact excellent temporal resolution in the right-hemisphere, but it is limited to a narrow range of low frequencies. However, it is not known to what extent the asymmetries demonstrated here might reflect the right-hemisphere's preference for spectral processing.

Previous studies investigating envelope representations

Previous studies of the human auditory system have described cortical encoding of slow temporal information in speech. In one study, it was shown that cortical phase-locking and frequency-matching to the speech envelope predicted speech comprehension using a set of compressed sentence stimuli (Ahissar et al., 2001b). There are a few important differences between the current work and Ahissar's. First, hemispheric specialization was not reported in Ahissar's work. Second, the analyses (i.e., phase-locking, frequency matching) were conducted on the average of multiple speech sentences with similar envelope patterns, which was necessary

given the parameters of the simultaneous speech comprehension task. In contrast, cortical responses in the current study represent activity measured to isolated sentence stimuli and enable a more detailed view of cortical following to individual sentences (Figs. 2-4).

The current results also show similarities to findings from a recent study that investigated rate processing in human auditory cortex in response to speech (Luo and Poeppel, 2007). In this study it was shown that different speech sentence stimuli elicited cortical activity with different phase patterns in the theta band (4-8 Hz), and theta-band dissimilarity was lateralized to the right hemisphere. A limitation of this work is that cortical responses were not compared to the stimulus; the analysis only compared cortical responses elicited by the various speech stimuli. Therefore, it was not transparent that the theta-band activity was driven by phase-locking to the speech envelope. Although many of the conclusions are the same as those described here, to our knowledge, our experiment is the first to explicitly show right-hemisphere dominance for phase-locking to the speech envelope.

Single-unit studies of auditory cortex in animal models suggest potential mechanisms underlying right-hemisphere dominance for coding the speech envelope. Across a variety of animal models, a sizable population of auditory cortical neurons is synchronized to the temporal envelope of species-specific calls (Wang et al., 1995; Nagarajan et al., 2002; Gourevitch and Eggermont, 2007) which show many structural similarities to human speech; one such study called these neurons “envelope peak-tracking units” (Gehr et al., 2000). One possible explanation for right-dominant asymmetry for envelope phase-locking is that a disproportionate number of envelope peak-tracking units exist in the right-hemisphere auditory cortex of humans. Future studies with

near-field recordings in humans (Liégeois-Chauvel et al., 2004) may be able to address this question.

A potential limitation of this work is that children served as subjects, and it is not known whether right-hemisphere speech envelope effects also occur in adults. While the current data cannot discount this possibility, we believe this is unlikely based on the fact that adults show cortical phase-locking to the speech envelope (Ahissar et al., 2001b) and have previously demonstrated a right-hemisphere preference for slow, non-speech acoustic stimuli (Boemio et al., 2005). An interesting possibility is that children have pronounced syllable-level processing relative to adults, reflecting a stage in language acquisition. Future studies may be able be better delineate the generality of this hemispheric asymmetry as well as possible interactions with language development in normal and clinical populations.

Across languages, the syllable is considered a fundamental unit of spoken language (Gleason, 1961), although there is debate as to its phonetic definition (Ladefoged, 2001) . The speech envelope provides essential acoustic information regarding syllable patterns in speech (Rosen, 1992) and psychophysical studies of the speech envelope have demonstrated that it is an essential acoustic feature for speech intelligibility (Drullman et al., 1994a). Results described here provide evidence that a cortical mechanism for processing syllable patterns in on-going speech is the routing of speech envelope cues to right-hemisphere auditory cortex. Given the universality of the syllable as an essential linguistic unit and the biological significance of the speech signal, it is plausible that discrete neural mechanisms, such as those described here, may have evolved to code this temporal feature in the human central auditory system.

Acknowledgements

Acknowledgements: We thank C. Warrier, D. Moore and three anonymous reviewers for helpful comments on a previous draft of this manuscript. We also thank the children who participated in this study and their families. This work is supported by the National Institutes of Health grant R01 DC01510-10 and National Organization for Hearing Research grant 340-B208.

CHAPTER III: ABNORMAL CORTICAL PROCESSING OF SYLLABLE RATES IN SPEECH IN READING-IMPAIRED CHILDREN

Submitted for Publication in *Proceedings of the National Academy of Sciences of the United States of America*

Daniel A. Abrams[1], Trent Nicol[1], Steven Zecker[1], and Nina Kraus[1,2]

[1]Auditory Neuroscience Laboratory, Department of Communication Sciences, and [2]Departments of Neurobiology and Physiology and Otolaryngology. Northwestern University, 2240 Campus Drive, Evanston, Illinois, 60208.

Abstract

Children with reading impairments (RI) have long been associated with impaired perception for rapidly presented acoustic stimuli and recently have shown deficits for slower features. It is not known whether impairments for slow acoustic features negatively impacts processing of speech in RI. Here we provide neurophysiological evidence that RIs have impaired representation of the speech envelope, the acoustical cue that provides syllable pattern information in speech. While normal-reading controls indicated consistent right-hemisphere dominance for speech envelope representation in auditory cortex, RI representations were symmetrical across the cerebral hemispheres in an acoustically challenging stimulus condition. Cortical representations were temporally delayed in RI and correlated with standardized measures of literacy across all subjects. Results support the hypothesis that asymmetric routing between cerebral hemispheres represents a mechanism for temporal encoding in the human auditory system, and argue for an

expansion of the temporal processing hypothesis for reading-disabilities to encompass impairments for both rapid and slow acoustic features.

Keywords: auditory cortex, dyslexia, speech

Introduction

Reading impairment (RI), a disorder that affects ~5% of the general population, is defined as a specific deficit in reading that is independent of intelligence, motivation, educational opportunity and overt neurological damage. There is general consensus that many RIs suffer from a phonological deficit, defined as an impairment in the representation and processing of speech sounds (Ramus, 2003), and a large body of evidence has accumulated that shows that many RIs are also impaired in the perception of rapid acoustic events in speech and non-speech signals (Tallal and Piercy, 1973; Tallal, 1980; Kraus et al., 1996). An influential hypothesis (Tallal et al., 1998) poses that abnormal perception of rapid acoustic events present in speech (on the order of tens of msec) precludes normal development of phonological systems since many phonological contrasts rely on resolving acoustic events occurring on this time scale (Phillips and Farmer, 1990).

A more recent hypothesis states that abnormal perception of slow temporal features in speech and non-speech signals (on the order of hundreds of msec) additionally contribute to reading impairments (Goswami, 2002). In speech, these slow acoustic features, known as the speech envelope, provide syllable pattern information and segmental cues for phoneme identity and are thought to be extremely important for normal speech perception (Drullman et al., 1994a). In

support of this hypothesis, it has been shown that deficits in the perception of slow temporal cues in non-speech acoustic signals accounts for significant variance in reading scores (Goswami et al., 2002; Witton et al., 2002), however it is not known whether this impairment impacts processing of the speech signal.

Recent work has provided new insight into the neural mechanisms governing temporal processing of speech in the unimpaired human auditory system. Specifically, it has been shown that a mechanism for processing temporal information in speech is the asymmetric routing of this information between the cerebral hemispheres (Poeppel, 2003): rapid features in speech are lateralized to left-hemisphere auditory areas (Belin et al., 1998; Liegeois-Chauvel et al., 1999; Joanisse and Gati, 2003; Abrams et al., 2006) while the speech envelope is lateralized to right-hemisphere auditory areas (Abrams et al., in press). A prediction of this model is that impaired perception of speech envelope cues, such as that demonstrated by RIs, will be accompanied by a disruption of this lateralization of slow features in speech. However, the rapid processing deficit of dyslexia would predict normal representation of the speech envelope in RI, and no relationship between speech envelope representation and phonological and reading abilities. To test these hypotheses, we measured cortical evoked-potentials in 12 normally-developing and 11 children with RI in response to sentence stimuli and compared activation patterns measured over left and right temporal cortices. We also examined the relationship between standardized measures of literacy and phonological processing and measures of cortical envelope processing.

Methods

The research protocol was approved by the Institutional Review Board of Northwestern University. Parental consent and the child's assent were obtained for all evaluation procedures and children were paid for their participation in the study.

Participants

Participants consisted of 23 children between 9-15 years old who reported no history of neurological or otological disease and were of normal intelligence (scores >85 on the Brief Cognitive Scale; Woodcock and Johnson, 1977). Children with reading impairments consisted of children who had been formally identified as such by an independent pyschoeducational diagnostician. Mean reading and spelling scores for RIs are below average (average is a score of 100 for these standardized test scores), but are not in the "impaired" range, defined as a score of <85 (Table 1). In addition, mean reading and spelling scores of NLs are considered above average. Nevertheless, controls differed significantly from RIs on measures of auditory processing (Woodcock and Johnson, 1989) (a composite score derived from the Incomplete Words and Sound Blending subtests), single-word reading and spelling ($P<0.001$ for these three tests) (Wilkinson, 1993). Since cerebral dominance for language is often reversed in left-handed and ambidextrous individuals, and many of the analyses performed here relate to cerebral asymmetries, all subjects completed a questionnaire to assess handedness.

Children were recruited from a database compiled in an ongoing project entitled Listening, Learning and the Brain. Children who had previously participated in this project and had indicated interest in participating in additional studies were contacted via telephone.

Neurophysiologic responses were collected in a single session and measures of literacy and phonological processing were measured in a subsequent session.

Stimuli

Stimuli consisted of the sentence stimulus "The young boy left home" produced in three modes of speech: conversational, clear and compressed modes (Fig. 1). These three modes of speech have different speech envelope cues and were used as a means to elicit a variety of cortical activation patterns. Conversational speech is defined as speech produced in a natural and informal manner. Clear speech is a well-described mode of speech resulting from greater diction (Uchanski, 2005). Clear speech is naturally produced by speakers in noisy listening environments and enables greater speech intelligibility relative to conversational speech. There are many acoustic features that are thought to contribute to enhanced perception of clear speech relative to conversational speech, including greater intensity, slower speaking rate and more pauses. Most importantly with respect to the current work, an established feature of clear speech is greater temporal envelope modulations at lower modal frequencies than conversational speech Modal frequency corresponds to the syllable rate of speech (1-4 Hz) (Krause and Braida, 2004). With respect to the particular stimuli used in the current study, greater amplitude envelope modulations are evident in the clear speech relative to the conversational stimuli. For example, there is no amplitude cue between "The" and "young" evident in the broadband conversational stimulus envelope, however an amplitude cue is present in the broadband clear stimulus envelope (Fig. 1, 0-450 msec). This phenomenon also occurs between the segments "boy" and "left" (450-900 msec). Compressed speech approximates rapidly-produced speech and is characterized by a higher modal frequency. Compressed speech is more difficult to perceive compared to

conversational speech (Beasley et al., 1980) and has been used in a previous study investigating cortical phase-locking to the speech envelope (Ahissar et al., 2001b).

Conversational and clear sentences were recorded in a soundproof booth by an adult male speaker at a sampling rate of 16 kHz. Conversational and clear speech sentences were equated for overall duration to control for slower speaking rates in clear speech (Uchanski, 2005). This was achieved by compressing the clear sentence by 23% and expanding the conversational sentence by 23%. To generate the compressed sentence stimulus, we doubled the rate of the conversational sample using a signal-processing algorithm in Adobe Audition (Adobe Systems Inc.). This algorithm does not alter the pitch of the signal. The duration of the clear and conversational speech sentences was 1500 msec, and the duration of the compressed sentence was 750 msec.

Neurophysiologic recording and data processing procedures

All of the recording and data processing techniques used to describe cortical representation of the speech envelope are identical to those described in detail in a recent publication (Abrams et al., in press). A PC-based stimulus delivery system (Neuroscan GenTask, Compumedics, Inc) was used to output the sentence stimuli through a 16-bit converter at a sampling rate of 16 kHz. Speech stimuli were presented unilaterally to the right ear through insert earphones (Etymotic Research ER-2) at 80 dB SPL. Stimulus presentation was pseudorandomly interleaved. The polarity of each stimulus was reversed for half of the stimulus presentations to avoid stimulus artifacts in the cortical responses. Polarity reversal does not affect perception of speech samples (Sakaguchi et al., 2000). An interval of 1 second separated the presentation of sentence stimuli.

Subjects were tested in a sound-treated booth and were instructed to ignore the sentences. To promote subject stillness during long recording sessions as well as diminish attention to the auditory stimuli, subjects watched a videotape movie of his or her choice and listened to the soundtrack to the movie in the non-test ear with the sound level set <40 dB SPL. This paradigm for measuring cortical evoked potentials has been used in previous studies investigating cortical asymmetry for speech sounds (Bellis et al., 2000; Abrams et al., 2006) as well as other forms of cortical speech processing (Kraus et al., 1996; Banai et al., 2005a; Wible et al., 2005). While it is acknowledged that cortical activity in response to a single stimulus presentation includes contributions from both the experimental speech stimulus and the movie soundtrack, auditory information in the movie soundtrack is highly variable throughout the recording session. Therefore, the averaging of auditory responses across 1000 stimulus presentations, which serves as an essential method for reducing the impact of noise on the desired evoked response, effectively removes contributions from the movie soundtrack. Cortical responses to speech stimuli were recorded with 31 tin electrodes affixed to an Electrocap (Electrocap International, Inc.) brand cap (impedance <5 Kohm). Additional electrodes were placed on the earlobes and superior and outer canthus of the left eye. These act as the reference and eye blink monitor, respectively. Responses were collected (Neuroscan Acquire, Compumedics Inc) at a sampling rate of 500 Hz for a total of 1000 repetitions each for clear, conversational and compressed sentences.

Processing of the cortical responses consisted of the following steps. First, excessively noisy segments of the continuous file (typically associated with subject movement) were manually rejected. The continuous file was high-pass filtered at 1 Hz and removal of eye-blink artifacts

was accomplished using the spatial filtering algorithm provided by Neuroscan. The continuous file was then low-pass filtered at 40 Hz to isolate cortical contributions and the auditory evoked potentials were then downsampled to a sampling rate of 200 Hz. All filtering was accomplished using zero phase-shift filters and downsampling was accompanied by IIR low-pass filtering to correct for aliasing (Compumedics USA, Inc). This goal of this filtering scheme was to match the frequency range of the speech envelope (Rosen, 1992). Responses were artifact rejected at a +/- 75 μV criterion. Responses were then subjected to noise reduction developed by our lab that has been used in improving the signal-to-noise ratio of brainstem and cortical evoked potentials (Abrams et al., in press). The theoretical basis for the noise reduction is that auditory evoked potentials are largely invariant across individual stimulus repetitions while the background noise is subject to variance across stimulus repetitions. Thus, the mean evoked response is significantly diminished by the fraction of repetitions that least resembles it. If these noisy responses are removed, the signal to noise ratio of the cortical response improves considerably with virtually no change to morphology of the average waveform. The algorithm calculated the average response from all sweeps for each stimulus condition at each electrode then performed Pearson's correlations between each of the individual stimulus repetitions and the average response. The 30% of repetitions with the lowest Pearson's correlations from each stimulus condition were removed from subsequent analyses, and the remaining repetitions were averaged and re-referenced to a common reference computed across all electrodes. Therefore, following the noise reduction protocol, cortical responses from each subject represent the average of ~700 repetitions of each stimulus. Data processing resulted in an averaged response for 31 electrode sites and 3 stimulus conditions measured in all 23 subjects.

Data analysis: measures of cortical speech envelope processing

All of the analyses techniques used to describe cortical representation of the speech envelope are identical to those described in detail in a recent publication (Abrams et al., in press), in press #134}. All data analyses were performed using software written in Matlab (The Mathworks, Inc). Broadband stimulus envelopes were determined by performing a Hilbert transform on the broadband sentence waveforms (Drullman et al., 1994a). The resulting amplitude envelopes were low-pass filtered at 40 Hz and resampled at 200 Hz to isolate the speech envelope (Rosen, 1992) and to match the frequency characteristics and sampling rate of the processed responses. We calculated the frequency of maximal power, known as the modal frequency (Ahissar et al., 2001b), of the envelope of each speech sentence stimulus by performing a fast Fourier transforms of the low-pass filtered Hilbert envelope. FFTs were calculated using windows of 1 s and overlaps of 0.5 s, consistent with a previous report (Ahissar et al., 2001b).

Data are presented for 3 temporal electrode pairs: (1) T3 and T4, (2) T5 and T6 and (3) Tp7 and Tp8 according to the modified International 10-20 recording system (Jasper, 1958). The modification is the addition of the Tp7-Tp8 electrode pair in which Tp7 is located midway between T3 and T5, and Tp8 is located midway between T4 and T6. Two types of analyses were performed on the data: cross-correlation and RMS analyses, resulting in three measures of cortical speech-envelope representation. First, cross correlations between the broadband speech envelope and cortical responses at each temporal electrode for the "envelope-following period" (250-1500 msec for conversational and clear stimuli, 250-750 msec for the compressed stimulus) were performed using the "xcov" function in Matlab. The peak in the cross-correlation function was found at each electrode between 50-150 msec lags, resulting in the first two measures. (1)

Phase-locking precision was defined as the peak r-value and (2) *phase-locking timing* was defined as the lag at the peak r-value. R-values were Fisher-transformed prior to statistical analysis. Finally, (3) *RMS amplitudes* at each electrode were calculated for 2 different time ranges: the "onset" period was defined by the time ranges 0-250 msec for all stimuli; the "envelope-following" period was defined as 250-1500 msec for conversational and clear stimuli and 250-750 msec for the compressed stimulus.

Data analysis: relationship between measures of cortical speech envelope processing and literacy and phonological processing

With respect to phase-locking precision, RIs were most clearly differentiated from control subjects based on the symmetry of their r-values in the compressed speech condition (see "Speech envelope phase-locking precision" in Results). Therefore, we first calculated mean left-hemisphere r-values (i.e., the mean of T3, T5 and Tp7) and right hemisphere r-values (i.e., the mean of T4, T6 and Tp8) from the cross-correlation analysis of the compressed speech condition, and entered these values into the asymmetry index (R – L) / (R + L) (Fig. 9). Collapsing across electrode pairs was justified in this instance since there was no main effect of electrode, or interaction involving electrode, in a 2 x (3 x 2) mixed-model ANOVA [group x electrode x hemisphere] with repeated-measures on the final 2 factors in which compressed speech r-value was the dependant variable ($F < 2.4$, $p > 0.10$ for main effects and interactions). The asymmetry index values were used in subsequent comparisons to standardized measures of literacy and phonological processing.

With regards to phase-locking timing, RIs were differentiated from control subjects based on a general delay in right-hemisphere cortical lags that was evident across all speech conditions (see "Speech envelope phase-locking timing" in Results). Therefore, we first averaged right-hemisphere cortical lags across T4, T6 and Tp8 electrodes for each speech condition (Fig. 9). Collapsing across right-hemisphere electrode pairs was justified since there was no main effect of electrode pair or [electrode x group] interaction in a 2 x (3 x 3) [group x stimulus x electrode] mixed-model ANOVA with repeated-measures on the final two factors with right-hemisphere lags serving as the dependant variable (main effect of electrode: $F_{2,42} = 0.032$, $p = 0.969$; electrode x group interaction: $F_{2,42} = 1.108$, $p = 0.340$). Since results from this ANOVA analysis also showed no main effect of stimulus ($F_{2,42} = 1.865$, $p = 0.168$) or [stimulus x group] interaction ($F_{2,42} = 0.264$, $p = 0.769$), we then averaged right-hemisphere lags for each subject across the three stimulus conditions. The resultant values are the averaged right-hemisphere lag across stimulus conditions for each subject, and these values were used in subsequent comparisons to standardized measures of literacy and phonological processing.

With respect to RMS amplitude, RIs were most clearly differentiated from control subjects based on the symmetry of their envelope-following amplitudes in the compressed speech condition (see "Onset and speech envelope period RMS amplitude analyses" in Results). Therefore, we first calculated mean left-hemisphere amplitudes (i.e., the mean of T3, T5 and Tp7) and right hemisphere amplitudes (i.e., the mean of T4, T6 and Tp8) from the analysis of the compressed speech condition, and entered these values into the asymmetry index $(R - L) / (R + L)$ (Fig. 9). Collapsing across electrode pairs was justified in this instance since there was no main effect of electrode ($F_{2,42} = 1.221$, $p = 0.305$) or [electrode x group] interaction ($F_{2,42} = 0.855$, $p = 0.433$) in

a 3 x 2 x 2 RMANOVA [electrode x hemisphere x group] in which compressed speech RMS amplitude was the dependant variable. The asymmetry index values were used in subsequent comparisons to standardized measures of literacy and phonological processing.

Statistical analysis

The statistical design used a series of mixed-model and repeated-measures ANOVAs to assess group differences for cross-correlation and RMS measures. When appropriate, *post-hoc* ANOVAs were conducted using repeated measures whenever possible to maximize statistical power. Repeated measures ANOVA p-values < 0.05 were considered statistically significant. Due to the relatively poor spatial resolution of EEG measures, hemispheric differences within and between groups were only considered significant if they occurred as main effects.

Pearson's correlations were performed between measures of cortical speech envelope processing and standardized measures of literacy and phonological processing. To prevent spurious results from correlation analyses and t-tests, which can be overly biased by outlying data points, all raw cortical asymmetry values beyond 2 standard deviations (SD) of the mean were moved to the 2 SD point for that particular measure. Across cortical measures, 2 data points (out of a total of 69) were moved to the 2 SD point. Pearson's correlations p-values < 0.05 were considered statistically significant.

Results

Cortical envelope representation in normal and RI children

Control subjects' data and most of the analyses used here were employed in a recent publication (Abrams et al., in press). Figure 10 shows grand average cortical responses to clear (top left) and compressed (top right) speech sentences from control (n=12) and RI (n=11) groups. These two conditions were plotted because they represent the most intelligible (clear) and most challenging (compressed) speech conditions, and impaired populations often reveal abnormal perception and representation of speech only in difficult listening conditions (Watson et al., 1990; Cunningham et al., 2001a; Bradlow et al., 2003).

Speech envelope phase-locking precision

To investigate *precision* and *timing* of envelope phase-locking, we cross-correlated the broadband envelope with left and right-hemisphere temporal electrode responses in all subjects, and plotted mean correlograms for clear (Fig. 10, bottom left) and compressed (Fig. 10, bottom right) conditions. For the clear speech condition, mean correlograms for both controls and RIs indicated asymmetry for phase-locking *precision*, seen as the y-axis of the correlogram: correlogram peaks were considerably greater in right-hemisphere electrodes for both groups. These observations were investigated with a 2 x (3 x 2) mixed-model ANOVA [group x electrode x hemisphere] with repeated-measures (RM) on the final two factors. The dependant variable in this analysis was the Z-value associated with the peak Pearson's "r" value for each subject from 3 paired electrode sites in response to the clear speech condition. Results confirmed our observations that Pearson's "r" values were greater in the right-hemisphere in the clear speech condition (main effect of hemisphere: $F_{1,21} = 13.267$, $p = 0.002$; Fig. 11; see Table 2 for descriptive statistics) and were similar across groups ([hemisphere x group] interaction: $F_{1,21} = 0.110$, $p = 0.774$). The same rightward asymmetric pattern was seen for the conversational

condition in both groups (main effect of hemisphere: $F_{1,21} = 9.944$, $p = 0.005$; [hemisphere x group] interaction: $F_{1,21} = 0.810$, $p = 0.378$).

In the compressed speech condition, controls appeared to show the same pattern of right-dominant asymmetry for precision of phase-locking, but RIs showed a strikingly different pattern of activation. Specifically, phase-locking precision appeared to be represented symmetrically across cerebral hemispheres (Figure 10, bottom right). These observations were confirmed with statistical analyses: in compressed speech, ANOVA results failed to show a main effect of hemisphere ($F_{1,21} = 2.019$, $p = 0.17$) but did show a significant group difference (hemisphere x group interaction: $F_{1,21} = 6.604$, $p = 0.018$). *Post-hoc* analyses confirmed that controls continued to show right-dominant asymmetry for precision of phase-locking in the compressed condition (3 x 2 RMANOVA [electrode x hemisphere] main effect of hemisphere: $F_{1,11} = 7.067$, $p = 0.022$). In contrast, r-values in RIs were statistically similar between right and left-hemisphere electrodes (main effect of hemisphere: $F_{1,10} = 0.786$, $p = 0.396$). This symmetric pattern in RIs to compressed speech was driven by both increases in left-hemisphere phase-locking relative to conversational speech (comparison between conversational vs. compressed conditions, main effect of stimulus: $F_{1,10} = 20.204$, $p = 0.001$) as well as a trend for decreases in right-hemisphere phase-locking (main effect of stimulus: $F_{1,10} = 4.312$, $p = 0.065$). These data indicate that in clear and conversational speech conditions, right hemisphere auditory cortex was dominant for phase-locking precision in both normal and RIs, but when presented with compressed speech, only controls showed right-hemisphere dominance.

Independent of asymmetry patterns, both groups appeared to show greater r-values in both hemispheres in the compressed condition compared to clear and conversational conditions (Fig. 11, top). This observation was confirmed in both groups (control compressed vs. clear: 3 x 2 x 2 RMANOVA [electrode x stimulus x hemisphere] main effect of stimulus: $F_{1,11} = 15.515$, $p = 0.002$; control compressed vs. conversational $F_{1,11} = 39.462$, $p < 0.0001$; RI compressed vs. clear: $F_{1,10} = 5.498$, $p = 0.041$; RI compressed vs. conversational: $F_{1,10} = 12.702$, $p = 0.005$). This finding likely reflects the fact that compressed stimuli were half the duration of the clear and conversational stimuli: studies have shown that auditory cortical neurons have reduced activity towards the end of a relatively long acoustic stimulus (Nagarajan et al., 2002), an effect that could conceivably translate into a reduction in phase-locking precision in far-field potentials.

Speech envelope phase-locking timing

We investigated the *timing* of cortical phase-locking to the speech envelope by analyzing the lags of subjects' correlogram peaks. A pattern that was evident across conditions was that controls had earlier lags in right-hemisphere electrodes (Figure 11, bottom left) while RIs showed later lags in right-hemisphere electrodes (Figure 11, bottom right). We performed a 2 x (3 x 3 x 2) mixed-model ANOVA [group x stimulus x electrode x hemisphere] with repeated-measures on the final three factors with lags as the dependent variable and found a significant interaction of hemisphere x group ($F_{1,21} = 5.959$, $p= 0.024$; see Table 2 for descriptive statistics). Since there was no main effect of stimulus condition ($F_{2,42} = 0.611$, $p= 0.547$) or hemisphere x stimulus x group interaction ($F_{2,42} = 0.020$, $p= 0.980$), we collapsed results across the three stimulus conditions for *post-hoc* analyses to improve statistical power. Consistent with our observation that cortical responses appeared to be delayed in the RIs, a *post-hoc* 2 x 3 mixed-model ANOVA

[group vs. right-hemi electrodes] with repeated-measures on the final factor showed that the initial interaction was driven by increased lags in right-hemisphere electrodes in RIs compared to controls (main effect of group: $F_{1,67} = 8.840$, $p= 0.004$). The same analysis performed on left-hemisphere electrodes showed that left-hemisphere lags were statistically similar between groups (main effect of group: $F_{1,67} = 0.793$, $p= 0.376$). This result indicates that the right hemisphere, which appears to be the dominant hemisphere for encoding the speech envelope (Abrams et al., in press), is delayed in its representation in RIs.

Onset and speech envelope period RMS amplitude analyses

To examine the magnitude of activation across left and right-hemisphere auditory cortex, we performed RMS amplitude analyses on the "onset" and "envelope-following" portions of the response (Abrams et al., in press). Mean amplitudes in the onset segment did not appear to indicate group differences (Figure 12), a result that was confirmed with a 2 x (3 x 3 x 2) mixed-model ANOVA [group x stimulus x electrode x hemisphere] with repeated-measures on the final three factors using onset RMS as the dependent variable (main effect of group and interactions involving group: $F \leq 1.768$, $p > 0.180$ for all group-related results). For envelope-following RMS, both groups appeared to show similar patterns of rightward asymmetry in clear and conversational conditions, but it appeared that this asymmetry was disrupted in the compressed condition for RIs only (Fig. 13). These observations were investigated with three separate 2 x (3 x 2) mixed-model ANOVAs [group x electrode x hemisphere] with repeated-measures on the final two factors in which the dependant variable was the envelope-following RMS amplitude measured in each stimulus condition. Consistent with our observations, right dominant asymmetry was evident across subjects for the clear condition (main effect of hemisphere: $F_{1,21} =$

7.634, p = 0.012) and showed a trend for significance in the conversational condition (main effect of hemisphere: $F_{1,21} = 3.506$, p = 0.075); there were no group differences for either the clear or conversational conditions (main effect of group, group x hemisphere interaction: $F_{1,21} \leq 1.591$, $p \geq 0.221$ for clear and conversational). In the compressed condition, there was a group difference with respect to asymmetry for envelope-following RMS amplitude (group x hemisphere interaction: $F_{1,21} = 5.341$, p = 0.031). *Post-hoc* RMANOVAs measured within the control group indicated a strong effect of asymmetry (main effect of hemisphere: $F_{1,11} = 19.131$, p = 0.001) while RIs showed symmetric responses (main effect of hemisphere: $F_{1,10} = 0.104$, p = 0.754). Similar to the Pearson's "r" values in the compressed condition, this symmetric pattern in RI amplitudes was driven by increases in left-hemisphere amplitudes (comparison between conversational vs. compressed conditions, main effect of stimulus: $F_{1,10} = 15.990$, p = 0.003).

Cortical representation of the speech envelope and measures of literacy and phonological processing

Cortical representation of the speech envelope differed significantly between RIs and controls on three measures: (1) phase-locking precision, (2) phase-locking timing, and (3) RMS amplitude. If cortical coding of the speech envelope represents an important factor for the development of normal reading and phonology, it should not only differ between control and RIs as we have shown, but variance in these measures of cortical representations should be able to predict scores on standardized measures of literacy and phonological processing across a wide range of behavioral abilities.

To compare cortical speech envelope representations to measures of literacy and phonological processing across subjects, we quantified values for phase-locking precision, phase-locking timing, and RMS amplitude for each subject (see Data Analysis). We performed Pearson's correlations between these three measures and the standardized measures. Of particular interest was a measure of literacy described in a previous work, which is the average of reading, spelling and non-word reading (Banai et al., 2005a). Results indicate that all three measures of cortical function showed significant correlations with this measure of literacy (Fig. 14, top), and these relationships accounted for 17% - 30% of the variability in literacy scores. Correlations were also significant when these cortical measures were compared to individual measures of reading, spelling and non-word reading with the exception of RMS asymmetry and reading/non-word reading which showed a trend for significance ($p \leq 0.12$; $p \leq 0.05$ for all other correlations). Significant relationships were also found between 2 of the cortical measures, phase-locking precision and RMS amplitude sensitivity, and standardized measures of phonological processing ($r > 0.5$, $p < 0.01$ for cortical-phonological processing correlations; Fig. 14, bottom). When cortical measures and measures of literacy and phonological processing were correlated within controls and RIs, correlations were not statistically significant ($p > 0.05$).

Handedness and measures of speech envelope representation

Left-handed and ambidextrous individuals often have reversed cerebral dominance (i.e., right-hemisphere dominance) for language. It is not known how dominance for language is related to cerebral asymmetries for temporal information in acoustic signals, and we wanted to rule out handedness as a contributing factor to group differences described here. To this end, we performed all statistical analyses after removing the two control and two RIs who reported as

either left-handed or ambidextrous (see Supplemental Results section)**.** Results were identical to those described above with one exception: the correlation between literacy and RMS asymmetry failed to reach statistical significance ($p = 0.153$) despite the fact that this cortical measure predicted 59% of the variance in phonological processing scores across subjects ($p < 0.0001$). This can possibly be explained by reduced statistical power as a result of removing 4 subjects from the analyses.

Supplemental results: right-handed subjects

To rule out handedness as a contributing factor to the abnormal patterns of cortical representation described in RI subjects, we performed the exact same analyses described in the manuscript after removing the two control and two RI subjects who reported themselves as either left-handed or ambidextrous. The following are the statistical results from this analysis:

Speech envelope phase-locking precision: We performed three separate 2 x (3 x 2) mixed-model ANOVAs [group x electrode x hemisphere] with repeated-measures (RM) on the final two factors. The dependant variable in this analysis was the Z-value associated with the peak Pearson's "r" value for each subject from 3 paired electrode sites in response to each speech condition.

- In the clear speech condition, results confirmed that Pearson's "r" values were greater in the right-hemisphere (main effect of hemisphere: $F_{1,17} = 10.822$, $p = 0.004$) and were similar across groups ([hemisphere x group] interaction: $F_{1,17} = 0.024$, $p = 0.878$).

- In the conversational condition, the same rightward asymmetric pattern was seen in both groups (main effect of hemisphere: $F_{1,17} = 8.845$, $p = 0.010$; [hemisphere x group] interaction: $F_{1,17} = 0.887$, $p = 0.359$).
- In the compressed speech condition, phase-locking precision was represented symmetrically across the left and right hemispheres in RI subjects. ANOVA results failed to show a main effect of hemisphere ($F_{1,17} = 1.864$, $p = 0.190$) but did show a significant group difference (hemisphere x group interaction: $F_{1,17} = 6.850$, $p = 0.018$). *Post-hoc* analyses confirmed that control subjects continued to show right-dominant asymmetry for precision of phase-locking in the compressed speech condition (3 x 2 repeated-measures ANOVA [electrode x hemisphere] main effect of hemisphere: $F_{1,9} = 10.090$, $p = 0.011$). In contrast, within the RI group, r-values were statistically similar between right and left-hemisphere electrodes (main effect of hemisphere: $F_{1,8} = 0.625$, $p = 0.452$).

Speech envelope phase-locking timing: We performed a 2 x (3 x 3 x 2) mixed-model ANOVA [group x stimulus x electrode x hemisphere] with repeated-measures on the final three factors with lags as the dependent variable.

- Results indicated a significant interaction of hemisphere x group ($F_{1,17} = 8.004$, $p= 0.012$
- There was no main effect of stimulus condition ($F_{2,34} = 1.496$, $p= 0.239$) or hemisphere x stimulus x group interaction ($F_{2,34} = 0.022$, $p= 0.978$)
- *Post-hoc* 2 x 3 mixed-model ANOVA [group vs. right-hemi electrodes] with repeated-measures on the final factor showed that the initial interaction was driven by increased

lags in right-hemisphere electrodes in the RI group compared to controls ($F_{1,55} = 5.077$, $p = 0.028$).

Onset period RMS amplitude analysis: We performed 2 x (3 x 3 x 2) mixed-model ANOVA [group x stimulus x electrode x hemisphere] with repeated-measures on the final three factors using onset RMS as the dependent variable.

- Results indicated no main effect of group and interactions involving group: $F \leq 1.588$, $p > 0.225$ for all group-related results).

Envelope-following period RMS amplitude analysis: We performed three separate 2 x (3 x 2) mixed-model ANOVAs [group x electrode x hemisphere] with repeated-measures on the final two factors in which the dependant variable was the envelope-following RMS amplitude measured in each stimulus condition.

- Results indicated that right dominant asymmetry was evident across subjects for the clear (main effect of hemisphere: $F_{1,17} = 5.585$, $p = 0.030$) condition and that this was similar between groups (group x hemisphere interaction: $F_{1,17} = 0.013$, $p = 0.910$).
- In the conversational condition, a main effect of hemisphere failed to reach statistical significance ($F_{1,17} = 2.089$, $p = 0.167$), and showed there was no group x hemisphere interaction ($F_{1,17} = 1.825$, $p = 0.194$). Within-group ANOVA analyses revealed that control subjects showed strong right-dominant asymmetry for RMS amplitude in the conversational condition (main effect of hemisphere: $F_{1,9} = 7.099$, $p = 0.026$) while the RI group did not show asymmetry in this condition (main effect of hemisphere: $F_{1,8} = 0.003$, $p = 0.959$).

- In the compressed condition, there was no main effect of hemisphere ($F_{1,17} = 2.515$, $p = 0.131$), and showed a trend for significance with respect to a group x hemisphere interaction ($F_{1,17} = 4.004$, $p = 0.062$). Within-group ANOVA analyses revealed that control subjects showed strong right-dominant asymmetry for RMS amplitude in the conversational condition (main effect of hemisphere: $F_{1,9} = 15.870$, $p = 0.003$) while the RI group did not show asymmetry in this condition (main effect of hemisphere: $F_{1,8} = 0.050$, $p = 0.829$).

Cortical representation of the speech envelope and measures of literacy and phonological processing*:* To compare cortical speech envelope representations to measures of literacy and phonological processing across all subject, we quantified values for phase-locking precision, phase-locking timing, and RMS amplitude for each subject (see Data Analysis). We performed Pearson's correlations between these three measures and standardized measures of literacy and phonological processing.

Pearson's correlations between cortical measures and literacy:

- Phase-locking asymmetry: $R = 0.524$, $p = 0.021$
- Phase-locking lag: $R = -0.455$, $p = 0.050$
- RMS asymmetry: $R = 0.341$, $p = 0.153$

Pearson's correlations between cortical measures and phonological processing:

- Phase-locking asymmetry: $R = 0.606$, $p = 0.006$
- Phase-locking lag: $R = -0.073$, $p = 0.767$

- RMS asymmetry: R = 0.768, p = 0.0001

Discussion

Summary

We tested the prediction that RI children, a group that has demonstrated abnormal perception for slow acoustic features (Goswami et al., 2002; Witton et al., 2002), would exhibit abnormal patterns of cerebral asymmetry in response to slow temporal features in speech. Across three speech conditions, controls showed consistent, right-dominant asymmetry in auditory cortex for multiple aspects of speech envelope representation. RIs indicated right-dominant asymmetry for phase-locking precision and RMS amplitude in clear and conversational speech conditions, however cortical responses were symmetric across the cerebral hemispheres in an acoustically challenging stimulus condition, compressed speech. Additionally, right-hemisphere cortical representations were delayed in RIs across all three stimulus conditions compared to controls. All three measures of cortical speech envelope representation correlated (r^2 = 0.17 - 0.30) with a standardized measure of literacy.

The AST hypothesis and RI

Speech is a highly complex acoustic signal that contains information on a number of time scales. Results from the current study add to an emerging body of literature describing a neural mechanism for processing these different temporal features in the human auditory system. This processing scheme is described in the "asymmetric sampling in time" (AST) hypothesis, and states that slow temporal features (3-5 Hz) in acoustic signals lateralize to right-hemisphere

auditory areas while rapid temporal features (20-50 Hz) lateralize to the left (Poeppel, 2003). Rates specified by this hypothesis are critical for speech perception: slow rates provide syllable pattern information and are essential for normal speech perception (Drullman et al., 1994b); fast rates correspond to the rate of temporal modulations that characterize many phonemic contrasts (e.g., formant transitions, voice onset time).

Prior research investigating the AST hypothesis has focused on the unimpaired auditory system and has shown that asymmetric processing of temporal information applies to both speech (Liegeois-Chauvel et al., 1999; Abrams et al., in press) and non-speech (Belin et al., 1998; Boemio et al., 2005) stimuli. Current results are the first to test the AST hypothesis in an impaired population, and results from RIs support this hypothesis under "good" listening conditions, represented here as the clear and conversational speech conditions. With respect to the AST hypothesis, this finding is important because it provides evidence that this temporal processing mechanism is a general and robust property of the human auditory system that is even present in a population associated with speech perception deficits (Kraus et al., 1996; Bradlow et al., 2003). Furthermore, normal patterns of asymmetry in RI in two speech conditions indicate that this mechanism can exist and function in the absence of a robust phonological system, a result that suggests that, in good listening conditions, this mechanism may not be strongly influenced by up-stream phonological processing mechanisms described in hierarchical models of speech perception (Hickok and Poeppel, 2007). There are still many important questions regarding this temporal processing mechanism that remain, including an examination of the functional independence of the rapid and slow temporal processors. Do left and right-hemisphere temporal processors represent separable neural mechanisms or a single mechanism

whose component processors function in conjunction with one another? Does rightward asymmetry for slow temporal features relate to the right hemisphere's preference for spectral processing (Zatorre et al., 2002)? It is hoped that future studies can identify the acoustic, linguistic and cognitive factors that play a role in this temporal processing mechanism.

Evidence for impaired speech envelope processing in RI

A major finding of this study was that the normal pattern of right-dominant activation of speech envelope representation was symmetric in RI for the compressed speech condition. Abnormal patterns of asymmetry have long been associated with RI, including abnormal manifestations of cerebral dominance with respect to handedness (Galaburda et al., 1985). A recent study provided evidence that RI was also associated with abnormal cortical asymmetry in response to rapid acoustic features in speech (Abrams et al., 2006). In that study, normal, left-dominant asymmetry in response to a consonant-vowel stimulus with a rapid formant transition was represented symmetrically in left and right auditory cortex in many RIs. In conjunction with the current results, these data suggest that both the rapid and slow mechanisms specified in the AST hypothesis are impaired in some RIs, and that abnormal function of both of these processors results in recruitment of the contralateral auditory cortex, rather than the "preferred" hemisphere.

An interesting observation across studies is that fast and slow acoustic processors appeared to "break down" in different manners in RI. Results from the current study are consistent with the possibility that the slow temporal processor in RIs is only impaired at the upper-end of the right-hemisphere's operating range, defined as 3-5 Hz (Poeppel, 2003). Patterns of asymmetry were only impaired in RIs when speech envelope frequencies were increased to a modal frequency of

4.3 Hz (Ahissar et al., 2001b) (see Methods) while slower envelope rates characteristic of clear and conversational stimuli appeared to elicit normal, right-dominant function. In contrast, RIs' left-hemisphere rapid acoustic processor (Abrams et al., 2006) may be impaired across its entire operating range: the formant transition in the consonant-vowel stimulus that elicited abnormal patterns of asymmetry was 40 msec in duration (25 Hz), a rate that corresponds to the lower-end of the proposed processing range for the left-hemisphere (20-50 Hz) (Poeppel, 2003). It is hoped that future studies will better delineate the processing range of the left- and right-hemispheres in RIs, and the relationship between these ranges and perception.

RIs showed delayed cortical response timing for speech envelope representations across all three stimulus conditions. Delayed cortical evoked-potentials in response to brief speech sound stimuli have been demonstrated in previous studies of RI (Cunningham et al., 2000) but these neurophysiological findings have not been directly supported in the behavioral literature since delayed reaction time in auditory tasks has typically not been associated with RI (Nicholson and Fawcett, 1994). A recent psychophysical study showed that dyslexic response accuracy was comparable to controls while response timing was significantly delayed in dyslexics (Chait et al., 2007). The authors concluded that reading impairments are associated with a general timing delay with respect to tracking acoustic change, while the accuracy of auditory representation is unimpaired until acoustic changes occur within a relatively short time window. The current neurophysiological results support these conclusions and together provide a framework for considering slow temporal processing impairments in RI.

Evidence for normal speech envelope processing in RI

RI speech envelope phase-locking and RMS amplitude showed the normal, right-dominant pattern of activation in clear and conversational speech conditions, and these results provide evidence regarding the nature of the RI auditory system. This suggests that the physiology underlying the RI auditory system is not categorically different from the unimpaired auditory system and that only relatively extreme acoustical conditions (e.g., compressed speech) disrupt cortical representations. This physiologic result reflects the widely-accepted notion that RI speech perception is not radically impaired and is characterized as a subtle deficit: previous studies have shown that RI speech perception is distinguished from unimpaired subjects primarily under difficult listening conditions (Watson et al., 1990; Bradlow et al., 2003; Ziegler et al., 2005). A separate observation is that the presence of normal neurophysiological responses in RI for two speech conditions makes it unlikely that a deficit in auditory attention, attentional state, or the specific methods used for data collection (i.e., watching a movie with simultaneous audio track) played a role in group differences for the compressed speech condition since speech conditions were randomly interleaved during data collection.

Theories of reading impairment

Despite decades of research, the neurobiological foundation of RI remains elusive. While there is near-universal agreement that many RIs suffer from a phonological deficit, whether or not phonological deficits are secondary to a more fundamental sensory deficit remains a source of debate (Rosen, 2003; Bishop, 2006; Goswami, 2006; White et al., 2006). Those who believe that phonological deficits are not causally related to sensory impairments would argue that abnormal neurophysiologic representation of the speech envelope and impaired perception for slow, non-speech signals (Goswami, 2002; Witton et al., 2002) share a common underlying

biological factor with impaired reading and phonology, but are otherwise unrelated phenomenon (White et al., 2006).

Alternatively, those who view sensory deficits as an underlying factor to phonological impairments in RI (Stein and Walsh, 1997; Tallal et al., 1998) would view the current results as evidence that abnormal auditory representations of temporal elements of speech are more pervasive than previously thought in this population. Specifically, current results provide strong evidence that acoustic deficits in RI are not isolated to rapid components of acoustic signals (Tallal et al., 1998). While abnormal representation of the speech envelope contradicts the temporal specificity of Tallal's hypothesis, this finding is consistent with the essential aspect of her hypothesis which states that RIs are impaired in temporal acoustic processing that contributes to phoneme discrimination. Both time scales are involved in phonemic contrasts. While segmental speech cues are typically associated with rapid temporal modulations (e.g., formant transitions, voice onset time), slower features in speech provide segmental cues as well, especially in manner of articulation, voicing and vowel identity (Rosen, 1992). Given that both rapid and slow aspects of the speech signal contribute to phoneme discrimination, a disruption in the neural processing of either of these aspects of the signal could contribute to the development of impaired phonological representations necessary for normal reading acquisition. Consequently, it is argued that the current results call for an expansion of the temporal processing hypothesis of reading-disabilities to encompass impairments for both rapid and slow acoustic features.

A limitation of the current work is that correlations linking impaired cortical representations of speech and reading ability cannot address causal relationships between these factors. Nevertheless, the current data are consistent with a causal link: objective measures of cortical speech envelope representation correlated with measures of literacy and phonological processing (Fig. 14). If speech envelope representation is unrelated to reading ability, results would indicate that the putative "common biological factor" that accounts for these two phenomena varies in severity along a continuum for all readers, and that both envelope processing and phonology are affected similarly at points across this continuum, but are still unrelated. An arguably more convincing hypothesis is that temporal processing abilities vary along a continuum and affect phonological abilities and reading in a proportional manner, and representations measured in the current study reflect an important aspect of speech processing.

Results indicate that not all RIs show abnormal cortical representation of the speech envelope (Fig. 9). Proponents of a non-sensory-based phonological deficit have argued that the failure for all RIs to exhibit auditory deficits indicates that these two factors cannot be related (White et al., 2006). Meanwhile, those who believe that auditory deficits contribute to RI have not generally acknowledged that many RIs do not suffer from auditory deficits (Tallal et al., 1998). Perhaps it is the case that the deficits underlying reading impairments are as heterogeneous as the collection of sensory, cognitive and mnemonic mechanisms required for normal reading acquisition. If so, auditory impairments such as those demonstrated here may only affect a sub-population of RIs, and may serve to exacerbate this subgroup's phonological deficits (Bishop, 2006).

Implications for remediation of RI

Much attention has been focused on how to remediate phonological disorders in RI, and success has been shown when training paradigms require subjects to practice discrimination of phoneme contrasts (Merzenich et al., 1996; Moore et al., 2005). A recent study showed that training on phoneme contrasts resulted in significant, long-lasting learning that generalized to measures of phonological awareness (Moore et al., 2005). Given that listeners use slow temporal information in speech to help identify phonemes (Rosen, 1992) and RIs show an impairment in the representation of these cues, additional benefit in phonological measures could potentially be achieved if training also involves discrimination of phonemes based on the speech envelope (Shannon et al., 1995).

Acknowledgements: We thank the children who participated in this study and their families. This work is supported by the National Institutes of Health grant R01 DC01510-10 and National Organization for Hearing Research grant 340-B208.

CHAPTER IV: AUDITORY BRAINSTEM TIMING PREDICTS CEREBRAL ASYMMETRY FOR SPEECH

Published in *Journal of Neuroscience,* 26(43):11131-11137

Daniel A. Abrams[1], Trent Nicol[1], Steven Zecker[1], and Nina Kraus[1,2]

[1]Auditory Neuroscience Laboratory, Department of Communication Sciences, and [2]Departments of Neurobiology and Physiology and Otolaryngology. Northwestern University, 2240 Campus Drive, Evanston, Illinois, 60208.

Abstract

The left hemisphere of the human cerebral cortex is dominant for processing rapid acoustic stimuli, including speech, and this specialized activity is preceded by processing in the auditory brainstem. It is not known to what extent the integrity of brainstem encoding of speech impacts patterns of asymmetry at cortex. Here we demonstrate that the precision of temporal encoding of speech in auditory brainstem predicts cerebral asymmetry for speech sounds measured in a group of children spanning a range of language skills. Results provide strong evidence that timing deficits measured at the auditory brainstem negatively impact rapid acoustic processing by specialized structures of cortex, and demonstrate a delicate relationship between cortical activation patterns and the temporal integrity of cortical input.

Keywords: auditory brainstem; auditory cortex; reading; dyslexia; cerebral asymmetry; speech

Introduction

The left hemisphere of the cerebral cortex is dominant in the processing of speech, and multiple lines of evidence have demonstrated a general preference of left auditory regions for the processing of rapid acoustic signals (Liegeois-Chauvel et al., 1999; Belin et al., 2000; Zatorre and Belin, 2001). Activation of specialized structures in the left hemisphere is preceded by processing in the auditory brainstem, a series of nuclei that receive input from the acoustic nerve and transmit this signal to the cortex via auditory thalamus. It is not known to what extent the integrity of brainstem encoding of speech is related to patterns of asymmetry at cortex. Here we demonstrate a correlation between the precision of temporal encoding of speech in auditory brainstem and cerebral asymmetry for speech sounds.

The auditory brainstem response (ABR) reflects neurophysiologic activity from synchronous neuronal ensembles in rostral and posterior brainstem structures. The ABR has emerged as an experimental tool to assess the integrity of brainstem processing of speech and other complex stimuli in normal and impaired populations (Kraus and Nicol, 2005). Speech-evoked ABRs represent temporal features of speech stimuli with great fidelity and delays in the response on the order of fractions of milliseconds have been linked to abnormal perception and linguistic abilities.

Auditory-evoked cortical responses reflect the summation of excitatory post-synaptic potentials originating from structures located primarily in the temporal lobe. Like ABRs, cortical potentials rely on stimulus-locked, synchronous firing from neuronal ensembles, however they provide an abstract representation of features in acoustic stimuli. The early components of cortical responses (<150 ms) reflect obligatory acoustic processing of speech stimuli (Sharma et al., 2000) and left-asymmetric responses in this time range is thought to describe its preference

for processing of rapid acoustic signals, including speech (Liegeois-Chauvel et al., 1999; Bellis et al., 2000).

To investigate a correspondence between brainstem encoding of speech and patterns of asymmetry at cortex, speech-evoked ABRs and cortical asymmetry for speech were evaluated in a group of children spanning a range in language skills, including children with language-based learning disabilities (LD). LD children were included in this study to provide the larger group with a wide range of neurophysiologic profiles: LDs have long been associated with abnormal cerebral asymmetry (Morgan, 1896), and more recently have demonstrated deficient encoding of speech-sounds in the auditory brainstem (Cunningham et al., 2001b; Banai et al., 2005a; Wible et al., 2005). We also tested subjects on behavioral measures, including speech-sound perception and tests of academic achievement, to assess a potential relationship between behavior and cortical asymmetry.

Methods

The research protocol was approved by the Institutional Review Board of Northwestern University. Parental consent and the child's assent were obtained for all evaluation procedures and children were paid for their participation in the study.

Children were acclimated to the testing circumstances prior to experimental data collection. They were allowed to visit the laboratory and interact with the tester on multiple occasions. Some children brought an electrode home with them to better familiarize themselves with the neurophysiological procedure.

Participants

All children were between 8-12 years old, reported no history of neurological or otological disease and were of normal intelligence (scores >85 on the Brief Cognitive Scale) (Woodcock and Johnson, 1977). In addition, all children had normal pure-tone hearing thresholds and click-evoked ABRs. The grouping of subjects in this work is based entirely on brainstem or cortical physiologic measures, however Normal (NL) and LD children are briefly described here, with group statistics provided in Table 3. LD children (n=30) were age-matched to NLs (n=37). Children with learning problems consisted of children who had been formally identified as such by an independent pyschoeducational diagnostician. Inspection of Table 3 reveals that the mean reading and spelling scores for LD subjects are below average (average is a score of 100 for these standardized test scores), but are not in the "impaired" range, defined as a score of <85. In addition, mean reading and spelling scores of NLs are considered above average. Nevertheless, the normal group differed significantly from the LD group on measures of auditory processing (Woodcock and Johnson, 1989) (a composite score derived from the Incomplete Words and Sound Blending subtests), single-word reading and spelling ($P<0.001$ for these three tests) (Wilkinson, 1993) and a measure of speech-sound discrimination ($P<0.03$) (Carrell et al., 1999).

Recording procedure

The procedures to measure brainstem and cortical responses were identical to those that have been described (Russo et al., 2004; Warrier et al., 2004). Brainstem and cortical responses were measured during different sessions. Brainstem responses were differentially recorded at a sampling rate of 20kHz using a vertex electrode referenced to the right earlobe. The forehead

served as ground. Three blocks of 1000 repetitions were collected at each polarity. For cortical responses, recording electrodes were placed at the vertex and over left and right temporal lobes; TL was located halfway between electrode sites T3 and T5 according to the international ten-twenty system (Jasper, 1958), and TR was located halfway between T4 and T6. The nose served as the reference electrode and the forehead served as ground. Cortical responses were sampled at 2kHz and 1000 repetitions were collected. For both brainstem and cortical recordings, speech sounds were presented to the right ear at 80 dB sound pressure level (SPL) through insert earphones. The inter-stimulus interval was 51 msec for brainstem responses and 590 msec for cortical responses. Since the side of stimulation was held constant for all subjects (right ear), we reasoned that subject differences in cortical asymmetry could not be attributable to stimulus delivery issues. The stimulus used to evoke brainstem and cortical responses was the speech syllable /da/ synthesized at a sampling rate of 10 kHz. The stimulus was 40 msec in duration and consisted of five formants with an onset burst during the first 10 ms at F3, F4, and F5.

Data analysis

Brainstem responses to the speech sound /da/ have been described in previous reports (Cunningham et al., 2001b; King et al., 2002; Russo et al., 2004; Wible et al., 2004; Banai et al., 2005a; Johnson et al., 2005; Russo et al., 2005; Wible et al., 2005) are extremely reliable between and within subjects (Russo et al., 2004). Amplitudes and latencies for brainstem onset (peaks V and A), offset (peak O) and fundamental frequency following (peaks D, E and F) were identified for each subject. To enable comparisons between peak latencies, Z-scores for all peak latencies were calculated. To prevent spurious results from regression analyses, which can be overly biased by outlying data points, all raw (i.e., prior to Z-score calculation) brainstem peak

and cortical asymmetry values beyond 2 standard deviations (SD) of the mean were moved to the 2 SD point for that particular measure. Across all brainstem peak and cortical asymmetry measures, 19 data points (out of a total of 455) were moved to the 2 SD point. Brainstem responses in 1 normal and 1 LD subject did not indicate a clear peak D and peak O, respectively. Since ANOVA and regression analyses explicitly relied on latency and amplitude data for all brainstem peaks, these subjects were omitted. Detailed descriptions of all analyses are provided throughout the Results section and are not repeated here.

Results

Figure 15 (bottom) shows the /da/ stimulus waveform and three overlaid brainstem responses. Common among these three responses is that temporal features of the stimulus are represented in brainstem responses: stimulus onset is evidenced by a large positive-negative peak complex at ~8.5 ms following stimulus onset (peaks V and A); phase-locking to the fundamental frequency of the stimulus is represented by negative peaks between 20-45 ms (peaks D, E, F); the offset of the stimulus is represented by a negative peak at ~49 ms (peak O). The presence of these particular peaks is extremely consistent between subjects and has been described in previous reports (King et al., 2002; Russo et al., 2004; Wible et al., 2004; Banai et al., 2005a; Johnson et al., 2005). It has been proposed that peaks representing acoustic transients in the stimulus (peaks V, A and O) are served by distinct neural mechanisms relative to those representing steady-state aspects of the stimulus, like the fundamental frequency (peaks D, E and F) (Kraus and Nicol, 2005).

Irrespective of the stereotyped morphology of responses, slight variations in the timing of brainstem peak latencies are evident across subjects, and previous studies have indicated that delays in the brainstem's representation of acoustic transients, such as speech onset and offset, is related to literacy in school-aged children (Cunningham et al., 2001b; King et al., 2002; Wible et al., 2004; Banai et al., 2005b). We ranked subjects' brainstem responses according to the latency of onset and offset peaks by transforming absolute latencies of peaks V, A and O into Z-scores, then averaged these Z-score values for each subject. This provided a single score for a subject that represented the brainstem's composite transient (onset and offset) response to the speech-sound stimulus. We then divided the subjects into 5 groups in ascending order of brainstem latency and averaged the responses within each group. Dividing all the subjects into 5 groups for this initial analysis enabled large enough samples (n=13) to avoid a single subject overwhelming an average across responses while also enabling a reasonable gradient by which to assess cortical activation patterns as a function of brainstem onset/offset latency. The zoomed-in plots of peaks V, A and O in Figure 15 show the latency differences, which are on the order of tenths of milliseconds, between first ("Early"), third ("Middle") and fifth quintile ("Late") brainstem responses to transients. Table 4 lists the number of subjects, as well as the breakdown of normal and LD subjects, in the 5 groups categorized according to brainstem transient (onset/offset) latency.

For cortical responses, we calculated the global field power (GFP), a measure defined as the standard deviation across multiple channels as a function of time. Peaks in the GFP serve to isolate and identify auditory evoked potential components and reflect a maximum of the total underlying brain activity that contributes to the surface potential field (Lehmann and Skrandies,

1980). A dominant peak identified in the GFP for both normal and LD subjects occurred at ~140 ms, consistent with a pattern of cortical activation that has been described in previous reports (Wolpaw and Penry, 1975, 1977; Tonnquist-Uhlen et al., 2003). This GFP result enabled us to focus our investigation on auditory cortical responses at this latency.

To investigate cortical activation patterns associated with delayed brainstem timing, cortical responses were averaged within each of the 5 brainstem timing-defined groups. Figure 15 (top) shows a distinct relationship between the relative timing of brainstem peaks and early cortical responses: subjects in the Early brainstem timing group show a striking left-dominant cortical activation pattern while Middle and Late groups showed progressively more similar responses between temporal electrodes, with subjects in the Late group showing virtually symmetric responses. This pattern was most prominent at latencies centered at 140 ms (shaded region), the latency range identified in the GFP analysis. The large amplitude difference between left and right temporal electrodes seen in the Early brainstem group has been described in previous reports as an indicator of cortical response asymmetry (Naatanen et al., 1997; Bellis et al., 2000) and provides evidence for diminished cortical asymmetry for speech sounds in children with delayed brainstem responses. One-way ANOVA statistics comparing individual subjects' cortical asymmetry, defined as the mean amplitude difference between TL and TR electrodes from 130-145 ms, for the three brainstem-defined groups displayed in Figure 15 (top) indicated a significant effect of brainstem timing on cortical asymmetry ($F_{2,36}$=4.805, p=0.014). *Post hoc* Tukey HSD comparisons revealed that the Early ABR group exhibited greater asymmetry than the Late group (p=0.013), however the Middle group was not statistically different from either the Early or Late group (P>0.05).

To further investigate the relationship between the timing of brainstem response onset/offset and cortical asymmetry, we evaluated this phenomenon across all subjects. To assign an asymmetry value to each subject, we calculated the mean amplitude difference between TL and TR electrodes for all subjects from 130-145 ms, producing a range of numbers from -150 μV through +50 μV with the most negative values indicating responses with the greatest leftward dominance. To quantify the relationship between brainstem timing for transients and cortical asymmetry, we performed a regression analysis on subjects' average brainstem onset/offset peak latency Z-score and asymmetry values. This relationship is displayed in Figure 16A, and results indicate that the timing of onset/offset transients in the brainstem response predicts the degree of cerebral asymmetry across subjects ($F_{1,63} = 10.365$, P=0.002, r^2=0.14). This relationship is markedly stronger when the 3 data points outside of the 95% confidence interval are removed from the regression analysis ($F_{1,60} = 15.947$, P=0.0002, r^2=0.21) (note that two of the data points outside of the 95% confidence interval fall on top of one another, making it appear that only 2 data points are outside this interval). The same relationship was found when the LD group was analyzed separately from the normal group ($F_{1,34} = 5.376$, P=0.027, r^2=0.14) and was marginally significant due to an outlying data point when the normal group was analyzed separately (with outlier: $F_{1,27} = 4.065$, P=0.054, r^2=0.13; without outlier: $F_{1,26} = 6.510$, P=0.017, r^2=0.20). The within-group results indicate that the relationship between brainstem timing and cortical asymmetry is a general property of the central auditory system irrespective of diagnostic category. Contrary to findings with respect to onset/offset peaks in the brainstem response, no relationship was seen between fundamental frequency phase-locking, defined as the average Z-transformed latency of peaks D, E and F, and cortical asymmetry (Figure 16B). Regression of

the timing of fundamental frequency phase-locking on cerebral asymmetry was neither significant when analyzed across all subjects ($F_{1,63}$= 0.936, P=0.337, r^2=0.01, not significant) nor when analyzed separately in the normal and LD groups (normal group only: $F_{1,28}$=0.147, P=0.704, r^2=0.005, not significant; LD only: $F_{1,34}$=0.935, P=0.34, r^2=0.027, not significant). Moreover, no relationship was seen between cortical asymmetry and any measure of brainstem peak amplitude or frequency-domain spectrum (P>0.15 for all measures, not significant). Consistent with previous reports, these data demonstrate dissociation between onset/offset and frequency following components of the brainstem response, further suggesting separate mechanisms for these features of the response (Kraus and Nicol, 2005).

It is not known to what extent abnormal brain processing of brief speech sound stimuli may be related to behavioral deficits on tasks that rely on normal auditory function. We reasoned that if the current measure of cortical asymmetry reflected temporal processing of the speech signal, then individuals with abnormal patterns of asymmetry should reveal deficits on two types of behavioral tasks: tasks that directly measure rapid acoustic processing (Johnsrude et al., 1997; Belin et al., 1998; Joanisse and Gati, 2003; Zaehle et al., 2004) such as speech sound discrimination, and linguistic measures, such as reading and phonological processing, which are thought to be negatively affected as a consequence of deficient acoustic processing (Tallal et al., 1993). To test this hypothesis, we first divided the combined normal and LD group into quintiles based on strength of cortical asymmetry (see Table 5). Then we compared individual subjects' scores of speech perception and academic achievement for those subjects with the strongest and weakest leftward asymmetry, determined by the bottom and top asymmetry quintiles for the combined normal and LD group (n=14; see Fig. 7). Results indicated that strong left-dominant

subjects performed better in discriminating the speech syllable /da/ from /ga/, a contrast that requires precise acoustic processing of rapid frequency transitions (t=2.575, P=0.016). Discrimination between the speech syllables /ba/-/wa/, a contrast that varies in the duration of the formant transition, serves as a task control to /da/-/ga/ discrimination (Kraus et al., 1996) which varies in the spectral content of the formant transition. There was no difference between strong and weak left-dominant subjects on the /ba/-/wa/ discrimination task (t=0.511, P>0.50, not significant). The strong left-dominant subjects also performed better on tests of phonological processing and spelling, and results from a single-word reading test narrowly missed statistical significance (Auditory Processing: t=2.106, P=0.045; spelling: t=2.608, P=0.015; single-word reading t=1.897, P=0.068). The majority of strong (n=12) and weak (n=13) left-dominant subjects were also tested on non-word reading, another important measure of phonological ability, and again strong left-dominant subjects performed better (t=2.366, P=0.027). A consideration is that the strong left-dominant subject group consists primarily of normal subjects while the weak left-dominant group has primarily LDs (see Table 5), which explains the discrepancy in the behavioral scores between these groups. Nevertheless, these data suggest that the current measure of cortical asymmetry reflects essential temporal processing of the speech stimulus, and that abnormal acoustic encoding of speech sounds by left-hemisphere auditory areas may contribute to reading deficits.

Discussion

In summary, we have shown that auditory brainstem timing of speech onset and offset is correlated to a measure of cortical asymmetry across, and within, normal and LD subjects. In addition, it was shown that subjects with greater left-asymmetric activation patterns were better

than subjects with weak asymmetry on a number of tasks that rely on phonological processing, including reading and spelling. Taken together, these data indicate a specific relationship between temporal acuity in the auditory brainstem and cerebral asymmetry for speech-sounds associated with phonological processing and reading ability.

Brainstem and cortical processing of speech sounds

The relationship between brainstem synchrony and cerebral asymmetry for speech sounds adds to a growing literature linking the auditory brainstem with cortical processing of speech. Wible et al. (Wible et al., 2005) demonstrated a strong correlation between synchronous onset timing of the speech-evoked ABR and the ability of cortical responses to maintain their representation of speech in the presence of background noise, suggesting that brainstem synchrony is related to the robustness of cortical representations. In another study, (Banai et al., 2005a) showed that asynchronous onset timing in the auditory brainstem was related to poor cortical sensitivity to acoustic change, measured in a group of normal and LD children. In conjunction with the findings described here, these data provide converging evidence that cortical function is closely related to brainstem timing for speech sounds.

An important consideration for the current data is the presence of temporal processing abnormalities across multiple levels of the auditory system. The dynamics of this system is an exciting topic, and while the results reported here do not prove causality between brainstem and cortical processing of speech sounds, the three following scenarios could account for their relationship. One plausible scenario is that neural deficits at a lower (i.e., more peripheral) level

of the auditory pathway cause abnormal cortical activation patterns, a tempting possibility based on signal flow in the afferent pathway of the auditory system. A piece of evidence in favor of this hypothesis is that auditory brainstem responses reach maturity many years before auditory cortex (Inagaki et al., 1987; Ponton et al., 2000), indicating a peripheral-to-central hierarchy in the development of this system. To its detriment, this interpretation would be an oversimplification of an extremely complex system that includes parallel (Young, 1998; Kaas and Hackett, 2000) and top-down (Xiao and Suga, 2002; Perrot et al., 2005) processing from cortex. While it is certainly plausible that "normal" cortical activation patterns would not develop if brainstem nuclei were not able to properly represent this signal, it remains to be seen if this is what actually occurs in the disabled auditory system.

Alternatively, a "top-down" scenario in which abnormal cortical function causes poor neural synchrony in the brainstem, and possibly more peripheral stations in the auditory system (Xiao and Suga, 2002), is also a possibility. A compelling hypothesis from the visual system, called the reverse hierarchy theory, poses that high-level cortical areas are responsible for directing plasticity in lower cortical levels, such as V1 (Ahissar and Hochstein, 2004). While it is conceivable that similar forms of plasticity could extend to brainstem nuclei, the authors of this theory do not explicitly address this possibility. Indeed studies in anesthetized bat have shown that cortical stimulation can drive plasticity in the inferior colliculus (Ma and Suga, 2001), the primary auditory nucleus of the midbrain. Moreover, it has been shown that auditory training (Russo et al., 2005) and language experience (Krishnan et al., 2005) impact brainstem encoding of acoustic elements of speech, which suggests that cortical mechanisms associated with directed attention and language can improve subcortical sensory encoding. While these pieces of

evidence describe particular top-down relationships between the cortex and brainstem, neither of them speaks directly to the etiology of abnormal acoustic processing in these two parts of the brain.

A third possible scenario to describe the dynamics of abnormal brainstem-cortical function lies somewhere between the previous two possibilities: abnormal function is truly systemic in nature, owing neither to aberrant brainstem or cortical function in particular. In this scenario, abnormal brainstem-cortical function represents a general failure of the system, possibly due to asynchronous activation patterns between the two auditory regions. A tenet of neuroscience proposed by Donald Hebb states that neurons that are active at the same time are mutually strengthened, and conversely, neural connections are weakened when the constituent neurons are activated asynchronously (Hebb, 1949). In the event that normal brainstem and cortical function are mutually reliant on synchronous activation between one another, then the current data could be explained as a general signaling disorder between these two regions of the brain. More work is required to specifically address the etiology of abnormal auditory function in the brainstem and cortex.

Feature dissociation in the auditory brainstem

The speech-evoked ABR provides discrete representations of many aspects of the acoustic structure of speech (Russo et al., 2004), including separate neural representations of speech-sound onset, phase-locking to the fundamental and formant frequencies and speech-sound offset. As mentioned previously, the current work adds to a growing body of evidence linking brainstem and cortical processing of speech sounds. A common thread among all of these studies is that

the portion of the ABR reflecting speech-sound onset is the structural feature that is associated with these various forms of cortical processing, in this case patterns of cortical asymmetry. The specificity of the brainstem's onset response with regards to cortical processing of speech reinforces the notion that features of the brainstem response are functionally dissociated from one another. This finding is consistent with a meta-analysis of a number of speech ABR experiments that showed a pattern in the dissociation of speech-sound representations in the auditory brainstem (Kraus and Nicol, 2005). Specifically, it was argued that auditory brainstem representations of the "source" of speech, which is generated by the vocal folds (i.e., the fundamental frequency), is differentiated from brainstem representations of features of speech introduced by the rest of the vocal tract, which serves as an acoustic "filter" (i.e., onset transients and formant structure). Future studies in both humans and animal models (King et al., 1999) may be able to better characterize the dissociation of the brainstem's representations of acoustic features in speech.

Components of cerebral asymmetry

A known limitation of this work is that stimulus presentation was provided to the right ear for all subjects, yet it is well-established that there is a stronger contralateral than ipsilateral cortical response independent of left-hemisphere asymmetries for speech sounds. Therefore, a reduction in the size of the asymmetry could be due either to (a) a reduction in the contralaterality of the right ear - left hemisphere response, or (b) a change in the degree of left-hemisphere specialization. It is argued that since all subjects were tested identically, it is reasonable to assume that effects of cerebral asymmetry are due to known specialized processing of the left-hemisphere for the speech signal. Nevertheless, while the current data clearly demonstrate a

general effect of cerebral asymmetry, future studies are needed to delineate the relative contributions of the contralateral response and the left-hemisphere specialization for acoustic features of speech.

Conclusion

Cerebral dominance is a fundamental organizing principle of the nervous system. It is hypothesized that the left-hemisphere's preference for rapid acoustic signals (Schwartz and Tallal, 1980; Belin et al., 1998) underlies cerebral dominance for language, and serves as a mechanism for the remarkable translation of acoustic speech information into complex linguistic constructs. Normal auditory function relies on the integrity of many auditory nuclei in the ascending pathway between the cochlea and the cortex, and findings described here demonstrate a relationship between extremely brief (~0.05 ms) delays at the brainstem and cortical activation patterns. Future studies addressing functional relationships between cortex and more peripheral loci in the auditory system will enable a more comprehensive understanding of the normal auditory system, and may serve to remediate auditory deficits in populations in which hearing function is of interest, such as individuals with learning disabilities, autism and the elderly.

Acknowledgements

Acknowledgements: We thank members of the Auditory Neuroscience Lab for help with data collection and analysis and C. Warrier and K. Banai for critical reviews of this manuscript. We also thank the children who participated in this study and their families. This work is supported

by the National Institutes of Health grant R01 DC01510-10 and National Organization for Hearing Research grant 340-B208.

CHAPTER V: AUDITORY BRAINSTEM TIMING IS NOT RELATED TO CORTICAL PROCESSING OF THE SPEECH ENVELOPE

Manuscript in preparation

Daniel A. Abrams[1], Trent Nicol[1], Steven Zecker[1], and Nina Kraus[1,2]

[1]Auditory Neuroscience Laboratory, Department of Communication Sciences, and [2]Departments of Neurobiology and Physiology and Otolaryngology. Northwestern University, 2240 Campus Drive, Evanston, Illinois, 60208.

Abstract

The human auditory system consists of a complex network of sub-cortical and cortical nuclei, and there is great interest in describing functional relationships among nuclei of this system. Recent work in the human auditory system has shown that a mechanism for processing temporal information in speech is the asymmetric routing of this information between the cerebral hemispheres: rapid features in speech are lateralized to left-hemisphere auditory areas while the slow features in speech, known as the speech envelope, are lateralized to right-hemisphere auditory areas. A recent study showed that temporal acuity in the auditory brainstem predicts left-dominant patterns of cortical asymmetry that reflect processing of rapid acoustic rate information. Here we investigate whether a similar relationship exists between brainstem

processing of speech and rightward cortical asymmetries for the slow temporal features in speech. To address this question, we measured brainstem and cortical responses to speech in a group of normal and reading0impaired children. Results indicate that there is no relationship between brainstem representation of speech and cortical representation of the speech envelope. These data suggest that a strong functional relationship is required between the auditory brainstem and left hemisphere of auditory cortex for the processing of rapid elements of speech signals but not for the processing of slower elements of speech.

Keywords: auditory cortex, brains asymmetries, speech, dyslexia

Introduction

The central auditory system consists of a highly complex network of sub-cortical and cortical nuclei characterized by an intricate pattern of connectivity between nuclei (Kaas and Hackett, 2000). An important step in acquiring a general understanding of this system is to describe functional relationships between constituent areas of the auditory system (Banai et al., 2005a; Wible et al., 2005).

Cerebral dominance represents a fundamental organizing principal of motor, sensory and cognitive systems (Geschwind and Galaburda, 1985b, a). In the auditory system, it has been shown that a mechanism for processing temporal information in speech is the asymmetric routing of this information between the cerebral hemispheres (Poeppel, 2003): rapid features in speech are lateralized to left-hemisphere auditory areas (Belin et al., 1998; Liegeois-Chauvel et al., 1999; Joanisse and Gati, 2003; Abrams et al., 2006) while the speech envelope is lateralized

to right-hemisphere auditory areas (Abrams et al., in press, submitted-a). The rates specified by this hypothesis are critical for speech perception: slow rates provide syllable pattern information and are essential for normal speech perception (Drullman et al., 1994b); fast rates correspond to the rate of temporal modulations that characterize many phonemic contrasts (e.g., formant transitions, voice onset time).

A recent study showed that temporal acuity in the auditory brainstem predicts left-dominant patterns of cortical asymmetry that reflect processing of rapid acoustic rate information (Abrams et al., 2006). Specifically, it was shown that brainstem responses to transient features (i.e., stimulus onset and offset) of a consonant-vowel stimulus predict the degree of left-hemisphere asymmetry for rapid acoustic processing. It is not known whether temporal acuity in the auditory brainstem also predicts right-dominant asymmetry that reflects slow temporal processing of acoustic signals. This is an important question for two reasons: first, results will provide important new information regarding the functional relationship between discrete structures in the ascending auditory system; second, it will address whether response timing in the auditory brainstem is important for all forms of temporal processing in auditory cortex, or alternatively whether brainstem timing is only important for processing rapid acoustic features, which may be more challenging for the auditory system to follow.

To investigate this question, we measured auditory evoked-potentials in 12 normally-developing and 11 children with RI in response to a variety of speech stimuli including a consonant-vowel stimulus that has been used to examine auditory brainstem function in many studies (King et al.,

2002; Wible et al., 2004; Johnson et al., 2005; Kraus and Nicol, 2005), and speech sentence stimuli that have been used to examine cortical representation of the speech envelope (Abrams et al., in press, submitted-a). LD children were included in this study to provide the larger group with a wide range of neurophysiologic profiles: LDs have long been associated with abnormal cerebral asymmetry (Morgan, 1896), and more recently have demonstrated deficient encoding of speech-sounds in the auditory brainstem (Cunningham et al., 2001b; King et al., 2002; Wible et al., 2004; Banai et al., 2005a). We then examined the relationship between established-measures of brainstem responses and cortical representations of the speech envelope.

Methods

The research protocol was approved by the Institutional Review Board of Northwestern University. Parental consent and the child's assent were obtained for all evaluation procedures and children were paid for their participation in the study.

Participants

Participants consisted of 23 children between 9-15 years old who reported no history of neurological or otological disease and were of normal intelligence (scores >85 on the Brief Cognitive Scale; Woodcock and Johnson, 1977). Children with reading impairments consisted of children who had been formally identified as such by an independent pyschoeducational diagnostician. Mean reading and spelling scores for RIs are below average (average is a score of 100 for these standardized test scores), but are not in the "impaired" range, defined as a score of <85 (Table 3). In addition, mean reading and spelling scores of NLs are considered above average. Nevertheless, controls differed significantly from RIs on measures of auditory

processing (Woodcock and Johnson, 1989) (a composite score derived from the Incomplete Words and Sound Blending subtests), single-word reading and spelling ($P<0.001$ for these three tests) (Wilkinson, 1993). Since cerebral dominance for language is often reversed in left-handed and ambidextrous individuals, and many of the analyses performed here relate to cerebral asymmetries, all subjects completed a questionnaire to assess handedness.

Children were recruited from a database compiled in an ongoing project entitled Listening, Learning and the Brain. Children who had previously participated in this project and had indicated interest in participating in additional studies were contacted via telephone. Brainstem and cortical responses were measured during different sessions.

Cortical protocol

Stimuli

Stimuli consisted of the sentence stimulus "The young boy left home" produced in three modes of speech: conversational, clear and compressed modes (Fig. 1). These three modes of speech have different speech envelope cues and were used as a means to elicit a variety of cortical activation patterns. Conversational speech is defined as speech produced in a natural and informal manner. Clear speech is a well-described mode of speech resulting from greater diction (Uchanski, 2005). Clear speech is naturally produced by speakers in noisy listening environments and enables greater speech intelligibility relative to conversational speech. There are many acoustic features that are thought to contribute to enhanced perception of clear speech relative to conversational speech, including greater intensity, slower speaking rate and more pauses. Most importantly with respect to the current work, an established feature of clear speech

is greater temporal envelope modulations at lower modal frequencies than conversational speech Modal frequency corresponds to the syllable rate of speech (1-4 Hz) (Krause and Braida, 2004). With respect to the particular stimuli used in the current study, greater amplitude envelope modulations are evident in the clear speech relative to the conversational stimuli. For example, there is no amplitude cue between "The" and "young" evident in the broadband conversational stimulus envelope, however an amplitude cue is present in the broadband clear stimulus envelope (Fig. 1, 0-450 msec). This phenomenon also occurs between the segments "boy" and "left" (450-900 msec). Compressed speech approximates rapidly-produced speech and is characterized by a higher modal frequency. Compressed speech is more difficult to perceive compared to conversational speech (Beasley et al., 1980) and has been used in a previous study investigating cortical phase-locking to the speech envelope (Ahissar et al., 2001b).

Conversational and clear sentences were recorded in a soundproof booth by an adult male speaker at a sampling rate of 16 kHz. Conversational and clear speech sentences were equated for overall duration to control for slower speaking rates in clear speech (Uchanski, 2005). This was achieved by compressing the clear sentence by 23% and expanding the conversational sentence by 23%. To generate the compressed sentence stimulus, we doubled the rate of the conversational sample using a signal-processing algorithm in Adobe Audition (Adobe Systems Inc.). This algorithm does not alter the pitch of the signal. The duration of the clear and conversational speech sentences was 1500 msec, and the duration of the compressed sentence was 750 msec.

Cortical recording and data processing procedures

All of the recording and data processing techniques used to describe cortical representation of the speech envelope are identical to those described in detail in a recent publication (Abrams et al., in press). A PC-based stimulus delivery system (Neuroscan GenTask, Compumedics, Inc) was used to output the sentence stimuli through a 16-bit converter at a sampling rate of 16 kHz. Speech stimuli were presented unilaterally to the right ear through insert earphones (Etymotic Research ER-2) at 80 dB SPL. Stimulus presentation was pseudorandomly interleaved. The polarity of each stimulus was reversed for half of the stimulus presentations to avoid stimulus artifacts in the cortical responses. Polarity reversal does not affect perception of speech samples (Sakaguchi et al., 2000). An interval of 1 second separated the presentation of sentence stimuli. Subjects were tested in a sound-treated booth and were instructed to ignore the sentences. To promote subject stillness during long recording sessions as well as diminish attention to the auditory stimuli, subjects watched a videotape movie of his or her choice and listened to the soundtrack to the movie in the non-test ear with the sound level set <40 dB SPL. This paradigm for measuring cortical evoked potentials has been used in previous studies investigating cortical asymmetry for speech sounds (Bellis et al., 2000; Abrams et al., 2006) as well as other forms of cortical speech processing (Kraus et al., 1996; Banai et al., 2005a; Wible et al., 2005). While it is acknowledged that cortical activity in response to a single stimulus presentation includes contributions from both the experimental speech stimulus and the movie soundtrack, auditory information in the movie soundtrack is highly variable throughout the recording session. Therefore, the averaging of auditory responses across 1000 stimulus presentations, which serves as an essential method for reducing the impact of noise on the desired evoked response, effectively removes contributions from the movie soundtrack. Cortical responses to speech stimuli were recorded with 31 tin electrodes affixed to an Electrocap (Electrocap International,

Inc.) brand cap (impedance <5 Kohm). Additional electrodes were placed on the earlobes and superior and outer canthus of the left eye. These act as the reference and eye blink monitor, respectively. Responses were collected (Neuroscan Acquire, Compumedics Inc) at a sampling rate of 500 Hz for a total of 1000 repetitions each for clear, conversational and compressed sentences.

Processing of the cortical responses consisted of the following steps. First, excessively noisy segments of the continuous file (typically associated with subject movement) were manually rejected. The continuous file was high-pass filtered at 1 Hz and removal of eye-blink artifacts was accomplished using the spatial filtering algorithm provided by Neuroscan. The continuous file was then low-pass filtered at 40 Hz to isolate cortical contributions and the auditory evoked potentials were then downsampled to a sampling rate of 200 Hz. All filtering was accomplished using zero phase-shift filters and downsampling was accompanied by IIR low-pass filtering to correct for aliasing (Compumedics USA, Inc). This goal of this filtering scheme was to match the frequency range of the speech envelope (Rosen, 1992). Responses were artifact rejected at a +/- 75 μV criterion. Responses were then subjected to noise reduction developed by our lab that has been used in improving the signal-to-noise ratio of brainstem and cortical evoked potentials (Abrams et al., in press). The theoretical basis for the noise reduction is that auditory evoked potentials are largely invariant across individual stimulus repetitions while the background noise is subject to variance across stimulus repetitions. Thus, the mean evoked response is significantly diminished by the fraction of repetitions that least resembles it. If these noisy responses are removed, the signal to noise ratio of the cortical response improves considerably with virtually no change to morphology of the average waveform. The algorithm calculated the

average response from all sweeps for each stimulus condition at each electrode then performed Pearson's correlations between each of the individual stimulus repetitions and the average response. The 30% of repetitions with the lowest Pearson's correlations from each stimulus condition were removed from subsequent analyses, and the remaining repetitions were averaged and re-referenced to a common reference computed across all electrodes. Therefore, following the noise reduction protocol, cortical responses from each subject represent the average of ~700 repetitions of each stimulus. Data processing resulted in an averaged response for 31 electrode sites and 3 stimulus conditions measured in all 23 subjects.

Cortical data processing: measures of cortical speech envelope processing

All of the analyses techniques used to describe cortical representation of the speech envelope are identical to those described in detail in a recent publication (Abrams et al., in press). All data analyses were performed using software written in Matlab (The Mathworks, Inc). Broadband stimulus envelopes were determined by performing a Hilbert transform on the broadband sentence waveforms (Drullman et al., 1994a). The resulting amplitude envelopes were low-pass filtered at 40 Hz and resampled at 200 Hz to isolate the speech envelope (Rosen, 1992) and to match the frequency characteristics and sampling rate of the processed responses. We calculated the frequency of maximal power, known as the modal frequency (Ahissar et al., 2001b), of the envelope of each speech sentence stimulus by performing a fast Fourier transforms of the low-pass filtered Hilbert envelope. FFTs were calculated using windows of 1 s and overlaps of 0.5 s, consistent with a previous report (Ahissar et al., 2001b).

Data are presented for 3 temporal electrode pairs: (1) T3 andT4, (2) T5 and T6 and (3) Tp7 and Tp8 according to the modified International 10-20 recording system (Jasper, 1958). The modification is the addition of the Tp7-Tp8 electrode pair in which Tp7 is located midway between T3 and T5, and Tp8 is located midway between T4 and T6. Two types of analyses were performed on the data: cross-correlation and RMS analyses, resulting in three measures of cortical speech-envelope representation. First, cross correlations between the broadband speech envelope and cortical responses at each temporal electrode for the "envelope-following period" (250-1500 msec for conversational and clear stimuli, 250-750 msec for the compressed stimulus) were performed using the "xcov" function in Matlab. The peak in the cross-correlation function was found at each electrode between 50-150 msec lags, resulting in the first two measures. (1) *Phase-locking precision* was defined as the peak r-value and (2) *phase-locking timing* was defined as the lag at the peak r-value. R-values were Fisher-transformed prior to statistical analysis. Finally, (3) *RMS amplitudes* at each electrode were calculated for 2 different time ranges: the "onset" period was defined by the time ranges 0-250 msec for all stimuli; the "envelope-following" period was defined as 250-1500 msec for conversational and clear stimuli and 250-750 msec for the compressed stimulus.

With respect to phase-locking precision, RIs were most clearly differentiated from control subjects based on the symmetry of their r-values in the compressed speech condition (see "Speech envelope phase-locking precision" in Results). Therefore, we first calculated mean left-hemisphere r-values (i.e., the mean of T3, T5 and Tp7) and right hemisphere r-values (i.e., the mean of T4, T6 and Tp8) from the cross-correlation analysis of the compressed speech condition, and entered these values into the asymmetry index (R – L) / (R + L) (Fig. 9). Collapsing across

electrode pairs was justified in this instance since there was no main effect of electrode, or interaction involving electrode, in a 2 x (3 x 2) mixed-model ANOVA [group x electrode x hemisphere] with repeated-measures on the final 2 factors in which compressed speech r-value was the dependant variable ($F < 2.4$, $p > 0.10$ for main effects and interactions). The asymmetry index values were used in subsequent comparisons to standardized measures of literacy and phonological processing.

With regards to phase-locking timing, RIs were differentiated from control subjects based on a general delay in right-hemisphere cortical lags that was evident across all speech conditions (see "Speech envelope phase-locking timing" in Results). Therefore, we first averaged right-hemisphere cortical lags across T4, T6 and Tp8 electrodes for each speech condition (Fig. 9). Collapsing across right-hemisphere electrode pairs was justified since there was no main effect of electrode pair or [electrode x group] interaction in a 2 x (3 x 3) [group x stimulus x electrode] mixed-model ANOVA with repeated-measures on the final two factors with right-hemisphere lags serving as the dependant variable (main effect of electrode: $F_{2,42} = 0.032$, $p = 0.969$; electrode x group interaction: $F_{2,42} = 1.108$, $p = 0.340$). Since results from this ANOVA analysis also showed no main effect of stimulus ($F_{2,42} = 1.865$, $p = 0.168$) or [stimulus x group] interaction ($F_{2,42} = 0.264$, $p = 0.769$), we then averaged right-hemisphere lags for each subject across the three stimulus conditions. The resultant values are the averaged right-hemisphere lag across stimulus conditions for each subject, and these values were used in subsequent comparisons to standardized measures of literacy and phonological processing.

With respect to RMS amplitude, RIs were most clearly differentiated from control subjects based on the symmetry of their envelope-following amplitudes in the compressed speech condition (see "Onset and speech envelope period RMS amplitude analyses" in Results). Therefore, we first calculated mean left-hemisphere amplitudes (i.e., the mean of T3, T5 and Tp7) and right hemisphere amplitudes (i.e., the mean of T4, T6 and Tp8) from the analysis of the compressed speech condition, and entered these values into the asymmetry index (R – L) / (R + L) (Fig. 9). Collapsing across electrode pairs was justified in this instance since there was no main effect of electrode ($F_{2,42} = 1.221$, $p = 0.305$) or [electrode x group] interaction ($F_{2,42} = 0.855$, $p = 0.433$) in a 3 x 2 x 2 RMANOVA [electrode x hemisphere x group] in which compressed speech RMS amplitude was the dependant variable. The asymmetry index values were used in subsequent comparisons to standardized measures of literacy and phonological processing.

Brainstem protocol

The procedures to measure brainstem and cortical responses were identical to those that have been described (Russo et al., 2004). Brainstem responses were differentially recorded at a sampling rate of 20kHz using a vertex electrode referenced to the right earlobe. The forehead served as ground. Three blocks of 1000 repetitions were collected at each polarity. Speech sounds were presented to the right ear at 80 dB sound pressure level (SPL) through insert earphones. The inter-stimulus interval was 51 msec. The stimulus used to evoke brainstem responses was the speech syllable /da/ synthesized at a sampling rate of 10 kHz. The stimulus was 40 msec in duration and consisted of five formants with an onset burst during the first 10 ms at F3, F4, and F5.

Data analysis

Analyses utilized previously described measures of the speech-evoked auditory brainstem response, including peak latency and amplitude measures, frequency domain and RMS measures, and measures describing parameters of the peak VA onset complex (Russo et al., 2004). We also calculated composite brainstem scores derived from a subset of theoretically-important measures of the speech-evoked brainstem response. Composite measures included measures of *formant frequency* representation (F1 and HF representation derived from frequency-domain measures; Russo et al., 2004), *acoustic transient* representation (latency of peaks V, A and O; Abrams et al., 2006), *fundamental frequency timing* (latency of peaks D, E and F; Abrams et al., 2006), and *VA complex measures* (VA duration, VA slope, VA amplitude). To calculate composite scores, each measure of brainstem function was transformed into Z-scores, and Z-scores for each of the measures included in a particular composite score were then averaged for each subject.

Cortical measures consisted of the three previously described measures of speech envelope representation, asymmetry for phase-locking precision, phase-locking lag, and RMS amplitude asymmetry. These particular measures are sensitive to standardized measures of literacy and phonological processing measured across individuals with a range of abilities (Abrams et al., submitted-a).

We performed the following 7 analyses to examine potential relationships between auditory brainstem representation of speech-sounds and cortical representation of the speech envelope:

Analysis #1: Pearson's correlations

(1) Pearson's correlations between the three measures of cortical speech envelope processing and individual measures of the speech-evoked auditory brainstem response (Russo et al., 2004). (2) Pearson's correlations between the three measures of cortical speech envelope processing and the four composite brainstem scores described above.

Analysis #2: Cortical response comparisons, grouping based on brainstem responses

(3) Wilcoxon rank-sum test on cortical response measures for the top and bottom ~1/3 of the combined normal and RI group (n=23) based on individual brainstem measures. (4) Rank-sum test on cortical response measures for the top and bottom ~1/3 of the combined normal and RI group based on composite brainstem scores.

Analysis #3: Brainstem response comparisons, grouping based on cortical measures

(5) Wilcoxon rank-sum test on individual brainstem measures for the top and bottom ~1/3 of the combined normal and RI group based on cortical response measures. (6) Rank-sum test on composite brainstem scores for the top and bottom ~1/3 of the combined normal and RI group based on composite brainstem scores.

Analysis #4: Chi-square analysis

We examined the propensity for subjects with abnormal brainstem responses to also exhibit abnormal cortical representation of the speech envelope. We considered a particular brainstem measure abnormal if it exceeded 2 standard deviations of the mean based on normative values generated by our lab (n = 88). We then identified the number of subjects with abnormal

brainstem responses that also had abnormal cortical measures; cortical values were considered abnormal if they exceeded 1.5 standard deviations of the mean defined by this group of subjects (we were less stringent with our criteria for cortical responses given that normative values do not exist for cortical measures of speech envelope representation). We performed chi-square tests to determine whether a statistically disproportionate number of subjects with abnormal brainstem responses also had abnormal cortical responses.

To prevent spurious results from correlation analyses, which can be overly biased by outlying data points, all raw brainstem and cortical values beyond 2 standard deviations (SD) of the mean were moved to the 2 SD point for that particular measure. Across cortical measures, 2 data points (out of a total of 69) were moved to the 2 SD point. For all analyses involving the comparison of means, we used the non-parametric Wilcoxon rank sum test to examine group differences since there were only 8 subjects in the two groups being compared. Pearson's correlations, rank sum tests and chi-square p-values < 0.02 were considered statistically significant; this p-value cutoff represents an adjustment to account for the large number of comparisons being performed.

Results

Analysis #1: Pearson's correlations

Pearson's correlations were performed between individual measures of the speech-evoked auditory brainstem response (19 brainstem measures) and cortical representation of the speech envelope (3 measures). Results indicated one significant correlation (out of 57 correlations performed) between brainstem RMS over the FFR period and R-value asymmetry ($r = 0.489$; $p =$

0.018). All other correlations (56) failed to meet our criteria for significance. Furthermore, none of Pearson's correlations between cortical representation of the speech envelope (3 measures) and composite measures of the speech-evoked auditory brainstem response, including formant frequency representation, acoustic transient representation, fundamental frequency timing, and VA complex measures met our criteria for significance.

Analysis #2: Cortical response comparisons

We performed Wilcoxon rank-sum tests on cortical response measures for the top and bottom ~1/3 of the combined normal and RI groups based on individual brainstem measures. Results from this analysis indicated no significant differences in any of the cortical measures between subjects in the top 33% and bottom 33% for a given brainstem measure. When we used the 4 composite brainstem scores to define the groups rather than individual measures and again there were no significant differences in any of the cortical measures between subjects in the top 33% and bottom 33%.

Analysis #3: Brainstem response comparisons

We performed Wilcoxon rank-sum tests on individual brainstem measures for the top and bottom ~1/3 of the combined normal and RI group based on cortical response measures. Results from this analysis indicated no significant differences in any of the brainstem measures between subjects in the top 33% and bottom 33% for a given cortical measure. There were again no significant differences when we performed Wilcoxon rank-sum tests on composite brainstem measures for the top and bottom ~1/3 of the combined normal and RI group based on cortical response measures.

Analysis #4: Chi-square analysis

We used previously established normative values generated by our lab to examine the possibility that subjects with abnormal brainstem responses may have been particularly likely to exhibit abnormal cortical representation of the speech envelope. Given that no more than 4 subjects were outside the normal range for a given brainstem measure, and that previous studies have indicated that groups of brainstem measures are more sensitive to cortical function compared to individual measures (Banai et al., 2005a; Abrams et al., 2006), we identified subjects who were outside the normal range for *any* measure among a subset of brainstem measures, including (1) Onset/Offset Latencies, (2) Onset/Offset Amplitudes, (3) peaks DEF Latency, (4) peaks DEF Amplitude, and (5) peak VA complex measures. We then identified the number of subjects who were also outside the normal range for any of the three cortical measures of speech envelope representation. We performed two separate chi-square tests to determine whether a statistically disproportionate number of subjects with abnormal brainstem responses also had abnormal cortical responses. The first chi-test examined this question over all 5 sub-groups of tests and failed to reach statistical significance ($X^2_{(4, N = 84)} = 6.10$, $p = 0.192$). A second chi-test examined this question over only Onset/Offset and VA complex measures, which have shown sensitivity to cortical function in previous studies (Banai et al., 2005a; Abrams et al., 2006), and again the statistical result failed to reach significance ($X^2_{(2, N = 51)} = 3.56$, $p = 0.168$).

Discussion

We examined whether there is a relationship between the auditory brainstem response to speech and cortical representation of the speech envelope. We failed to see any relationship between

these two regions of the auditory system. Although one correlation showed statistical significance, the relationship between brainstem RMS amplitude and cortical asymmetry for the precision of envelope following does not represent a theoretically-important relationship. Furthermore, given that there were dozens of correlations performed here, the fact that this one correlation barely reached statistical significance is enables us to consider this finding negligible. In addition to correlation results, a variety of analyses involving the grouping of brainstem and cortical responses also failed to show statistical significance. Based on the overwhelming consistency of these negative results, we argue that there is no relationship between the fidelity of auditory processing in the auditory brainstem and cortical speech envelope representation.

Acknowledgements

Acknowledgements: We thank the children who participated in this study and their families. This work is supported by the National Institutes of Health grant R01 DC01510-10 and National Organization for Hearing Research grant 340-B208.

CHAPTER VI: A ROLE FOR A PARALEMNISCAL AUDITORY PATHWAY IN THE CODING OF SLOW TEMPORAL INFORMATION

Submitted for Publication in *Experimental Brain Research*

Daniel A. Abrams[1], Trent Nicol[1], Steven Zecker[1], and Nina Kraus[1,2]

[1]Auditory Neuroscience Laboratory, Department of Communication Sciences, and

[2]Departments of Neurobiology and Physiology and Otolaryngology.

Northwestern University, 2240 Campus Drive, Evanston, Illinois, 60208.

Abstract

Low frequency temporal information present in speech is critical for normal perception, however the neural mechanism underlying the differentiation of slow rates in acoustic signals is not known. Recent data from the rat trigeminal system suggests that the paralemniscal pathway may be specifically tuned to code low-frequency temporal information. We tested whether this phenomenon occurs in the auditory system by measuring the representation of temporal rate in lemniscal and paralemniscal auditory thalamus and cortex in guinea pig. Similar to the trigeminal system, responses measured in auditory thalamus indicate that slow rates are represented differently in lemniscal and paralemniscal pathways, with paralemniscal responses indicating sensitivity to slow rates. In cortex, both lemniscal and paralemniscal neurons indicated sensitivity to slow rates. We speculate that a paralemniscal pathway in the auditory system may be specifically tuned to code low frequency temporal information present in acoustic signals. Moreover, these data suggest that somatosensory and auditory

modalities have parallel sub-cortical pathways that separately process slow rates and the spatial representation of the sensory periphery.

Keywords: auditory thalamus; auditory cortex; guinea pig; paralemniscal auditory pathway; non-primary auditory pathway; temporal coding.

Introduction

A feature of the speech signal that is important for normal perception is the speech envelope (Rosen, 1992). The speech envelope is defined as slow amplitude modulations present in the signal occurring between 2-50 Hz. Perceptual studies of the speech envelope indicate that it is both sufficient (Shannon et al., 1995) and necessary (Drullman et al., 1994a) for normal perception. Consequently, understanding how slow rates, such as the rates present in the speech envelope, are coded in the central auditory system is of great interest.

In the auditory system, it has been suggested that temporal rate information is represented with a two-stage mechanism in auditory thalamus (Bartlett and Wang, 2007) and cortex (Lu et al., 2001). While these studies suggest disparate mechanisms for the representation of slow and fast acoustic rates, this model does not include details on how slow rates may be differentially represented in the auditory system, a processing stage that would presumably be essential for the discrimination of these rates. Results from the rat trigeminal system demonstrate that slow rates (between 2-8 Hz) are differentially coded by lemniscal and paralemniscal pathways (Ahissar et al., 2000). Specifically, lemniscal neurons in both thalamus and cortex code stimulation rate with constant latencies while paralemniscal

neurons code stimulation frequency as systematic changes in latency. Based on its unique sensitivity for slow rates, it is suggested that the paralemniscal pathway is "optimally tuned for temporal processing of vibrissal information around the whisking frequency range (8 Hz)."

An implication of the somatosensory findings is that a paralemniscal pathway in the auditory system may be optimally tuned to code slow rates present in acoustic signals and could serve as a neural mechanism for speech envelope coding. Surprisingly, there has been no systematic investigation of paralemniscal representation of acoustic rate and whether it differs from the lemniscal representation. To investigate how the auditory system codes slow rates, we measured local field potentials (LFPs) from lemniscal and paralemniscal neurons in guinea pig thalamus and cortex in response to click trains with rates between 2-8 Hz, which represent the most important frequencies in the speech envelope for normal speech perception (Drullman et al., 1994a). Moreover, these rates are also prevalent in the temporal envelope of many guinea pig calls (Wagner and Manning, 1976), and are likely critical for their perception.

Methods

The research protocol was approved by the Animal Care and Use Committee of Northwestern University.

Animal preparation

The experimental materials and procedures were similar to those reported previously (McGee et al., 1996; Cunningham et al., 2002). Eighteen pigmented guinea pigs of either sex, weighing between 400-600 g, were used as subjects. Animals were initially anesthetized with ketamine hydrochloride (100 mg/kg) and xylazine (8 mg/kg). Smaller supplemental doses (25 mg/kg ketamine; 4 mg/kg xylazine) were administered hourly throughout the rest of the experiment. Following the induction of anesthesia, the animal was mounted in a stereotaxic device for the duration of the experiment. Body temperature was maintained at 37.5° C by using a thermistor-controlled heating pad on the guinea pig's abdomen (Harvard). Normal hearing sensitivity was confirmed by auditory brainstem response (ABR). The ABR was elicited by a click stimulus at 70 and 40 dBHL (referenced to normal guinea pig click thresholds) from a recording site located at the posterior vertex/midline of the scalp using an EMG needle electrode. A rostro-caudal incision was made along the scalp surface and the tissue was retracted to expose the skull. Holes were drilled in the skull under an operating microscope. The dura was removed with a cautery to prevent damage to the recording electrode, and the cortical surface was coated with mineral oil.

Anatomical structures

The lemniscal and paralemniscal auditory nuclei investigated in the present study were selected because they have the same reciprocal and parallel connectivity patterns as the lemniscal and paralemniscal pathways in the rat trigeminal system (somatosensory: (Diamond and Armstrong-James, 1992; Woolsey, 1997); auditory: (Redies et al., 1989b). The lemniscal pathway described here consists of the ventral nucleus of the medial geniculate body of thalamus (MGv) and primary auditory cortex. MGv is tonotopically

organized, shows a preference for tonal stimuli and indicates short-latency responses (Redies and Brandner, 1991; He, 2002). Primary auditory cortex in guinea pig consists of two areas, A1 and the dorsocaudal field (DC), which are characterized by tonotopic organization, sharp frequency tuning, a preference for tonal stimuli and short response latencies (Redies et al., 1989a; Wallace et al., 2000). Lemniscal cortex receives its afferent thalamic input from MGv (Redies et al., 1989b).

The paralemniscal nucleus of thalamus described here is the shell nucleus of the medial geniculate body (MGs) (Redies et al., 1989b; He, 2002). The shell nucleus is a band of neurons that surround the MGv dorsally, laterally and ventrally. Neurons of the MGs are generally characterized by broad frequency tuning curves and long-latency responses (Redies and Brandner, 1991; He, 2002). This nucleus projects to the ventral caudal belt of cortex (VCB), the paralemniscal cortical area described in the present work. Neurons of VCB show broad frequency tuning, are more responsive to noise compared to pure tones, and have long-latency responses (Wallace et al., 2000). Based on these particular anatomical connections, it is postulated that these thalamocortical connections represent parallel pathways in the ascending auditory system.

We performed histology on MGv and MGs using hematoxylin and eosin staining and only subtle differences were evident between MGv and MGs with respect to the density of cell bodies. This corroborates previous anatomical investigations of guinea pig MGB using Nissl-staining indicating that a distinct cytoarchitectonic division between these nuclei was not present; it was proposed that the cell population of these nuclei is intermingled (Redies et

al., 1989b). During data collection, we relied on the relative anatomical locations of these nuclei (the Shell subdivision surrounds the Ventral subdivision dorsally, laterally and ventrally), as well as the substantial differences between their responses to various probe stimuli, as means to distinguish these two nuclei.

Acoustic stimuli

Acoustic stimuli were generated digitally and presented in Matlab (Mathworks). Acoustic stimuli were delivered to the contralateral ear using Etymotic insert earphones (ER2) through the hollow earbars of the stereotaxic device. The sound pressure level (SPL, expressed in dB re 20 mPa) was calibrated over a frequency range of 20-20 kHz using a condenser microphone (Brüel and Kjaer). Three second-long click train stimuli were delivered at a level of 75 dB SPL (peak intensity). Click trains of 2, 5 and 8 Hz were randomly presented with an inter-stimulus interval of 1 s. 100 repetitions of each click rate were presented at all electrode penetration. Clicks consisted of 100 μsec rectangular pulses. Clicks with alternating polarities were presented to remove any possibility of a stimulus artifact within the response. The delivery system output the signal through a 16-bit converter at a sampling rate of 16 kHz. That system triggered the PC-based collection computer. All stimuli were presented in a sound-treated booth (IAC). Third-octave tone-pips were used to map auditory cortex. Mapping of auditory cortex was essential to properly locating the paralemniscal cortical nucleus. Tones were 100 msec in duration with a rise-fall time of 10 msec.

Neurophysiologic recording

Both thalamus and cortex were accessed with a vertical approach using tungsten microelectrodes (Micro Probe) with impedance between 1-2 MΩ at 1 kHz. An electrode was advanced perpendicular to the surface of cortex using a remote-controlled micromanipulator (Märzhäuser-Wetzlar). The coordinates of the electrode were determined at a point slightly above cortex at the first penetration, and these coordinates were kept for the remainder of the experiment.

Typically, recordings of MG and cortex were performed on different animals and recording sessions. For recording MGv, locations were approximately 4.8 mm rostral to the interaural line, 4.0 mm left or right of the sagittal suture and 7.2 mm ventral to the surface of the brain. Visual inspection of the response size and waveform morphology was considered. If the response was small in amplitude and broad in shape, electrode penetration was continued. This process was repeated until the morphology of the waveform conformed to the large amplitude, sharp onset response commonly observed in recordings obtained from the ventral division of the MG. Previous use of this technique has shown a 100% hit rate for the ventral division of the medial geniculate using the stereotaxic and physiological criteria described above (King et al., 1999; Cunningham et al., 2002). For recording MGs, the same recording procedures described above were used, however the recording electrode was gradually moved laterally from MGv until neural responses to clicks increased in latency, a known characteristic of MGs (Redies and Brandner, 1991; He, 2002). Due to the relatively large volume measured with LFPs and the proximity of lemniscal and paralemniscal nuclei, on many occasions the responses to both nuclei were recorded at a single penetration.

Responses from these nuclei were often separable in single trial observations, as monitored on our oscilloscope, and were always separable in the 100 trial averaged waveforms.

For recording lemniscal areas A1 and DC, locations were approximately 3 mm caudal to bregma and 10 mm lateral of the sagittal suture. Recordings were made at depths of 500-900 μm, corresponding to cortical layers III and IV in guinea pig. Auditory cortex was mapped using third-octave tones to enable description of best frequency for lemniscal penetrations. To determine the location of paralemniscal area VCB, we used the frequency map of lemniscal regions A1 and DC to identify the ventral border of area DC which abuts area VCB. VCB was easily identified when the electrode had entered this cortical region: neural responses became significantly later, showed negligible and non-frequency-specific responses to pure-tones, and moderate-to-large responses to noise bursts (Wallace et al., 2000).

The electrode signal was amplified using Grass amplifiers with filters set between 1-20,000 Hz. The analog signal was digitized at 33.3 kHz by an A-D card (MCC) attached to a PC-based computer system. Responses were logged and stored using Matlab routines designed by our lab. Recorded brain responses were off-line filtered between 10-100 Hz to isolate local field potentials (LFPs) (Eggermont and Smith, 1995) and downsampled to 10 kHz using an algorithm that applies an anti-aliasing (lowpass) FIR filter during the downsampling process (Matlab, Mathworks, Inc). Following each experiment, the final thalamic recording location was marked with an electrolytic lesion (35 mA for 10 s) to enable correlation between electrophysiological recordings and histology.

Data analysis

LFPs to each click rate were averaged across 100 repetitions for all penetrations. First, we isolated LFPs in response to clicks two and three for each click rate. The justification for analyzing responses to the second and third clicks in a train is that accurate coding of the speech envelope likely requires an efficient neural mechanism that can rapidly track speech envelope frequencies in a continuous signal. In addition, we analyzed "steady-state" LFPs, defined here as the average LFP measured during the final second of the click train at each rate. The reason for this analysis is that it enabled direct comparison to the results from the rat trigeminal system.

The first major deflection in the LFP was identified automatically using software code in Matlab. Correct identification of all peaks was verified visually. Amplitudes and latencies at both LFP peak and half-peak were identified for statistical analysis, however since the goal of this paper is to investigate effects of click rate on LFP latency, we are only presenting peak amplitude statistics. Peak and half-peak latencies as well as peak amplitudes were statistically evaluated using a repeated-measures design which enabled a description of within-penetration changes in LFP latency and amplitude for increasing stimulus rate. When comparison of responses between nuclei is reported, these analyses were also conducted using a repeated-measures design.

Results

Thalamus: lemniscal and paralemniscal responses

The averaged LFPs in response to the second and third clicks within a train are plotted at each of the three repetition rates as a function of time in Figure 18.

In the lemniscal nucleus of the MG (Fig. 18a), response characteristics were similar irrespective of click rate: response latency remained nearly constant while LFP amplitude decreased slightly with increasing stimulus rate, consistent with an adaptation process. Alternatively, in the paralemniscal nucleus of MG (Fig. 18b), response latency systematically shifted later in time with increasing stimulus rate, with a similar amplitude decrement as that seen in the MGv. These particular response characteristics were seen across many penetrations measured in seven different experimental animals.

To quantify the effect of click rate on response latency and amplitude for lemniscal and paralemniscal thalamic nuclei for all recording penetrations (n=35 for MGv; n=15 for MGs), we identified the peak of the LFP at each click rate and recorded its latency and amplitude values. Since paralemniscal auditory neurons typically respond later than lemniscal neurons (Redies and Brandner, 1991; He, 2002) (see Table 6 for descriptive statistics for lemniscal and paralemniscal thalamic responses), we wanted to ensure that statistical differences between lemniscal and paralemniscal LFPs were not due to proportionally similar latency increases. Therefore, we normalized response latencies measured at each penetration to the latency measured in the 2 Hz condition for that penetration.

To test whether LFP latency characteristics distinguish lemniscal from paralemniscal thalamus, we performed a repeated measures ANOVA on normalized LFP peak latencies

measured from MGv and MGs. A significant difference between lemniscal and paralemniscal response latency for increasing stimulus rates was confirmed by a significant effect of pathway in this analysis ($F_{1,48}$ = 13.518, P=0.001; Figure 19a). Additional repeated measures ANOVA analyses revealed that there was no effect of click rate on LFP latencies measured from lemniscal thalamus ($F_{2,68}$ = 1.89; P=0.159), however a significant effect of click rate on paralemniscal response latency was evidenced with increased latencies seen for increasing stimulus rates ($F_{2,28}$ = 12.107, p=0.0001). *Post hoc* pairwise comparisons of paralemniscal latencies indicated significantly later responses for 5 Hz compared to 2 Hz click rates (t = 2.097, P=0.055) as well as 8 Hz compared to 5 Hz click rates (t = 2.405, P=0.03). Results from latency at half-peak were identical to those described for latency at peak (effect of pathway: $F_{1,48}$ = 11.660; P=0.001; effect of click rate on lemniscal LFPs: $F_{2,68}$ = .538, P=0.586; effect of click rate on paralemniscal LFPs: $F_{2,28}$ = 24.752, P<0.0001). The same effects were also prevalent in steady-state peak latencies (Figure 19a inset, effect of pathway: $F_{1,48}$ = 19.722, P<0.0001; effect of click rate on lemniscal LFPs: $F_{2,68}$ = 1.102, P=0.338; effect of click rate on paralemniscal LFPs: $F_{2,28}$ = 24.752, P<0.0001). These results indicate that at the level of thalamus, a paralemniscal auditory nucleus systematically encodes temporal rate information in the form of a latency shift. This encoding is seen no later than the third click in a train and persists for up to 3 seconds.

To determine whether the effect of click rate on LFP amplitude distinguishes lemniscal from paralemniscal thalamus, the same statistical analyses were performed on peak amplitudes values (Figure 19b). All amplitude values from a given penetration were normalized with respect to amplitudes measured in the 2 Hz condition. There was no statistical difference

between lemniscal and paralemniscal thalamus with respect to LFP amplitude changes for increasing stimulus rate (LFP amplitude pathway effect: $F_{1,48} = 0.005$, $p = 0.945$). Consistent with an adaptation process, increases in stimulus rate resulted in significant decreases in LFP amplitude in lemniscal thalamus (MGv: $F_{2,68} = 30.076$, $P<0.0001$). There was a trend for the same effect in paralemniscal thalamic nucleus (MGs: $F_{2,28} = 1.931$; $P=0.160$). The failure to reach statistical significance in paralemniscal thalamus was due to responses measured from a single penetration which showed a preference (i.e., larger LFP amplitudes) for rates greater than 2 Hz. When this outlying data point was removed from this analysis, an effect of stimulus rate on normalized LFP amplitude (i.e., adaptation) was evident (MGs: $F_{2, 26} = 5.852$; $P=0.001$). These data indicate similarity between lemniscal and paralemniscal LFP amplitudes in response to different click rates.

Cortex: lemniscal and paralemnsical responses

To determine if lemniscal and paralemniscal nuclei of the cortex also showed differential encoding of rate information, we measured LFPs from lemniscal auditory cortex (fields A1 and DC; n=29) and the paralemniscal ventral caudal belt (VCB; n=24) in response to the same click trains as described for the thalamus (see Table 7 for descriptive statistics for lemniscal and paralemniscal cortical responses). All latency and amplitude values were normalized with respect to values measured in the 2 Hz condition. The averaged LFPs in response to the second and third clicks within a train are plotted at each of the three repetition rates as a function of time in Figure 20.

For lemniscal responses, the relationship between click rate and response latency was varied: one third of lemniscal cortical responses were latency-invariant for increasing click rate (Fig. 20a) while the remaining responses were latency-variant in a manner similar to paralemniscal thalamus, with increasing click rates inducing longer response lags (Fig. 20b). All latency invariant lemniscal responses had short (absolute) latency responses. Statistical analysis across all lemniscal cortex responses revealed a significant increase in response latency for increasing click rate ($F_{2,56} = 73.189$; $P<0.0001$; Fig. 21a). *Post hoc* pairwise comparisons of lemniscal latencies indicated significantly later responses for 5 Hz compared to 2 Hz click rates ($t = 7.693$, $P<0.0001$) as well as 8 Hz compared to 5 Hz click rates ($t = 6.608$, $P<0.0001$). Paralemniscal cortical response latencies also varied according to click rate, with increasing stimulus rates resulting in longer response latencies ($F_{2,46} = 78.573$; $P<0.0001$; Fig. 20c). Unlike lemniscal cortex, all paralemniscal penetrations followed this pattern. *Post hoc* pairwise comparisons of paralemniscal latencies indicated significantly later responses for 5 Hz compared to 2 Hz click rates ($t = 7.230$, $P<0.0001$) as well as 8 Hz compared to 5 Hz click rates ($t = 6.346$, $P<0.0001$). Results from latency at half-peak were identical to those described for latency at peak (effect of click rate on lemniscal LFPs: $F_{2,56} = 52.020$, $P<0.0001$; effect of click rate on paralemniscal LFPs: $F_{2,46} = 77.559$, $P<0.0001$). There was no statistical difference between lemniscal and paralemniscal cortical response latencies for increasing stimulus rates (effect of pathway: $F_{1,51} = 0.002$; $P=0.966$, not significant).

We quantified the effect of click rate on peak LFP amplitudes measured at cortex in response to the second and third click in the trains (Fig. 21b). For lemniscal cortex, there was a significant effect of rate on normalized LFP amplitude ($F_{2,56} = 11.356$; $P<0.0001$).

Specifically, lemniscal cortical LFPs indicated a preference for the 5 Hz condition, with LFP amplitudes increasing significantly between 2 Hz and 5 Hz ($t = 3.289$, $P=0.003$) and decreasing between 5 Hz and 8 Hz ($t = 4.752$, $P<0.0001$). The preference for 5 Hz based on LFP amplitude may reflect a rate code in lemnsical auditory cortex. Despite a similar trend of preference for 5 Hz for paralemniscal cortex, there was no effect of click rate on normalized LFP amplitude ($F_{2,46} = 0.384$; $P=0.683$). This failure to reach statistical significance can be explained by the fact that approximately half of the paralemniscal LFP amplitudes indicated a preference for the 5 Hz condition, similar to lemniscal cortical responses, while the other half of the responses preferred the 2 Hz condition. There was no statistical difference between lemniscal and paralemniscal cortical response amplitudes for increasing stimulus rates (LFP amplitude rate * pathway interaction, $F_{1,51} = 0.314$, $P=0.577$).

Two important differences were identified between cortical LFPs to the second and third clicks in the train and steady-state LFPs. First, some paralemniscal LFP amplitudes diminished into the noise floor towards the end of the 5 and 8 Hz click trains, consistent with an adaptation process. This only affected steady-state responses and occurred in 25% of paralemniscal sites in response to the 5 Hz condition and 42% of paralemniscal sites in response to 8 Hz click trains. Responses from these sites were not included in the steady-state statistical analysis. The second important difference between steady-state cortical LFPs and LFPs to the second and third clicks in the train was that steady-state LFP latencies consistently decreased between the 5 Hz and 8 Hz conditions for both lemniscal and paralemniscal recording sites (Fig. 21a inset) in contrast to the systematic increase in latency

with increasing click rate seen in LFPs to the second and third clicks in a train (Figure 21a, main figure).

Due to the different rate-latency pattern between LFPs to the second and third clicks in the train and steady-state cortical LFPs, we investigated the dynamics of lemniscal LFPs. Results indicate two distinct patterns of activation across the three click rate conditions for lemniscal recording sites. In one pattern, lemniscal peak LFP latencies remained extremely consistent, in this case between 10-11 msec, across all clicks in all three rate conditions (Fig. 22). This occurred in 34% of lemniscal recording sites. A second pattern of activation revealed dynamic shifts in LFP latency that were evident both within a click rate condition and across rate conditions (Fig. 23). For example, in the 2 Hz click rate condition, following the first click, LFP peak latencies are relatively static throughout the click train. In contrast, LFP peak latencies in the 5 Hz condition gradually shift during the first 5 clicks in the train before settling at a latency of 24 msec. In the 8 Hz click train condition, LFP peak latencies initially jump to a latency of 27 msec before settling to a latency of 19 msec by click nine. These particular patterns were seen in 66% of lemniscal recording sites. We also investigated the dynamics of paralemniscal LFPs (Fig. 24). In the 2 Hz click rate condition, peak LFP latencies again are fairly static throughout the click train. In the 5 Hz click rate condition, peak LFP latencies make a large shift to 30 msec by 1 second into the 3 second click train. In the 8 Hz click rate condition, LFPs adapt immediately after the first click in the train. These particular patterns were seen in 85% of paralemniscal recording sites.

Discussion

To investigate central auditory system coding of slow acoustic rates, we measured responses from lemniscal and paralemniscal auditory thalamus and cortex in guinea pig to click train stimuli with rates between 2-8 Hz. At the level of thalamus, lemniscal neurons were latency invariant regardless of click rate while paralemniscal neurons showed systematic latency shifts with increasing temporal rate. Similar latency shifts have been demonstrated in the rat paralemniscal trigeminal system and presumably indicate sensitivity to input rate (Ahissar et al., 2000). At the level of cortex, some lemniscal responses were rate invariant while others showed dynamic latency shifts. In contrast, paralemniscal sites in cortex consistently showed latency shifts with increasing rate.

Auditory coding of slow rates: a new hypothesis

Previous studies have demonstrated that acoustic rate information may be represented by the auditory thalamus (Bartlett and Wang, 2007) and cortex (Lu et al., 2001) with a two-stage mechanism. These studies suggest that slow acoustic rates (< ~50 Hz in cortex; < ~200 Hz in thalamus) are represented "explicitly" according to the temporal discharge patterns of synchronized lemniscal auditory neurons, while faster rates are represented "implicitly" according to the average discharge rate of non-synchronized neurons. The goal of this work was to investigate neural mechanisms that underlie the perception of slow acoustic rates.

Results described in the present study suggest that temporally synchronized neurons of a paralemniscal pathway, as well as lemniscal cortex, differentially represent specific low-frequency stimulus rates (<10 Hz) with discrete shifts in response latencies. These latency shifts may provide a neural code for slow rate information. Rate-sensitive neurons showed

selectivity to particular rates within a short time period (by the third click in a train) in both thalamus and cortex. Consequently, results from paralemniscal neurons of thalamus and cortex enable a new hypothesis for the role of this pathway in the representation of biologically relevant acoustic signals. Specifically, paralemniscal pathways may be involved in the representation of the low-frequency temporal envelope of complex acoustic signals, including speech and animal communication calls. Moreover, lemniscal cortex is also implicated in the processing of these slow temporal features.

Previous studies have proposed neural codes for the representation of slow temporal rates (Horst et al., 1986; Cariani and Delgutte, 1996; Eggermont, 1998). For example, if a neuron synchronizes to all of the clicks within a click train, then the interspike or interburst interval will be different for different repetition rates and these different intervals could conceivable code the stimulus rate. A counterargument to this hypothesis is the notion that if a stimulus variable is "processed" by a neuron or group of neurons, the neural response should be characterized by some sort of alteration. A stimulus variable that is kept constant from station to station in the nervous system preserves the stimulus information but does not suggest processing, leaving the processing for higher levels. For example, if a group of neurons activate every 500 msec in response to a 2 Hz click train and every 100 msec in response to a 10 Hz click train, one might conclude that these neurons are faithfully relaying this temporal information to downstream nuclei. However, if a group of neurons shows systematic latency shifts for increasing stimulus rates as the current data show, then it can be argued that these neurons are actively processing the stimulus rate in some meaningful fashion. This argument is based on the assumption that neural processing is done in closed-

loops, which is thought be a fundamental characteristic of the auditory thalamocortical system. In this case, processing involves dynamic changes in the value of the processed variable, such as the systematic latency shifts for increasing click rates shown here.

Comparison to the rat trigeminal system

Previous findings from the rat trigeminal system indicate discrete properties for lemniscal and paralemniscal neurons of thalamus and cortex (Ahissar et al., 1997; Ahissar et al., 2000; Sosnik et al., 2001), for review see (Ahissar and Zacksenhouse, 2001). In both thalamus and cortex, the lemniscal pathway showed sensitivity to spatial organization of the periphery (i.e., vibrissae location) and fixed time-locking for various stimulus rates while the paralemniscal system was insensitive to spatial organization of the periphery but showed variable time-locking. Based on these response characteristics it was proposed that these parallel pathways serve different neural functions related to exploration and active touch during whisking. Specifically, it was proposed that lemniscal neurons perform spatial processing of the sensory periphery with sensitivity to whisker location but fixed time-locking while paralemniscal neurons perform low-frequency temporal processing with variable time-locking, but lack spatial processing.

If the current findings in the auditory system are considered with respect to response characteristics of the trigeminal system, as well as their putative functional roles, these data provide novel hypotheses regarding the processing of spectral and temporal acoustic cues in the central auditory system. First, in lemniscal thalamus, it has been shown that neurons reflect the spatial organization of the auditory periphery (i.e., frequency organization of the

cochlea) while paralemniscal thalamic neurons typically show poor frequency selectivity (Redies and Brandner, 1991; He, 2002). Results from the present study suggest that lemniscal thalamus lacks slow temporal rate processing, while paralemniscal thalamus processes rates with latency shifts. Therefore, it is hypothesized that lemniscal neurons of thalamus code spectral frequency with fixed time-locking and sensitivity to the cochleotopic arrangement of the auditory periphery while paralemniscal neurons perform low-frequency temporal processing with variable time-locking and insensitivity to cochleotopy.

Similar to thalamus, lemniscal neurons of cortex are sensitive to spectral frequency while paralemniscal neurons are again insensitive to spectral frequency (Merzenich et al., 1975; Wallace et al., 2000). Results from the present study suggest that the majority of lemniscal, as well as paralemniscal cortical neurons, are sensitive to slow rates. Therefore, based on these response properties, it is hypothesized that the majority of lemniscal neurons of cortex process both spectral frequency and slow temporal rate (i.e., rate-variant neurons) while paralemniscal cortex exclusively processes low-frequency temporal information.

An important finding of this work is that there appear to be two discrete populations of cortical sites: in one population, absolute response latencies are long and response latencies increase with increasing temporal rates, and in the other population, absolute latencies are short and are unaffected by click rate (Fig. 20). The latter population is exclusively in the lemniscal cortex while the former population is in both cortices. These results allow the possibility that lemniscal cortex contains a "lemniscal core" domain, in which latency is short and insensitive to temporal rate, and an integrative domain, which also received

paralemniscal information. This is comparable to results in the rat trigeminal system, in which the same cortical area, barrel cortex, contains domains (layers, in this case) in which neurons are sensitive to the temporal rate (layers 5a and 2/3) and other domains in which they are not (layers 4 and 5b). Therefore, one possibility is that the lemniscal cortex in the auditory system contains both domains, and that the long-latency domains might reflect integration of both pathways. This would be analogous the somatosensory case in which layer 2/3 neurons show signs of integration of both pathways (Ahissar et al., 2001a; Bureau et al., 2006).

An important difference between results from rat trigeminal system and the current results from the auditory system is the response of lemniscal cortical neurons to slow temporal rates. In the somatosensory system, lemniscal cortex (layers 4 and 5b) was insensitive to slow temporal rate information while in the auditory system the majority of these sites were sensitive to slow temporal rate. A potential explanation for this discrepancy is that sensory cortex is more remote from the periphery in the auditory system relative to somatosensory cortex. Consequently, lemniscal auditory cortex could represent a higher level of processing relative to the somatosensory cortex that is responsible for the integration of lemniscal and paralemniscal temporal rate responses. Moreover, the additional processing of slow rates in lemniscal cortex of the auditory system could reflect the import of the temporal envelope of communication calls for perception (Drullman et al., 1994a).

It should be noted that there are significant anatomical differences in the configuration of lemniscal and paralemniscal cortex in the somatosensory relative to the auditory system. In

the somatosensory system, lemniscal and paralemniscal cortex share the same area of barrel cortex but occupy different layers (lemniscal: layers 4 and 5b; paralemniscal: layer 5a) while in the auditory system, lemniscal and paralemniscal cortex are separate areas altogether. Layer 2/3 neurons of the somatosensory system also show latency shifts consistent with paralemniscal cortex (Ahissar et al., 2001a). A simple explanation linking the aforementioned anatomical and physiological differences between somatosensory and auditory cortices is not straightforward.

A major conclusion of the work by Ahissar and colleagues is that the paralemniscal thalamo-cortical loop may operate as a phase-locked loop (PLL) to convert temporal information to a rate code. The PLL is a closed-loop circuit in which an internal oscillator matches the frequency of an input signal, and changes in input frequency appear as a change in the phase relationship between the oscillator and input frequency. All of the elements necessary for a PLL are present in the somatosensory system: the physical loop is formed by reciprocal connections in the paralemniscal pathway between the posterior nucleus of the thalamus and layer 5a of barrel cortex; internal oscillators are provided by individual cortical cells in layers 5/6 of barrel cortex which produce spontaneous oscillations around 10 Hz (Ahissar et al., 1997); an input signal is provided by the spinal nucleus interpolaris of the midbrain. Ahissar and colleagues have argued that these elements in the paralemniscal somatosensory pathway function as a PLL: “this circuit establishes a negative feedback loop in which the latency of the cortical oscillations determines the latency and spike count of the thalamic neurons, and the thalamic spike count determines the subsequent latency of the cortical oscillations” (Ahissar et al., 2000).

Given previous findings as well as results described here, it is possible that the paralemniscal auditory pathway in thalamus and cortex may also operate as a phase-locked loop (PLL). All of the elements necessary for a PLL are likely present in the guinea pig paralemniscal auditory pathway: a physical loop is formed between paralemniscal nuclei MGs and VCB; across animal models, spontaneous rhythmic activity has been demonstrated in auditory cortex (Eggermont, 1992; Lakatos et al., 2005) and this is likely the case in guinea pig as well; an input signal is provided presumably by the auditory midbrain nucleus, the inferior colliculus. Furthermore, results from paralemniscal thalamo-cortical neurons in the current work indicate response latency characteristics predicted by the PLL. If PLL circuits operate in the auditory paralemniscal pathway, they could be useful in decoding the information contained in the temporal envelope of acoustic signals such as conspecific calls. An additional advantage is that such PLLs can provide the brain with an internal timing reference signal, which signals the beginning of a chunk of informative spectral information. This internal reference adaptively locks to the input rate of information and follows slow changes in that rate (Ahissar and Ahissar, 2005). Additional work is required to further investigate whether a PLL is responsible for response characteristics in the auditory system.

As mentioned previously, an essential component of a PLL is an internal oscillator that matches the frequency of the input signal. In the somatosensory system, the frequency of the internal oscillators is ~ 10Hz which matches the whisking frequency. In the auditory system, the low frequency temporal envelope is variable in frequency. Therefore, in order for a PLL to function in the auditory system, the internal oscillators would have to have similar

frequency variability. Indeed it has been shown that spontaneous oscillations measured from auditory cortical neurons cover a range of frequencies [6-11 Hz (Eggermont, 1992); 1-32 Hz in awake macaque (Lakatos et al., 2005); up to 20 Hz in anesthetized rat (Gaese and Ostwald, 1995)] and this range shows a reasonable correspondence to the range of frequencies present in the temporal envelope of species-specific calls and other naturally-occurring acoustic stimuli. While the lack of a stable "locking" frequency in the auditory system would likely indicate a more complex system of PLLs than the somatosensory system, we do not believe that the variability of the auditory input would preclude the existence of PLLs in this system.

While many features of this system are consistent with the possibility that the paralemniscal auditory pathway in thalamus and cortex may operate as a phase-locked loop (PLL), the latency shifts described in this work for increasing stimulus rates may simply represent adaptation. In fact, it has been shown that $GABA_B$-mediated feedback inhibition from the reticular thalamic nucleus can explain the latency shifts in thalamus of the somatosensory system (Golomb et al., 2006). This may also be the case in the auditory system, and it is hoped that future studies can address the cellular mechanisms underlying the latency shifts demonstrated in this work.

A methodological difference between the somatosensory results (Ahissar et al., 1997; Ahissar et al., 2000; Ahissar et al., 2001a; Sosnik et al., 2001) and results presented in the current work is that the metric for neural activity in the somatosensory work is based on spikes while LFPs were measured here. LFPs and spikes result from the interaction of synaptic and cellular mechanisms with LFPs representing the compound input to the cells

while spiking represents the firing of action potentials (Logothetis et al., 2001). These two events are closely related, as spiking results if the right balance of excitatory and inhibitory post-synaptic potentials is achieved (Kaur et al., 2004). Moreover, it has been shown that there is a high correlation between LFPs and spiking activity for rate coding in primary auditory cortex (Eggermont and Smith, 1995). While it seems unlikely that comparing LFPs in the auditory system to spikes in the somatosensory system presents a major interpretive confound, investigating these particular auditory phenomena by measuring spikes would represent another step towards describing analogous response properties for rate coding in the auditory and somatosensory systems.

Comparison to previous studies in the auditory system

Lemniscal pathway

Slow temporal rates have been studied extensively in the lemniscal pathway of the auditory system (thalamus: (Vernier and Galambos, 1957; Creutzfeldt et al., 1980; Rouiller et al., 1981; Rouiller and de Ribaupierre, 1982; Preuss and Muller-Preuss, 1990; Miller et al., 2002); cortex: (Goldstein et al., 1959; Creutzfeldt et al., 1980; Phillips et al., 1989; Eggermont, 1991; Bieser and Muller-Preuss, 1996; Steinschneider et al., 1998; Lu and Wang, 2000; Lu et al., 2001). At the level of thalamus, many studies have investigated the representation of repetition rate with an emphasis on identifying the maximum frequency that these neurons are able to synchronize, however the current study is the first systematic description of onset latencies for various click rates in lemniscal thalamus. On the other hand, the current results from lemniscal cortex corroborate previous descriptions of latency shifts for increasing stimulus rates. Specifically, single-unit responses measured in auditory cortex

of anesthetized cat showed systematic latency shifts for increasing stimulus rates similar to those shown here (Phillips et al., 1989). In a similar study in unanesthetized rat, however, no latency shifts were shown for rates below 20 Hz (Anderson et al., 2006). A possible explanation for this discrepancy is that spike timing of a particular unit in the latter study did not appear to be analyzed with respect to its own timing at other rates (i.e., repeated measures ANOVA statistics apparently were not used). Therefore, relatively small increases in spike-timing for increasing rates seen within individual neurons may have been overwhelmed by the variability of spike times between neurons.

It remains a possibility that the pattern of latency shifts for increasing click rates is due to the use of anesthesia. Previous studies have shown that neural adaptation at lower frequencies in the somatosensory thalamus and cortex increases when the experimental animal is in a quiescent state (i.e., when anesthetized) relative to an active state (Castro-Alamancos, 2002). Therefore, it is conceivable that LFP latency shifts seen at higher-frequency click rates in the current study are influenced by anesthetic state. This appears to be an unlikely explanation of the current results, however, since only some of the neural responses in thalamus and cortex exhibited latency shifts despite controlling for anesthetic state (to the best of our ability) across all neurophysiological measurements.

Paralemniscal pathway

Physiological properties of paralemniscal nuclei of thalamus and cortex in the guinea pig auditory system have been described in a number of previous studies (Redies et al., 1989a; Redies and Brandner, 1991; Wallace et al., 2000; He, 2002, 2003b, a). Studies in

paralemniscal thalamus have shown that these nuclei may be responsible for coding acoustic onset and offset information and play a role in the gating of acoustic information to and from cortex (He, 2002, 2003b, a). To our knowledge, the current study is the first to provide a description of paralemniscal processing of acoustic rate information. Moreover, these data are the first to suggest that paralemniscal nuclei may be responsible for the processing of slow rate information in acoustic signals.

A final question raised by this work involves the hierarchy of the paralemniscal auditory pathway at the levels of auditory thalamus and cortex. While it is tempting to consider paralemniscal pathways as separate neural entities from the lemniscal pathway, this may be an oversimplification of a more complex junction in the ascending auditory system. While it has been shown that the VCB in guinea pig receives its primary thalamic input from MGs and MGcm, it also receives projections from adjacent lemniscal cortex. Similar connections have been shown in marmosets (de la Mothe et al., 2006b, a). Since the current results show that paralemniscal thalamus/cortex and lemniscal cortex show similar sensitivity to slow rates, these data do not address which input might dominate paralemniscal cortex. However, given that both paralemniscal thalamus and cortex are insensitive to spectral frequency, it could be argued that paralemniscal thalamus provides the dominant input to paralemniscal cortex. Future studies may begin to disentangle the relative interactions of these auditory structures.

Implications for speech envelope processing in humans

Slow rates present in speech, known as the speech envelope, are important for normal speech comprehension (Drullman et al., 1994a; Shannon et al., 1995). The current results indicate that a paralemniscal pathway appears to be specifically tuned to process low-frequency temporal information, such as that present in the speech envelope, and in the human auditory system, may play an important role in the perception of this speech feature. In addition, these data suggest that the lemniscal auditory cortex is the first nucleus in the lemniscal pathway to process the low frequency content of the speech envelope.

The click stimuli used in the current experiment are, by design, considerably different from speech. Speech has enormous variability in both temporal and spectral domains whereas the stimuli used here are broadband, discrete click trains. Despite the gross acoustic differences between these stimuli, they both contain slow temporal rates between 2-8 Hz, which are the most important frequencies in the speech envelope for normal speech perception (Drullman et al., 1994a). Therefore, an interpretative limitation of this work is how these mechanisms would respond to more complex acoustic stimuli such as speech. Moreover, it is difficult to predict how more spectrally complex acoustic stimuli might affect these particular responses. Future studies may be able to test how these particular response properties might change when a more dynamic acoustic signal is used to stimulate auditory neurons.

Abnormal perception of slow amplitude modulations, such as those present in the speech envelope, are associated with reading disabilities (Lorenzi et al., 2000; Ahissar et al., 2001b; Goswami et al., 2002). Given that a paralemniscal auditory pathway indicates sensitivity to speech envelope frequencies, an interesting possibility is that reading impairment is

associated with abnormal paralemniscal auditory function. Little is known about the relative contributions of lemniscal and paralemniscal pathways to the auditory deficits shown in reading-disabled individuals or other clinical populations in which hearing function is in question. Future studies of the auditory system using imaging techniques may be able to begin disentangling the question of relative contributions of lemniscal and paralemniscal function in clinical populations.

The existence of discrete sub-cortical networks for the analysis of temporal and spectral acoustic information represents an exciting possibility for human encoding of speech. The existence of separate pathways for encoding the discrete temporal and spectral cues in speech would provide an elegant mechanism for their coding in the central auditory system.

Comparison to previous studies in the auditory system

In conclusion, we have shown dissociation between lemniscal and paralemniscal pathways in the encoding of slow rates in the auditory thalamus, similar to findings from the rat trigeminal system. Moreover, we have shown convergence of these two rate representations at the level of lemniscal auditory cortex. Results suggest that at the level of thalamus, spectral frequency and temporal rate are processed in lemniscal and paralemniscal pathways, respectively. Additionally, data suggest that lemniscal auditory cortex is the first nucleus in the lemniscal pathway that is sensitive to both spectral information and low-frequency temporal information. These data suggest a mechanism for the encoding of the speech envelope in humans.

Acknowledgements: This work is supported by the National Institutes of Health grant R01 DC01510-10 and National Organization for Hearing Research grant 340-B208. We thank M. Hartmann and E. Ahissar for critical reviews of this manuscript and C. Warrier for assistance with data collection.

CHAPTER VII: REPRESENTATION OF THE SPEECH ENVELOPE AND PERIODICTY: EVIDENCE FROM CORTICAL ENSEMBLES

Manuscript in preparation

Daniel A. Abrams[1], Trent Nicol[1], Mitra Hartmann[2], Steven Zecker[1], and Nina Kraus[1,2,3]

[1]Auditory Neuroscience Laboratory, Department of Communication Sciences

[2]Departments of Neurobiology and Physiology

[3]Department of Otolaryngology.

Northwestern University, 2240 Campus Drive, Evanston, Illinois, 60208.

Abstract

Temporal features in speech represent an important class of acoustical cues for normal speech perception. The primary goal of the current study was to investigate cortical representation of two of these temporal features, the speech envelope and periodicity cues. We measured neural representation to speech stimuli in primary auditory cortex and an adjacent non-primary cortical field in anesthetized guinea pig. With respect to speech envelope representation, results showed that in primary auditory cortex, the speech envelope is represented by the envelope of population responses and that these ensembles tended to show more precise representation of wider bandwidths of the speech envelope compared to narrow-bands. The fundamental frequency was represented robustly by neuronal populations in primary auditory cortex with phase-locked responses across all three speech conditions. Responses measured from non-primary auditory cortex were heterogeneous in response to temporal features in speech: some cortical sites represented temporal features of the speech

signal in a time-locked fashion while other sites did not. Results show that the auditory system is adept at simultaneously resolving multiple temporal features in complex acoustic signals using discrete coding mechanisms.

Keywords: auditory thalamus; auditory cortex; guinea pig; paralemniscal auditory pathway; non-primary auditory pathway; temporal coding.

Introduction

Human speech perception depends critically on temporal features of the speech signal. Rosen presented a framework for these temporal features which segregated the speech signal into three frequency ranges: the speech envelope (2-50 Hz), periodicity cues (50 – 500 Hz), and the temporal fine structure (600 – 10,000 Hz) (Rosen, 1992). The speech envelope is dominated by the syllable rate of speech. Periodicity cues include the representation of the fundamental frequency (F0) of the speaker's voice, conveys prosodic information and, in the case of tonal languages, semantic information. The temporal fine-structure provides information about the spectrum and formant structure of speech sounds. The primary goal of the current study is to investigate the primary and non-primary cortical representations of the speech envelope and periodicity cues in speech sentence stimuli.

The neural representation of the speech envelope has not been investigated in near-field recordings from auditory cortex. Information about how cortical neurons might respond to the speech envelope can be extrapolated from data provided in studies investigating the cortical representation of temporal modulations of conspecific animal calls. Results from

marmoset (Wang et al., 1995; Nagarajan et al., 2002) and cat (Gehr et al., 2000; Gourevitch and Eggermont, 2007) indicate that a sizable population of primary cortical neurons phase-lock to the temporal envelope of conspecific communication calls. All studies investigating conspecific calls have analyzed single unit activity, or multiunit activity measured from relatively small populations of cortical neurons, and little is known regarding how the temporal envelope of conspecific calls is represented by ensemble population responses from cortical neurons in auditory cortex, which will be referred to here as multi-unit activity (MUA). MUAs provide an alternative method for examining neuronal representation of acoustic signals (Steinschneider et al., 1998; Steinschneider et al., 2003), and it has been shown that neuronal ensembles indicate different temporal properties relative to single units (Steinschneider et al., 1980).

Results from studies investigating ensemble cortical representation of periodic aspects of the speech signal have shown that MUAs phase-lock to periodicities up to ~200 Hz in auditory cortical neurons (Steinschneider et al., 1980). This result is in contract to the literature describing single unit spiking activity by auditory cortical neurons, which generally shows that phase-locked responses are limited to periodicities less than ~50 Hz (Lu et al., 2001). A limitation in the existing literature is that the stimuli used to probe the representation of periodicity in speech have been brief consonant-vowel stimuli produced by a synthesizer, and it has not been shown to what extent cortical populations represent periodicities in naturally produced speech sentence stimuli.

There are two major goals for this work. First, given the importance of the speech envelope for normal perception (Drullman et al., 1994a), a major goal of the current work was to examine the representation of the speech envelope by neuronal ensembles in auditory cortex. A second goal of this work was to examine the representation of periodic aspects of the speech signal using naturally produced speech sentences as a means to verify findings from studies using isolated consonant-vowel stimuli. To address the representation of speech envelope and periodicity in cortical neurons, we measured responses to a speech sentence produced in three different modes of speech: conversational speech, "clear" speech (Uchanski, 2005) and compressed speech (Watson et al., 1990; Abrams et al., in press). We used these three modes of speech as a means to vary speech envelope cues in an ecologically-relevant fashion: clear speech is a mode of speech naturally produced by talkers in noisy environments, and, among other acoustical variations, is characterized as having enhanced amplitude modulation depth relative to conversational speech (Uchanski, 2005). Compressed speech, which approximates rapidly produced speech, has faster speech envelope cues relative to conversational speech. Furthermore, as a means to vary periodicity cues in our stimulus set, we used relatively long sentence stimuli (1.5 sec) which exhibited natural fluctuations in F0 within and across utterances.

We measured neural representation to these speech stimuli in primary auditory cortex and an adjacent non-primary cortical field, the ventral caudal belt (VCB), in anesthetized guinea pig. The justification for measuring from primary auditory cortex is that the vast majority of the literature describing neural representation of isolated acoustic features in speech and conspecific communication calls has been performed in primary auditory cortex. We also

measured from a non-primary auditory cortical field, the ventral-caudal belt region, which is adjacent to primary auditory cortex in guinea pig (Redies et al., 1989b). A recent study showed that VCB neurons are sensitive to slow click rates (2-8 Hz), and consequently it was proposed that slow rates present in complex acoustic stimuli (i.e., speech) may be represented by these neurons (Abrams et al., submitted-b). It is not known how neurons in VCB would respond to the complex acoustic features inherent to the speech signal, and this aspect of the study was exploratory in nature.

Methods

The research protocol was approved in advance by the Animal Care and Use Committee of Northwestern University.

Animal preparation

The experimental materials and procedures were similar to those reported previously (McGee et al., 1996; Cunningham et al., 2002; Abrams et al., submitted-b). Seven pigmented guinea pigs of either sex, weighing between 325-650 g, were used as subjects. Animals were initially anesthetized with ketamine hydrochloride (100 mg/kg) and xylazine (8 mg/kg). Smaller supplemental doses (25 mg/kg ketamine; 4 mg/kg xylazine) were administered hourly throughout the rest of the experiment. Following the induction of anesthesia, the animal was mounted in a stereotaxic device for the duration of the experiment. Body temperature was maintained at 37.5° C by using a thermistor-controlled heating pad on the guinea pig's abdomen (Harvard). Normal hearing sensitivity was confirmed by auditory brainstem response (ABR). The ABR was elicited by a click stimulus at 70 and 40 dBHL

(referenced to normal guinea pig click thresholds) from a recording site located at the posterior vertex/midline of the scalp using an EMG needle electrode. A rostro-caudal incision was made along the scalp surface and the tissue was retracted to expose the skull. Holes were drilled in the skull under an operating microscope. The dura was removed with a cautery to prevent damage to the recording electrode, and the cortical surface was coated with mineral oil.

Anatomical structures

We measured neural activity from adjacent structures of auditory cortex: primary auditory cortex, which in guinea pig consists of two areas, A1 and the dorso-caudal field (DC), and a non-primary cortical region, the ventral caudal belt of cortex (VCB). Primary cortical responses are characterized by tonotopic organization, sharp frequency tuning, a preference for tonal stimuli and short response latencies (Redies et al., 1989a; Wallace et al., 2000). Neurons of VCB show broad frequency tuning, are more responsive to noise compared to pure tones, and have long-latency responses (Wallace et al., 2000).

Acoustic stimuli

Stimuli consisted of the sentence stimulus “The young boy left home” produced in three modes of speech: conversational, clear and compressed speech modes (Fig. 1). These three modes of speech have different speech envelope cues and were used as a means to elicit a variety of cortical activation patterns. Conversational speech is defined as speech produced in a natural and informal manner. Clear speech is a well-described mode of speech resulting from greater diction; clear speech enables greater speech intelligibility in noisy acoustic

environments relative to conversational speech in part due to greater amplitude modulation (i.e., speech envelope cues) (Uchanski, 2005). Compressed speech replicates rapidly-produced speech and is characterized by a higher-frequency speech envelope. Compressed speech is more difficult to perceive compared to conversational speech (Beasley et al., 1980).

Conversational and clear sentences were recorded in a soundproof booth by an adult male speaker at a sampling rate of 16 kHz. Conversational and clear speech sentences were normalized for overall duration to control for slower speaking rates in clear speech (Uchanski, 2005). Normalization was achieved by compressing the clear sentence by 23% and expanding the conversational sentence by 23%. To generate the compressed sentence stimulus, we doubled the rate of the conversational sample using a signal-processing algorithm in Adobe Audition (Adobe Systems Inc.). This algorithm does not alter the fundamental frequency or perceived pitch of the signal. The duration of the clear and conversational speech sentences was 1500 msec, and the duration of the compressed sentence was 750 msec.

30 repetitions of each speech stimulus were delivered to the contralateral ear using Etymotic insert earphones (ER2) through the hollow earbars of the stereotaxic device. The sound pressure level (SPL, expressed in dB re 20 mPa) was calibrated over a frequency range of 20-20 kHz using a condenser microphone (Brüel and Kjaer). Sentence stimuli were delivered at a level of 60 dBA. Stimuli were presented with alternating polarities to remove any possibility of a stimulus artifact within the response. The delivery system output the signal through a 16-bit converter at a sampling rate of 16 kHz. That system triggered the PC-based

collection computer. All stimuli were presented in quiet in a sound-treated booth (IAC). Third-octave tone-pips were used to map auditory cortex. Mapping of auditory cortex was essential to properly locating the paralemniscal cortical nucleus. Tones were 100 msec in duration with a rise-fall time of 10 msec.

Neurophysiologic recording

Cortex was accessed with a vertical approach using tungsten microelectrodes (Micro Probe) with impedance between 1-2 MΩ at 1kHz. An electrode was advanced perpendicular to the surface of cortex using a remote-controlled micromanipulator (Märzhäuser-Wetzlar). The coordinates of the electrode were determined at a point slightly above cortex at the first penetration, and these coordinates were kept for the remainder of the experiment.

For recording lemniscal areas A1 and DC, locations were approximately 3 mm caudal to bregma and 10 mm lateral of the sagittal suture. Recordings were made at depths of 500-900 μm, corresponding to cortical layers III and IV in guinea pig. Auditory cortex was mapped using third-octave tones to enable description of best frequency for lemniscal penetrations. To determine the location of paralemniscal area VCB, we used the frequency map of lemniscal regions A1 and DC to identify the ventral border of area DC which abuts area VCB. VCB was easily identified when the electrode had entered this cortical region: neural responses became significantly later, showed negligible and non-frequency-specific responses to pure-tones, and moderate-to-large responses to noise bursts (Wallace et al., 2000).

The electrode signal was amplified using Grass amplifiers with filters set between 1-20,000 Hz. The analog signal was digitized at 33.5 kHz by an A-D card (MCC) attached to a PC-based computer system. Responses were logged and stored using Matlab routines designed by our lab. Recorded brain responses were averaged and off-line filtered between 500-3000 Hz to isolate multiunit activity (MUAs; e.g., Steinschneider et al., 1990). A caveat to using this measure of MUA is that there is no way to determine the different cell types contributing to the overall response (Fishman et al., 2000). While the goal is to measure exclusively the contribution of cortical pyramidal cells, it is likely that MUA recordings also include contributions from stellate cells and thalamo-cortical axons, the latter of which are known to phase-lock to faster stimulus rates than pyramidal neurons.

Data analysis

All data analyses were performed using software written in Matlab (The Mathworks, Inc).

Periodicity analysis, frequency domain: We calculated the spectrum of MUAs for specific time segments of the response based on the presence of pronounced periodicity in MUAs, and the spectrum of corresponding time segments of the stimuli were also calculated. The spectrum was calculated using a 20000 point fast Fourier transform (FFT) and a rectangular window. In all stimulus and MUA FFTs, there were peaks at ~100 Hz corresponding to the F0 of the stimulus, and the frequency and magnitude of the FFT maxima were recorded.
Statistical analysis: A one-way ANOVA was performed to determine if there were statistical differences in MUA periodicity representation between clear, conversational and compressed speech conditions.

Periodicity analysis, time domain: For those segments of MUAs that indicated pronounced periodicity, we analyzed the latency difference between local maxima in the MUAs. Local maxima were picked automatically and were subsequently verified visually. A similar procedure was performed for corresponding segments of the speech stimuli.

Speech envelope extraction: A goal of this work was to investigate whether auditory cortical neurons, across the tonotopic map, phase-lock to temporal features in narrow or wide-band segments of the stimulus. To address this question, we extracted the temporal envelope from the stimulus across a number of stimulus bandwidths including: (1) 3 octaves; (2) 2 octaves; (3) 1 octave; (4) 1/2 octave; (5) 1/3 octave; (6) 1/8 octave (7) 1/10 octave (Fig. 25). An important consideration is that we used each cortical site's best-frequency as the center frequency for these filtering conditions. The steps involved in extracting the envelope across these band-width conditions were the following. First, the best-frequency of the cortical site was identified using the results from the pure-tone probes. The raw stimulus waveform was then filtered at the above listed band-widths using the best-frequency of the cortical site as the center frequency. The band-pass filters were Butterworth filters with 30 dB/octave rolloffs. The upper cutoff frequency for each band-width condition was calculated according to the equation $2^{bandwidth/2} \times BF$ and the lower cutoff frequency was calculated according to the equation $BF / 2^{bandwidth/2}$ (Everest, 2001). The Hilbert transform of the temporal envelope was then extracted from each of the band-filtered stimuli by calculating the absolute value of the Hilbert transform. The resulting waveforms were then low-pass filtered at 50 Hz to

extract the slowly-varying temporal envelopes of the speech signal (Rosen, 1992). The widest band envelope never exceeded the highest frequency in the speech signal.

Speech envelope cross-correlation analysis, primary MUAs: We performed cross-correlations between primary MUAs and the 7 iterations of speech envelopes described above. We automatically selected the peak in the resulting correlograms for lags between 10-50 msec to account for conduction and propagation delays in the auditory system.

Non-primary cortex, spike triggered stimulus analysis: The goal of this analysis was to address what stimulus features elicit neural activity in VCB neurons. First, for each non-primary MUA, we set a threshold based on the mean of the prestimulus MUA + 1.5 SD and identified all instances in the poststimulus period in which MUAs exceeded this threshold, and the times of these instances was recorded. We then identified the corresponding times in the stimulus and plotted the preceding 15 msec of the stimulus.

Speech envelope cross-correlation analysis, non-primary MUAs: We performed cross-correlations between non-primary MUAs and the 7 iterations of speech envelopes described above. Since a well-known characteristic of non-primary auditory cortical nuclei is that they do not show spectral frequency preference (Merzenich et al., 1975; Wallace et al., 2000), we were not able to use the cortical site's best-frequency as the center frequency for stimulus envelope filtering as we did for primary MUAs. In lieu of a best frequency, and as a means to be as thorough as possible in this analysis, we used a variety of center frequencies for speech envelope filtering, and cross-correlated all of these filtered versions of the speech

envelope to non-primary MUAs. Specifically, we used 125, 250, 500, 1k, 2k, and 4kHz as center frequencies in this analysis. For each of these center frequencies, we extracted the temporal envelope from the stimulus across a number of stimulus bandwidths including: (1) 3 octaves; (2) 2 octaves; (3) 1 octave; (4) 1/2 octave; (5) 1/3 octave; (6) 1/8 octave (7) 1/10 octave (Fig. 25). Therefore, the cross-correlation analysis for the non-primary MUAs varied in two dimensions: (1) the center frequency used for envelope extraction and (2) the bandwidth of the envelope. We automatically selected the peak in the resulting correlograms for lags between 10-50 msec to account for conduction and propagation delays in the auditory system.

Results

Primary cortex MUAs and periodicity

The clear speech stimulus (top) and three representative MUAs measured in response to this speech sample are plotted in Figure 26. A prominent feature visible throughout the stimulus waveform is a high amplitude periodicity near 100-120 Hz. This periodicity represents the fundamental frequency of speech (F0), the acoustic feature that determines the pitch of a speaker's voice. Visual inspection of the MUA activity reveals a periodicity that seems to correspond to F0 of the stimulus, most prominent at latencies between 160-260 msec and 900-1000 msec (Figure 26, highlighted area). The correspondence between the period of the neural activity and the stimulus F0 is even more evident in the magnified views of the waveforms (Figure 26, right). This periodicity appeared consistently in MUAs regardless of the best frequency (BF) of the MUA.

We investigated the correspondence between stimulus F0 periodicity and the periodic MUA activity in two manners. First, we performed a frequency analysis of the stimulus and MUAs over the time ranges 160-260 msec and 900-1000 msec (Fig. 27a and 27b, respectively), as MUAs in these particular time ranges show substantial periodic activity. Results of this analysis indicate considerable overlap between the stimulus F0 and periodic activity in cortical MUAs over these time ranges. We then determined the peak frequency of the F0 and all MUAs over these time segments, as well as the frequency difference between stimulus F0 and periodic MUA, and a close correspondence between these peaks is evident (Fig. 27c). Considering that the stimulus F0 varied within this speech sample (~120 Hz for 160-260 msec segment; ~100 Hz for 900-1000 msec segment) these data suggest that cortical MUAs faithfully follows F0.

A second analysis was performed in the time-domain as a means to further describe the correspondence between stimulus F0 periodicity and periodic activity in MUAs. We analyzed the latency difference between peaks in the raw stimulus waveform and the MUAs (Fig. 28, top). Peaks in the speech waveform preceded peaks in the MUA by approximately 12 msec (Figure 28, top). Results indicate that the periodic activity present in MUAs largely represents the stimulus F0, however MUAs fail to represent all of the latency variations expressed in the stimulus waveform. The most obvious discrepancy between peak latency differences in the speech sample and MUAs is between peaks 3-5 (Fig. 28, bottom). During this interval, peaks in the stimulus waveform vary between 8.0 - 9.5 msec while MUA peaks occur in a relatively small range between 8.0 - 8.5 msec. This discrepancy suggests that

cortical MUAs are not simply tracking the peaks of the stimulus waveform, as this would have resulted in a closer correspondence between stimulus F0 and MUA interval latencies.

Similar to the clear speech condition, correspondences between periodic features of the raw stimulus waveforms and MUAs were also evident in the conversational (Fig. 29) and compressed speech (Fig. 30) conditions. Specifically, frequency analyses of the conversational and compressed stimuli and responses indicated a close correspondence between the stimulus F0 and periodic neural activity (Fig. 31 and 32 for conversational and compressed speech conditions, respectively). We performed an ANOVA to determine if F0 frequency tracking, defined here as the frequency difference between stimulus F0 and periodic MUA, was different between the three stimulus conditions over the segments highlighted in Figures 27, 31 and 32. ANOVA results indicate no statistical difference between F0 tracking between these stimulus conditions ($F_{2,67} = 0.178$, $P = 0.837$). Furthermore, time-domain analyses showed a consistent association in peak latency differences between the raw stimulus waveform and the peaks in the MUA (Figures 33 and 34). MUA peak latency differences did not deviate more than 0.5 msec with the exception of peaks at the beginning and end of the response segments.

In summary, results from the analysis across multiple speech conditions – clear, conversational, and compressed – suggest that the F0 in on-going speech sentences is represented with periodic neural activity in primary cortex. This representation is resistant to variations in the frequency of the F0, as well as to variations in temporal envelope characteristics across the varying speech conditions.

Primary cortex MUAs and the speech envelope

A second prominent temporal feature present in the speech stimuli is the low-frequency amplitude envelope, also known as the speech envelope. Previous work has shown that cortical responses phase-lock to the temporal envelope of complex stimuli, such as the temporal envelope of communications calls (Wang, 2000). In the current results, it appeared that the temporal envelope of the stimulus was represented in low-frequency temporal activation patterns (i.e., the envelope) of cortical MUAs. To investigate this possibility, the envelope of all MUAs was extracted to compare to the stimulus envelope. A characteristic of the speech envelope is that it varies between different pass-bands in a given stimulus, and it is not known if cortical neurons represent the envelope of narrow (Gourevitch and Eggermont, 2007) or wide-bands (Nagarajan et al., 2002) of speech. We investigated this question by extracting 7 iterations of the speech envelope from increasingly wide bands of the stimulus, using the best frequency of each MUA as the center frequency (Fig. 35, bottom plots). We performed cross-correlations between each envelope iteration and cortical MUAs and identified the maximum in the correlogram between 10-50 msec for each stimulus bandwidth. The MUA from the recording site of Figure 35 showed a clear preference for wider bands of the speech envelope, with the largest stimulus-response correlations for stimulus envelopes greater than 1 octave (Fig. 35, inset).

Not all MUAs showed a preference for wide-bands of the stimulus envelope. Another representative MUA showed no preference for any of the bandwidth conditions (Fig. 36). Preference for envelope bandwidth in the three stimulus conditions for all cortical MUAs is plotted in Figure 37. All bandwidth-correlation functions are either flat or upward sloping,

indicating either no preference or preference for wider bandwidths. Very few of these functions are downward sloping, which would indicate a preference for narrow-band envelopes. Mean correlation functions indicate slight increased preference for wide bandwidths in all three stimulus conditions, and this preference appears to be greatest in the clear stimulus condition (Fig. 37).

Non-primary cortex MUAs

So far, our results have shown that MUAs from primary auditory cortex reliably encode the fundamental frequency F0 of the speech stimulus, and tend to preferentially respond to wideband speech envelopes. Previous work on a non-primary auditory cortical region, the ventral caudal belt (VCB), has shown that it is sensitive to slow click rates (Abrams et al., submitted-b). Consequently, it was suggested that VCB may have a role in coding the slow temporal rates present in the speech envelope. To explore this possibility, MUAs from VCB were also recorded in response to the speech sentences.

The clear speech stimulus (top) and three representative MUAs measured in VCB to this speech sample are plotted in Figure 38. There were three basic response type measured in this nucleus to the speech sentences. One response type (Fig. 38, top) showed similar temporal characteristics as primary MUAs, although VCB response amplitudes were generally smaller than primary MUAs. These responses showed periodic representation of the F0 in the speech stimulus and evidence of envelope representation. A second response type (Fig. 38, middle) showed considerably different activation patterns, with very large activations at onsets and the offset of features in the speech sentences. In contrast to the

previous response type, there was virtually no periodic neural activity present in MUAs measured from VCB, but activity in these responses did appear to align with temporal features in the stimulus. A third response type present in VCB (Fig. 38, bottom) indicated activation patterns that appeared to have little correspondence to temporal features present in the speech sentence. Unlike the first two VCB response types, activation in this population of neurons was neither periodic nor temporally aligned to stimulus features. These three patterns of VCB activation were present and consistent across the three stimulus conditions (Figures 39 and 40 for conversational and compressed speech sentences).

A question raised by responses from the third population of neurons in VCB is what stimulus feature elicits neural activity in this population. To address this question, we performed a spike-triggered stimulus analysis on this population of neurons and compared results to the first population of neurons, which appeared to represent temporal characteristics of the stimulus more directly. For the first VCB response form (Fig. 41, left), MUA activity appeared to be elicited by periodic portions of the stimulus waveforms in a time-locked manner. The stimulus components that elicited this activity show a large amplitude peak ~8 msec prior to spikes. A frequency analysis of the segments of the stimulus waveforms indicated clear peaks at the F0 and harmonics of the fundamental, confirming that neural activity was elicited by periodic stimulus components. In comparison, it was not clear what features in the stimulus elicited neural activity in VCB response type #3. When a frequency analysis was performed on the stimulus segments that elicited activity in VCB population #3, results were similar to those described in population #1, with clear peaks for the F0 and harmonics. These phenomenon are shown in a second representative example (Fig. 41, right).

These results suggest that periodic portions of the stimulus elicit non-periodic activity in the third population of VCB neurons.

We also evaluated the representation of the speech envelope by VCB neurons by performing cross-correlations between the same envelope iterations described above and MUAs measured from VCB neurons. Since non-primary neurons generally do not indicate spectral frequency preferences (Merzenich et al., 1975; Wallace et al., 2000), we used a variety of center frequencies for speech envelope filtering, and cross-correlated all of these filtered versions of the speech envelope to non-primary MUA (see Data Analysis). Therefore, the cross-correlation analysis for the non-primary MUAs varied in two dimensions: (1) the center frequency used for envelope extraction and (2) the bandwidth of the envelope. MUAs from VCB showed poor representations of the envelope as measured by cross-correlation analysis. Not only did maximum correlation coefficients fail to peak above r=0.25, but all of these MUAs failed to show a preference for a particular bandwidth of the speech envelope (Fig. 42). This was the case across all predetermined center frequencies and bandwidth iterations.

Discussion

The goals of this study were to examine cortical representation of the speech envelope and periodicity cues in two auditory cortical areas in guinea pig using naturally-produced speech stimuli. The results are summarized as follows: (1) In primary auditory cortex, the speech envelope is represented by the envelope of population responses. (2) Ensembles in primary auditory cortex are variable in their representation of the speech envelope: some population

responses represented the speech envelope with excellent fidelity, measured by cross-correlation analysis, while others showed poor fidelity in representing this aspect of the signal. (3) Ensembles in auditory cortex tended to show more precise representation of wider bandwidths of the speech envelope compared to narrow-bands. (4) The fundamental frequency of the speech signal, a prominent periodic aspect of the speech signal, was represented robustly by neuronal populations in primary auditory cortex with phase-locked responses across all three speech conditions. (5) A non-primary auditory cortical field, the ventral-caudal belt, was heterogeneous in response to temporal features in speech: some cortical sites represented temporal features of the speech signal in a time-locked fashion while other sites did not.

Rationale for studying cortical speech representation in animal models

Our understanding of how the human central auditory system represents temporal features in speech has been facilitated primarily by studies that have examined central auditory coding of specific, rudimentary acoustic features present in speech sounds. A second experimental approach that has facilitated knowledge about central auditory representation of speech sounds is the investigation of neural representation of conspecific communicational calls in animals (Wang et al., 1995; Gehr et al., 2000; Nagarajan et al., 2002; Gourevitch and Eggermont, 2007). However, if the ultimate research objective is to understand auditory processing of speech sounds, it can be argued that studying near-field neural responses to speech in an animal model is a necessary method. The rationale for this approach is that, given the complexity of the speech signal and well-established non-linearity of the auditory system, the neural representation of isolated acoustic features present in speech cannot

predict the representation of the actual speech signal: acoustic features often occur simultaneously in the on-going speech signal and the complex interaction of features may result in different response characteristics than those predicted by simple stimuli. Furthermore, while conspecific communication calls share acoustic features with speech, the stimuli used in these studies are not sufficiently complex to model the speech signal (nor is this the stated intent of these studies).

An important caveat of the current work is that speech does not represent a behaviorally significant signal to a guinea pig, and consequently these auditory responses would not be expected to produce response enhancements that are known to result from auditory learning of complex acoustic signals (Kilgard and Merzenich, 1998a, b; Wang and Kadia, 2001; Beitel et al., 2003). Furthermore, anytime an animal model is used to investigate speech processing mechanisms in humans there is no way of knowing whether the same mechanisms are utilized in the human auditory system. It has been shown that temporal aspects of speech are represented similarly in human and monkey auditory cortices, and an assumption of the current work is that cortical mechanisms are conserved across mammalian auditory systems.

Comparison previous studies: periodicity

A major finding of the current work is that cortical ensembles phase-locked to the fundamental frequency of the speech signal. Phase-locking was evidenced both in the frequency domain, where there was a close correspondence between the f0 of the stimulus and the peak frequency in the cortical response, as well as the time domain, where it was shown that the latency difference between peaks in the cortical responses closely mimicked

those in the stimulus waveform. Ensemble phase-locking to the fundamental frequency of speech has been demonstrated in a number of previous studies. For example, this phenomenon has been shown by Steinschneider in studies examining the voice-onset time temporal characteristics of stop consonants in an awake primate preparation (Steinschneider et al., 1995; Steinschneider et al., 2003). These studies have shown that cortical and thalamocortical ensembles are able to represent the fundamental frequency of brief speech segments with phase-locked responses. The current work establishes this phenomenon using naturally-produced speech sentence stimuli. Moreover, results also indicates that ensemble phase-locking to the f0 occurs irrespective of variations in the speech envelope characteristics of the stimulus. For example, f0-locking was equally prevalent in the clear speech and compressed speech conditions, which have markedly different speech envelope characteristics.

The propensity of ensembles of cortical neurons to follow periodic acoustic signals at these relatively fast rates (~100 Hz) is not consistent with single unit data. Single unit studies in auditory cortex have repeatedly shown that phase-locking is extremely poor for periodicities > 50 Hz, even in awake preparations (Lu et al., 2001). Moreover, it has been shown that single units of auditory cortex represent periodicities > 50 Hz with an unsynchronized rate code (Wang et al., 2003). One possible explanation for this discrepancy in the literature is that ensembles are able to phase-lock to faster rates than single units because their activity represents a "volley" of activity from a host of single units, similar to what has been proposed for the auditory nerve. Unfortunately, it is not straightforward how the "implicit" mode of rate representation seen in single unit studies for faster periodicities might be related

to a volley of phase-locked activity; perhaps the cortical "volley" and implicit rate representation provide two separate mechanisms for the representation of the f0. Unfortunately, the idea that ensembles of cortical neurons are able to phase-lock to faster rates is generally not discussed in studies investigating single-unit activity in auditory cortex. The result is that, despite decades of research on the auditory system, there is no consensus as to whether phase-locking provides a viable code for representing the fundamental frequency of speech in the human auditory cortex. It is hoped that future studies will allow for a more comprehensive hypothesis regarding the cortical representation of behaviorally important periodicities that takes into consideration all of the available data, including both single-unit and population responses.

Comparison to previous studies: speech envelope

Another finding from the current study is that the speech envelope appeared to be represented by the envelope of population responses measured from auditory cortical neurons. Specifically, it was shown that ~50% of primary cortical sites showed correlation coefficients > 0.50 with respect to the speech envelope. To our knowledge, cortical representation of the speech envelope has never been investigated in an animal model, although a number of studies have investigated the cortical representation of conspecific animal calls in cat (Gehr et al., 2000; Gourevitch and Eggermont, 2007), marmoset (Wang et al., 1995; Nagarajan et al., 2002) and guinea pig (Wallace et al., 2005a; Wallace et al., 2005b). Conclusions from these studies are largely consistent with those presented here: many cortical neurons across the tonotopic axis represent the temporal envelope of complex communication calls with excellent fidelity. For example, in a study of the representation of marmoset twitter calls in

primary auditory cortex of marmoset, it was shown that peristimulus time histograms from a population of auditory cortical neurons showed strong correlations to the temporal envelope of the twitter calls (correlation coefficients ~0.75). Similarly, in responses to purr calls in cat, it was shown that 40% of the neurons in primary auditory cortex could be categorized as "envelope peak-tracking units." The current work adds to this literature by showing that many cortical ensembles showed a preference for wider bands of the speech envelope (1-3 octaves) compared to narrow bands (< 1 octave). This is an important distinction since the literature is somewhat conflicted about this issue. For example, it was stated in a recent study of cortical representation of cat meows that there is a strong relationship between the cortical responses and the envelope of the vocalization band-pass filtered around the characteristic frequency of the neuron (Gourevitch and Eggermont, 2007). Contradicting this statement, it has been shown that wide swaths of auditory cortex, including portions of auditory cortex that would not be predicted to be activated based on the frequency content of the stimulus, show synchronous response patterns in response to temporal features in marmoset twitter calls (Nagarajan et al., 2002). The current results support this latter view, and are the first to quantify the bandwidth of the envelope that is preferred by auditory neurons. It is hoped that future studies will be able to confirm this result in other animal preparations, and whether this is a phenomenon specific to population responses or whether this also occurs in single units.

Comparison to previous studies: non-primary cortical responses

An exploratory aspect of the current study was to investigate temporal responses to speech sentences in a non-primary cortical region, the ventro-caudal belt region of guinea pig

auditory cortex. The reason for measuring from the particular belt region in auditory cortex is that a recent study showed that VCB neurons are sensitive to slow temporal modulations in simple acoustic stimuli, and consequently it was proposed that slow rates present in complex acoustic stimuli (i.e., speech) may be represented by these neurons (Abrams et al., submitted-b). Results from the current study showed heterogeneity of response patterns in VCB following speech stimulation. The three major response forms included (1) ensembles that showed similar temporal characteristics as primary MUAs, although amplitudes were generally smaller than primary MUAs, (2) onset and the offset responders, and (3) populations that responded to periodic portions of the stimulus with non-periodic activity. None of these populations showed robust encoding of the speech envelope. The interpretation of these results is not straightforward. In the study examining cortical representation of slow temporal modulations (Abrams et al., submitted-b), it was shown that increases in click rates are represented in increases in response latency shifts measured from VCB. The stimuli in the previous study were periodic acoustic clicks, which are dramatically more simple stimuli than the speech stimuli employed in the current study. Therefore, one interpretive complication with respect to the current results is that it is not known how the introduction of a more spectrally complex signal like speech would affect the temporal response properties of VCB. Previous studies have shown that some non-primary auditory regions are be more responsive to more spectrally complex acoustic stimuli (Rauschecker et al., 1995), so there was reason to think that VCB neurons would be sufficiently activated by the spectrotemporally complex speech stimuli. It is hoped that future studies will address more basic response properties in VCB and other belt regions as a means to better describe the role of non-primary auditory regions in the coding of complex acoustic signals.

Cortical representation of multiple temporal features in speech

Results show that neural ensembles of auditory cortex are capable of coding multiple aspects of the speech signal simultaneously. Specifically, periodicity in the speech signal provided by the fundamental frequency is represented with phase-locked responses from populations of auditory neurons, while the slowly varying speech envelope is represented in the envelope of cortical responses. From one point of view, this finding might be considered unremarkable: the auditory neurophysiology literature might have predicted this phenomenon based on cortical representation of simple, isolated acoustic features such as periodicity (Steinschneider et al., 1990) and slow amplitude modulations (Joris et al., 2004). From another standpoint, however, this finding represents one of the truly astounding capabilities of the auditory system: the auditory system is adept at simultaneously resolving multiple temporal features in complex acoustic signals using discrete coding mechanisms. In the speech-perception literature, it has been shown that these particular temporal features are essential for various aspects of speech perception, including providing information about speaker identity (i.e., f0 periodicity) as well as providing critical information regarding the content of the signal (i.e., speech envelope). The contribution of the current study is that we have demonstrated how these multiple temporal aspects of speech might be represented in the human auditory system. It is possible that the cortical mechanisms described here represent a neural correlate to these important perceptual phenomena.

CHAPTER VIII: DISCUSSION

Summary

These studies addressed various aspects of temporal processing in the mammalian auditory system, with an emphasis on the processing of temporal cues in speech in the human auditory system. The central hypothesis for the human work was that temporal processing of discrete elements of speech is essential for normal perception of speech and impacts phonological systems necessary for normal reading acquisition. Results from the current studies support this hypothesis with respect to crucial aspects of the speech signal. First, we provided the first demonstration that right-hemisphere auditory cortex is dominant for temporal processing of the slow rates inherent to the speech signal, the speech envelope. The behavioral importance of speech envelope processing was demonstrated in the second study where we showed that RI is associated with abnormal cortical response patterns for the speech envelope, and that cortical responses correlated with measures of reading and phonological processing across all subjects. In the third study we showed functional connectivity between the auditory brainstem and cortex for the processing of rapid acoustic features in speech. Specifically, we showed that temporal acuity in the auditory brainstem predicts cortical asymmetry for rapid temporal feature in speech. Furthermore, we showed that cortical asymmetry for rapid speech features is related to measures of speech perception and standardized measures of phonological processing and reading. In the fourth study, we showed that there does not appear to be a functional relationship between the auditory brainstem and cortical asymmetry for the slow temporal features in speech.

With respect to the animal work, the hypothesis was that a fundamental property of the mammalian auditory system is that discrete neuronal populations represent specific temporal

features in acoustic signals. Results from the animal work support this hypothesis. In the fifth study, we provided evidence that non-primary pathways in the auditory system may be important for the processing of slow temporal rates in the auditory system. In the sixth study, we showed how ensembles of neurons in primary auditory cortex are able to simultaneously represent multiple temporal features present in the speech signal as a model for the unimpaired human auditory system.

Given the importance of temporal information for the processing of speech and other biologically important signals, there is great interest in understanding how the central auditory system encodes temporal information. Here, we examined two questions regarding central auditory representations of temporal information in acoustic signals. First, we tested an hypothesis from the somatosensory system of rat that states that non-primary pathways may be important for the encoding of low frequency temporal information. Second, we examined the ability of cortical neurons to simultaneously represent multiple temporal features in speech.

With respect to this first question, results from the current work support the somatosensory hypothesis by showing that a non-primary pathway in the auditory system appears to preferentially encode the low-frequency temporal information in acoustic signals (<10 Hz) with a latency code. This is an important finding for two reasons. First, the neural mechanisms underlying acoustic rate differentiation and discrimination is not known, and the current results provide an enticing candidate. Second, these data provide a compelling hypothesis for a functional role of non-primary pathways in the encoding of important biological signals like speech and communication calls: the encoding of slow temporal information. By analogy,

results suggest that non-primary pathways in the human auditory system may be important for the coding of the speech envelope.

With respect to the second question, we showed that ensembles of primary cortical neurons are able to represent the fundamental frequency of speech (f0), a prominent acoustic cue that provides information regarding speaker identity, and the speech envelope, which provides syllable pattern information in speech. Specifically, cortical ensembles phase-lock to the f0, and the envelope of the ensemble responses often show a close correspondence to the speech envelope of the signal. From one point of view, this finding might be considered unremarkable: the auditory neurophysiology literature might have predicted this phenomenon based on cortical representation of simple, isolated acoustic features such as periodicity (Steinschneider et al., 1980; Steinschneider et al., 1990; Steinschneider et al., 1994; Steinschneider et al., 1995; Steinschneider et al., 2003) and slow amplitude modulations (Joris et al., 2004). From another standpoint, however, this finding represents one of the truly astounding capabilities of the auditory system: the auditory system is adept at simultaneously resolving multiple temporal features in complex acoustic signals using discrete coding mechanisms. The contribution of the current work is that we have demonstrated how these multiple temporal aspects of speech might be represented in the human auditory system. It is possible that the cortical mechanisms described here represent a neural correlate to these important perceptual phenomena.

Temporal information processing in the human auditory system

With respect to the human work, results provide new insight into how the human auditory system is able to simultaneously represent multiple temporal aspects of the speech signal. Results

support Poeppel's hypothesis that a mechanism for processing temporal information in the human auditory system is the routing of rapid temporal information (20-40 Hz) to left-hemisphere auditory cortex and slow temporal information (3-6 Hz) to right-hemisphere auditory cortex (Poeppel, 2003). An important consideration is that these particular rates are important for normal speech perception: the range of frequencies described as "rapid" in this context is important because it corresponds to the range of temporal information in speech relevant for encoding formant transitions in stop consonants; the range of frequencies described as "slow" in this context is important for the perception of the syllable rate of speech. Results provide evidence for neural mechanisms underlying the perception of these temporal features.

Current results are the first to test the AST hypothesis in an impaired population, and results from RIs support this hypothesis under "good" listening conditions, represented as the clear and conversational speech conditions. With respect to the AST hypothesis, this finding is important because it provides evidence that this temporal processing mechanism is a general and robust property of the human auditory system that is even present in a population associated with speech perception deficits (Kraus et al., 1996; Bradlow et al., 2003). An interesting avenue for future research with respect to the AST hypothesis is investigating the AST hypothesis in other clinical populations such as the elderly. It has been hypothesized that there may be a general "slowing" of the central nervous system in the elderly, and a previous study showed that elderly individuals have abnormal cerebral asymmetry associated with rapid speech sound processing (Bellis et al., 2000). Therefore, an interesting question is whether a similar deficit would be found in rightward asymmetry for slow temporal processing shown here. This would investigate the extent to which slowing of the central auditory system in the elderly impacts rapid processing

of acoustic signals in specific or, whether this slowing has a general effect on the temporal processors specified in the AST hypothesis. One interesting possibility is that the elderly may have intact processing of slow temporal features in speech, and consequently may rely more heavily on the speech envelope than younger listeners.

In addition to the study of clinical populations and the temporal processing mechanisms described in the current work, there are also many important questions regarding this temporal processing mechanism in the unimpaired auditory system. For example, do left and right-hemisphere temporal processors represent separable neural mechanisms or a single mechanism whose component processors function in conjunction with one another? Does rightward asymmetry for slow temporal features relate to the right hemisphere's preference for spectral processing (Zatorre et al., 2002)? It is hoped that future studies can identify the acoustic, linguistic and cognitive factors that play a role in this temporal processing mechanism.

Functional connectivity in the human auditory system

Two of the studies described in the current work investigated functional connectivity in the human auditory system. Previous work in the human auditory system has shown important relationships between temporal acuity in the auditory brainstem and cortical processing of speech. First, it was shown that brainstem onset responses are related to the robustness of cortical responses in the presence of background noise (Wible et al., 2005). Second, it was shown that brainstem timing was related to cortical representations that reflect fine-grained acoustic change (Banai et al., 2005a). Here, we examined the extent to which temporal processing in the auditory brainstem correlated with cortical asymmetries for rapid and slow temporal features in speech.

Results from these studies present a dichotomy in the functional relationship between auditory brainstem and cortex: while a strong relationship was shown between auditory brainstem timing and the strength of leftward asymmetry for rapid acoustic processing, no relationship was apparent when brainstem timing was correlated to rightward asymmetry for slow temporal features. With respect to the previous studies investigating brainstem-cortical relationships (Banai et al., 2005a; Wible et al., 2005), it was indeed surprising that an important aspect of cortical function, the processing of the speech envelope, was not related to brainstem timing.

This finding has a number of interesting implications. First, it may be the case that rapid processing of the acoustic signal represents a greater challenge to the auditory system than processing of the slow speech envelope, and therefore more exquisite timing in the auditory brainstem may be necessary for rapid processing. Another interesting possibility with respect to these findings is perhaps it is the case that the left-hemisphere's well-established asymmetry for language processing (i.e., phonological and semantic processing) has a top-down influence on the strength of connections between left-hemisphere auditory regions and auditory brainstem nuclei. It is hoped that future studies may be able to explain the origin of this brainstem-cortex dichotomy for temporal processing in the auditory system.

Hierarchical models of speech perception

Current findings provide important information regarding the neurobiological foundation of language and hierarchical models of speech perception. Recent studies using functional imaging techniques have enabled a detailed description of the functional neuroanatomy of spoken language, and the accumulated results have yielded models of speech perception consisting of a

number of discrete processing stages, including acoustic, phonological and semantic processing of speech (Hickok and Poeppel, 2007; Obleser et al., 2007).

An interesting point regarding the preferential routing of slow temporal information to right-hemisphere auditory cortex is that hierarchical models of speech perception propose that all aspects of speech processing, including acoustic, phonological and semantic processing of the speech signal, are dominated by left-hemisphere auditory and language areas (Hickok and Poeppel, 2007; Obleser et al., 2007). Results from the current study are among the first to show that the right-hemisphere of cerebral cortex is dominant during speech processing, and this shown in both the unimpaired auditory systems as well as the RI auditory system under "clear" and "conversational" speech conditions. Results presented here contradict the conventional thinking that language processing consists of neural operations largely confined to the left-hemisphere of the cerebral cortex. Moreover, results from the current study show right-dominant asymmetry for the speech envelope despite these other well-established forms of leftward asymmetry.

A second interesting point with regards to models of speech perception is results support the notion that the anatomical basis of speech perception is initially governed by the component rates present in the speech signal. This raises a number of interesting question regarding hierarchical models of speech perception. What is the next stage of processing for syllable pattern information in right-hemisphere auditory areas? Does slow temporal information in speech follow a parallel processing route relative to phonological processing? Does it follow the same

processing route as the processing of rhythm in music? It is hoped that these questions will receive additional consideration and investigation.

A third interesting consideration with respect to hierarchical models of speech perception is that the current findings show normal patterns of rightward asymmetry in RI for the speech envelope in two speech conditions (clear and conversational speech). This result suggests that this temporal processing mechanism can exist and function in the absence of a robust phonological system, a result that suggests that, in good listening conditions, this mechanism may not be strongly influenced by up-stream phonological processing mechanisms described in hierarchical models of speech perception (24).

Theories of reading impairment

Despite decades of research, the neurobiological foundation of RI remains elusive. While there is near-universal agreement that many RIs suffer from a phonological deficit, whether or not phonological deficits are secondary to a more fundamental sensory deficit remains a source of debate (Rosen, 2003; Bishop, 2006; Goswami, 2006; White et al., 2006). An influential hypothesis (Tallal et al., 1998) poses that abnormal perception of rapid acoustic events present in speech (on the order of tens of msec) precludes normal development of phonological systems since many phonological contrasts rely on resolving acoustic events occurring on this time scale (Phillips and Farmer, 1990). A major finding of this study was that the normal pattern of right-dominant activation of speech envelope representation was symmetric in RI for the compressed speech condition, a finding that provides strong evidence that acoustic deficits in RI are not

isolated to rapid components of acoustic signals. Consequently, it is argued that the current results call for an expansion of the temporal processing hypothesis of reading-disabilities to encompass impairments for both rapid and slow acoustic features.

Some researchers who study RI do not believe that there is a causal relationship between sensory deficits, such as impairments for both rapid and slow temporal features in speech, and phonological and reading deficits seen in individuals with RI. These researchers believe that the failure for all RIs to exhibit auditory deficits indicates that these two factors cannot be related (White et al., 2006), and that the auditory impairments evident in ~35% of RIs can be attributed to a common biological factor that also affects reading and phonology, but are otherwise unrelated to these important language skills (White et al., 2006). It is argued that the current results do not support the "common biological factor" hypothesis based on two pieces of evidence. First, it was shown that cortical speech envelope representation correlated with measures of literacy and phonological processing; second, it was shown that individuals with strong left-hemisphere dominance for rapid acoustic processing had better scores on measures of phonological processing and reading compared to individuals with weak left-hemisphere dominance. Taken together, these two findings show a link between the relative strength of cortical processing of temporal information in the auditory system for both rapid and slow temporal features and reading and phonological abilities.

If these forms of temporal processing are unrelated to reading ability, results would indicate that the putative "common biological factor" that accounts for these two phenomena varies in severity for all readers, and that temporal processing and phonology are affected similarly at

points across this continuum, but are still unrelated. An arguably more convincing hypothesis is that temporal processing abilities vary along a continuum and affect phonological abilities and reading in a proportional manner, and representations for both rapid and slow temporal processing measured in the current study reflect important aspects of speech processing.

Conclusion

Although the studies described here have provided a host of novel information regarding temporal information processing in the central auditory system, there is still much work to be done. With respect to the unimpaired auditory system, it is hoped that the details surrounding the AST hypothesis (Poeppel, 2003) are addressed in future studies, with an emphasis on describing the functional relationship between the rapid and slow temporal processors described in this hypothesis. With respect to the RI population, it is hoped that the necessary work is done to properly address the question of causality between temporal auditory processing and phonological and reading ability. This work would likely involve a longitudinal study to examine auditory, phonological and reading abilities in a large group of children across many formative years. With respect to the mechanisms proposed in the animal component of this work, it is hoped that future studies may be able to describe the influence of auditory training on mechanisms described here. For example, if an animal is trained to discriminate between low-frequency rates, is it the case that non-primary latency representations would be preferentially strengthened relative to primary pathway representations? Nevertheless, results from the six studies described here have revealed a number of interesting mechanisms involved in temporal processing in the auditory system, and it is hoped that this work will inspire exciting new work in the field of central auditory research.

TABLES

	NL	RI	P
N	12	11	
Age	11.7 (1.6)	12.5 (1.9)	>0.25
IQ	128.6 (9.0)	99.0 (11.1)	<0.001
Reading	110.8 (9.0)	92.8 (8.7)	<0.001
Spelling	113.8 (11.4)	89.1 (9.9)	<0.001
Word Attack	120.0 (13.0)	95.0 (6.9)	<0.001
Auditory Processing	105.3 (11.6)	86.8 (9.8)	<0.001

Table 1. Normal and reading-impaired subject characteristics

Stimulus	Group	Hemisphere	Pearson's "r" (STE)	Lag in ms (STE)	Envelope-following amplitude (STE)
Clear	Normal	Left	0.282 (.055)	91.25 (7.2)	1.066 (.10)
		Right	0.450 (.035)	85.28 (5.3)	1.277 (.11)
	LD	Left	0.318 (.057)	88.49 (7.6)	1.096 (.11)
		Right	0.438 (.036)	101.52 (5.5)	1.328 (.12)
Conversational	Normal	Left	0.200 (.054)	95.28 (8.4)	1.083 (.13)
		Right	0.388 (.027)	93.06 (5.3)	1.428 (.12)
	LD	Left	0.251 (.057)	86.67 (7.3)	1.218 (.14)
		Right	0.348 (.028)	103.18 (5.5)	1.285 (.12)
Compressed	Normal	Left	0.404 (.067)	95.28 (8.4)	1.039 (.11)
		Right	0.646 (.054)	80.42 (6.7)	1.494 (.15)
	LD	Left	0.564 (.070)	90.30 (8.7)	1.542 (.12)
		Right	0.495 (.056)	96.82 (7.0)	1.475 (.15)

Table 2. Three physiologic measures of speech envelope representation

Values represent the mean and standard error for each measure, calculated across left (T3, T5, Tp7) and right-hemisphere (T4, T6, Tp8) electrodes

	NL	LD	P
N	30	37	
Age	10.6 (1.6)	10.0 (1.5)	NS
IQ	120.5 (13.7)	105.6 (14)	<0.01
Reading	114.4 (14.1)	93.7 (14.6)	<0.01
Spelling	112.9 (12.8)	92.8 (15.6)	<0.01
Word Attack	112.6 (13.2)	93.6 (14.3)	<0.01
Auditory Processing	98.8 (11.0)	89.5 (9.2)	<0.01
Incomplete Words	95.8 (12.5)	92.4 (11.1)	NS
Memory For Words	105.8 (14.9)	95.5 (11.5)	<0.01
Sound Blending	101.7 (12.4)	90.1 (11.8)	<0.01
Listening Comprehension	123.0 (16.7)	112.7 (18.3)	<0.05
Cross Out	111.9 (11.5)	106.6 (14.1)	NS
Just Noticable Difference			
/da/ - /ga/ (Hz)	95.6 (53.1)	124.8 (51.0)	<0.01
/ba/ - /wa/ (msec)	8.5 (2.9)	7.7 (2.9)	NS

Table 3. Normal and learning-disabled subject characteristics

	Quintile 1 ("Early ABR")	Quintile 2	Quintile 3	Quintile 4	Quintile 5 ("Late ABR")
N, total	13	13	13	13	13
N, NL	8	4	6	6	5
N, LD	5	9	7	7	8
Age	10.2 (1.4)	9.9 (1.5)	11.3 (1.7)	10.3 (1.4)	9.9 (1.5)
IQ	107.5 (15.8)	115.9 (16.0)	113.5 (12.1)	116.2 (15.8)	108.8 (19.2)
Peak V Latency	7.4 (0.2)	7.6 (0.1)	7.8 (0.1)	7.9 (0.1)	8.0 (0.1)
Peak A Latency	8.4 (0.1)	8.6 (0.2)	8.6 (0.1)	8.8 (0.2)	9.0 (0.2)
Peak O Latency	48.7 (0.8)	48.7 (0.4)	49.1 (0.5)	49.1 (0.5)	49.9 (0.7)
Asymmetry, 130-145 msec (μV)	-47.9 (38.1)	-58.0 (44.1)	-36.4(33.8)	-34.1(57.9)	-8.0 (28.7)

Table 4. ABR grouping

	Quintile 1 ("Strong Left Dominant")	Quintile 2	Quintile 3	Quintile 4	Quintile 5 ("Weak Left Dominant")
N, total	14	13	13	13	14
N, NL	9	6	6	6	3
N, LD	5	7	7	7	11
Asymmetry, 130-145 msec (μV)	-101.3 (26.1)	-51.0 (6.2)	-33.3 (3.8)	-17.1 (7.3)	18.5 (15.5)
Age (years)	10.2 (1.4)	9.9 (1.5)	11.3 (1.7)	10.3 (1.4)	9.9 (1.5)
IQ	107.5 (15.8)	115.9 (16.0)	113.5 (12.1)	116.2 (15.8)	108.8 (19.2)
Reading	112.0 (9.2)	94.7 (17.5)	105.0 (21.1)	100.8 (18.4)	101.8 (17.9)
Spelling	112.9 (11.9)	95.4 (19.9)	103.5 (16.9)	97.8 (18.7)	98.9 (16.3)
Auditory Processing	102.4 (9.2)	87.8 (8.3)	94.1 (11.2)	89.8 (8.5)	93.6 (12.4)
Incomplete Words	99.2 (11.2)	86.1 (11.2)	98.3 (9.3)	91.2 (12.6)	94.2 (10.9)
Memory For Words	106.4 (16.4)	91.2 (9.2)	99.8 (13.2)	100.8 (17.0)	101.9 (9.9)
Sound Blending	106.6 (12.5)	90.4 (7.8)	93.0 (13.0)	91.7 (10.7)	94.1 (15.7)
Listening Comprehension	123.6 (14.1)	112.1 (20.1)	121.4 (17.7)	111.1 (20.4)	117.9 (17.6)
Cross Out	109.4 (11.0)	115.3 (10.5)	107.6 (14.3)	112.3 (15.1)	101.2 (12.1)
Just Noticable Difference					
/da/ - /ga/ (Hz)	83.9 (31.1)	123.2 (47.3)	97.8 (48.2)	133.7 (78.7)	121.2 (43.5)
/ba/ - /wa/ (msec)	7.5 (2.8)	7.0 (3.2)	10.0 (2.8)	8.0 (2.8)	8.1 (2.7)

Table 5. Asymmetry grouping and measures of academic achievement

	2Hz	5 Hz	8 Hz
Clicks 2 and 3 LFP			
Peak Amplitude			
Lemniscal mean (SD)	30.526 (22.776) µV	28.623 (22.833) µV	26.692 (23.554) µV
Paralemniscal mean (SD)	11.621 (8.143) µV	8.702 (6.002) µV	6.035 (3.892) µV
Peak Latency			
Lemniscal mean (SD)	9.500 (1.175) ms	9.523 (1.159) ms	9.523 (1.173) ms
Paralemniscal mean (SD)	21.293 (3.751) ms	21.713 (3.943) ms	22.313 (3.640) ms
Half-Peak Latency			
Lemniscal mean (SD)	7.823 (1.032) ms	7.831 (1.0197) ms	7.806 (1.017) ms
Paralemniscal mean (SD)	18.887 (3.937) ms	19.213 (3.842) ms	19.527 (3.876) ms
Steady-State LFP			
Peak Amplitude			
Lemniscal mean (SD)	31.339 (22.773) µV	24.633 (17.023) µV	22.314 (15.531) µV
Paralemniscal mean (SD)	11.1 (8.050) µV	8.377 (4.594) µV	6.646 (3.465) µV
Peak Latency			
Lemniscal mean (SD)	9.474 (1.132) ms	9.549 (1.199) ms	9.517 (1.227) ms
Paralemniscal mean (SD)	22.193 (3.727) ms	21.813 (3.731) ms	22.34 (3.926) ms
Half-Peak Latency			
Lemniscal mean (SD)	7.843 (1.069) ms	7.8571 (1.068) ms	7.8286 (1.090) ms
Paralemniscal mean (SD)	18.893 (3.913) ms	19.200 (3.960) ms	19.560 (4.112) ms

Table 6. Thalamic LFP statistics

	2Hz	5 Hz	8 Hz
Clicks 2 and 3 LFP			
Peak Amplitude			
Lemniscal mean (SD)	227.21 (245.492) µV	301.33 (299.519) µV	171.09 (170.369) µV
Paralemniscal mean (SD)	128.14 (148.968) µV	157.09 (192.412) µV	115.16 (156.257) µV
Peak Latency			
Lemniscal mean (SD)	20.431 (5.272) ms	22.986 (6.892) ms	25.248 (7.329) ms
Paralemniscal mean (SD)	27.604 (6.806) ms	30.45 (7.111) ms	33.671 (7.823) ms
Half-Peak Latency			
Lemniscal mean (SD)	16.069 (3.973) ms	18.331 (5.168) ms	19.576 (5.101) ms
Paralemniscal mean (SD)	21.612 (5.392) ms	24.229 (5.814) ms	26.525 (6.609) ms
Steady-State LFP			
Peak Amplitude			
Lemniscal mean (SD)	249.98 (233.654) µV	160.31 (154.973) µV	111.06 (119.945) µV
Paralemniscal mean (SD)	134.1 (144.903) µV	182.36 (261.226) µV	55.82 (46.950) µV
Peak Latency			
Lemniscal mean (SD)	20.01 (5.207) ms	22.183 (5.941) ms	20.7 (5.602) ms
Paralemniscal mean (SD)	26.688 (6.756) ms	30.589 (9.107) ms	27.686 (9.515) ms
Half-Peak Latency			
Lemniscal mean (SD)	15.603 (3.714) ms	16.607 (4.055) ms	15.148 (3.211) ms
Paralemniscal mean (SD)	21.2 (5.472) ms	24.428 (7.962) ms	22.822 (8.173) ms

Table 7. Cortical LFP statistics

FIGURES

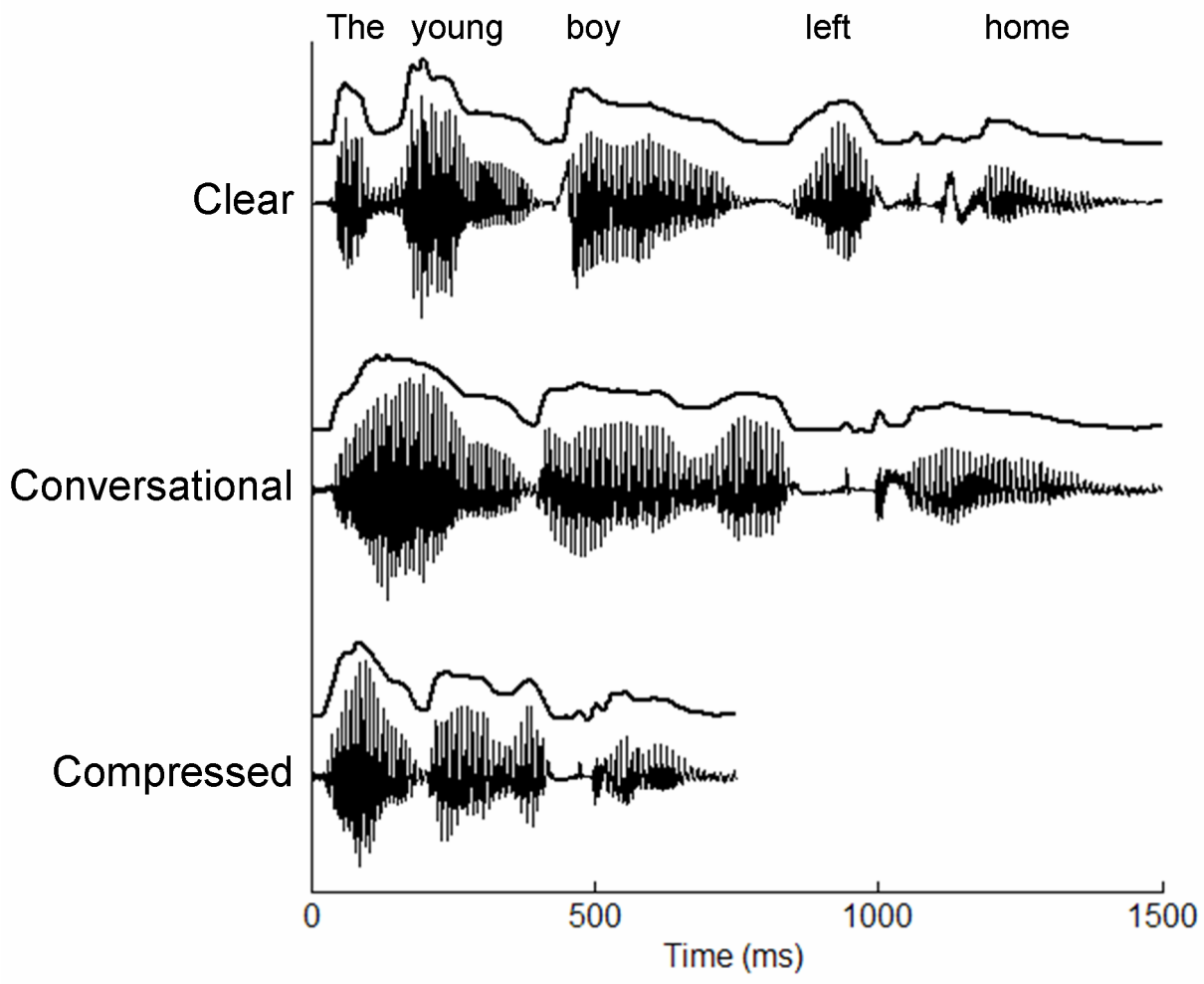


Fig. 1: Speech waveforms for the sentence "The young boy left home" in clear (top), conversational (middle) and compressed (bottom) speech modes. The speech envelopes for these stimuli are plotted immediately above each speech waveform.

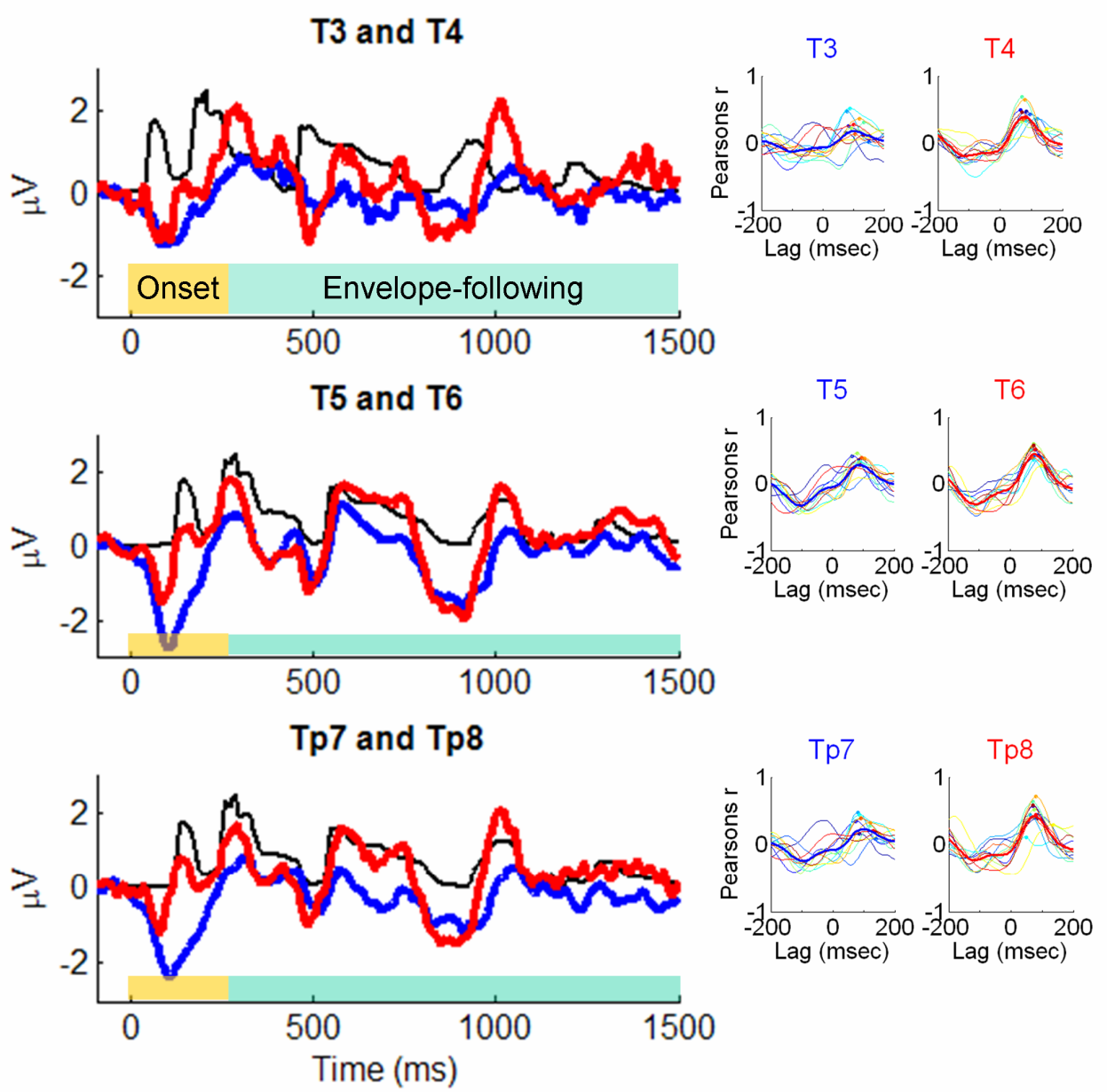


Fig. 2a-c, left column: Grand average cortical responses for three matched electrode pairs and broadband speech envelope for "clear" stimulus condition. The black lines represent the broadband speech envelope for the clear speech condition, the red lines represent cortical activity measured at right hemisphere electrodes and the blue lines represent activity from left-hemisphere electrodes. 95 msec of the pre-stimulus period is plotted. The speech envelope was

shifted forward in time 85 msec to enable comparison to cortical responses; this time shift is for display purposes only. Figure 2d-f, right column: cross-correlograms between "clear" speech envelope and individual subjects' cortical responses for each electrode pair. A small dot appears at the point chosen for subsequent stimulus-to-response correlation analyses.

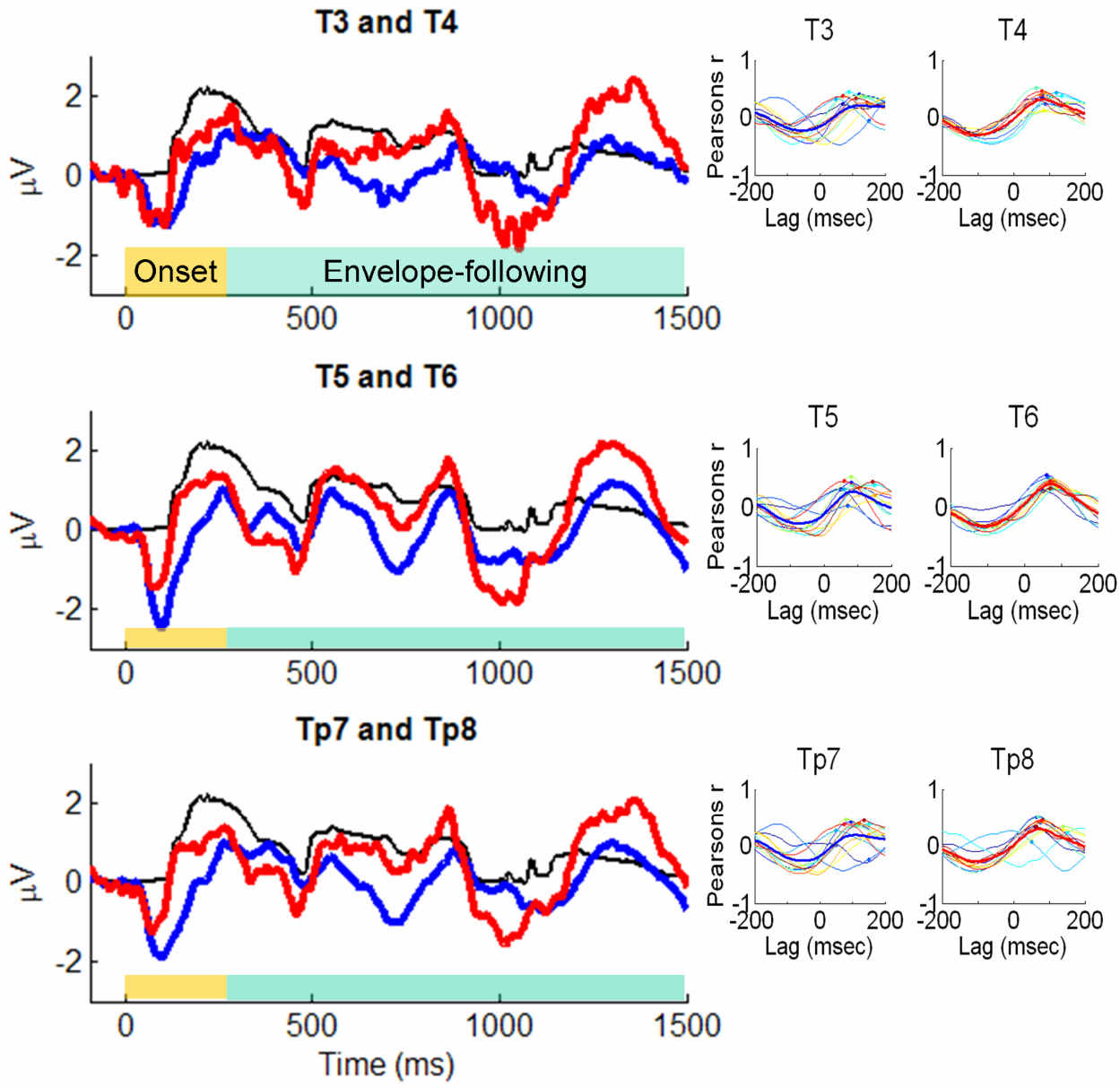


Fig. 3a-c, left column: Grand average cortical responses and broadband speech envelope for "conversational" stimulus condition. The black lines represent the broadband speech envelope for the conversational speech condition, the red lines represent cortical activity measured at right hemisphere electrodes and the blue lines represent activity from left-hemisphere electrodes. The speech envelope was shifted forward in time 85 msec to enable comparison to cortical responses.

Fig. 3, right column: cross-correlograms between “conversational” speech envelope and raw cortical responses.

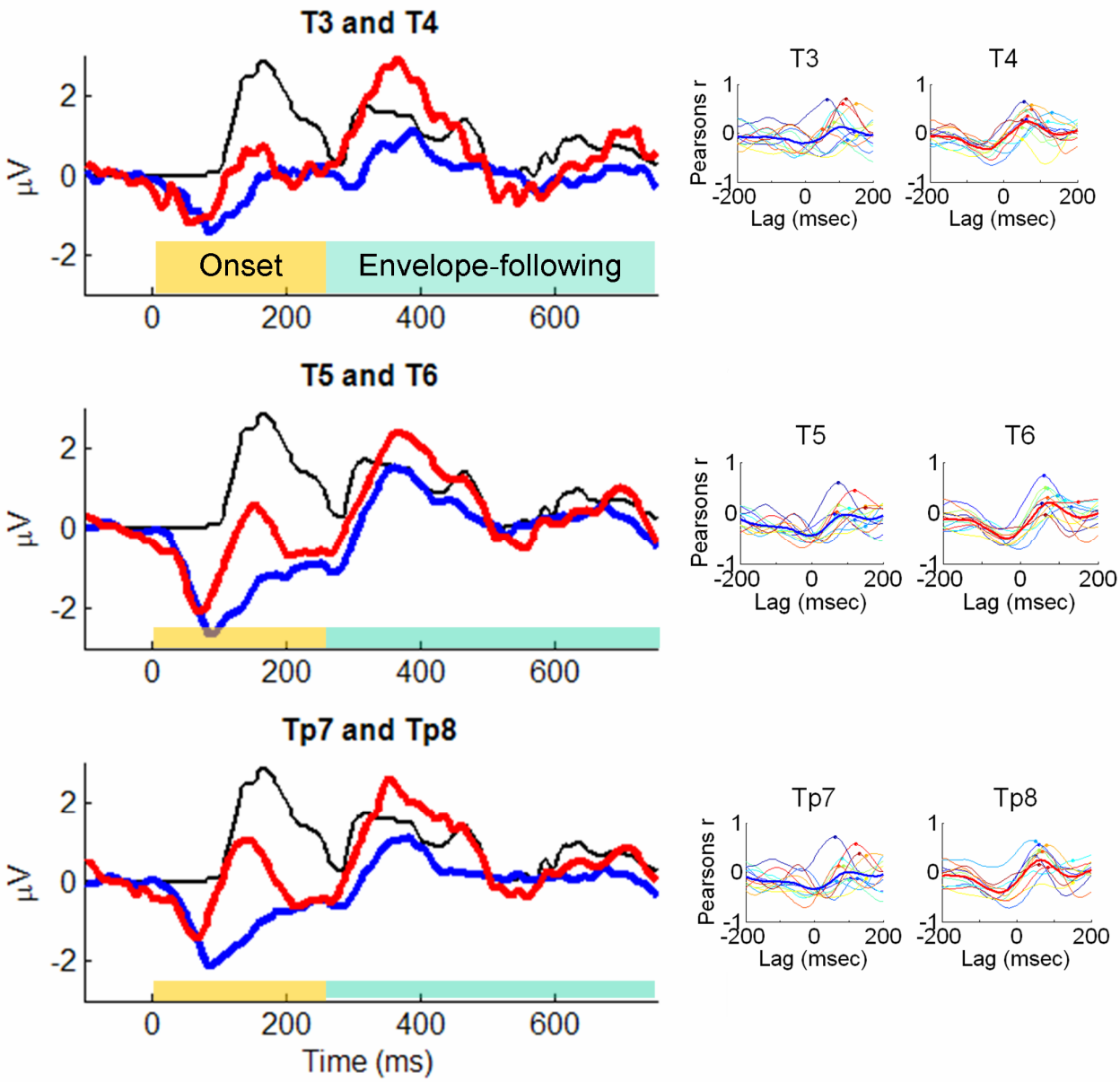


Fig. 4a-c, left column: Grand average cortical responses and broadband speech envelope for "compressed" stimulus condition. The black lines represent the broadband speech envelope for the compressed speech condition, the red lines represent cortical activity measured at right hemisphere electrodes and the blue lines represent activity from left-hemisphere electrodes. The

speech envelope was shifted forward in time 85 msec to enable comparison to cortical responses. Fig. 4, right column: cross-correlograms between “compressed” speech envelope and raw cortical responses.

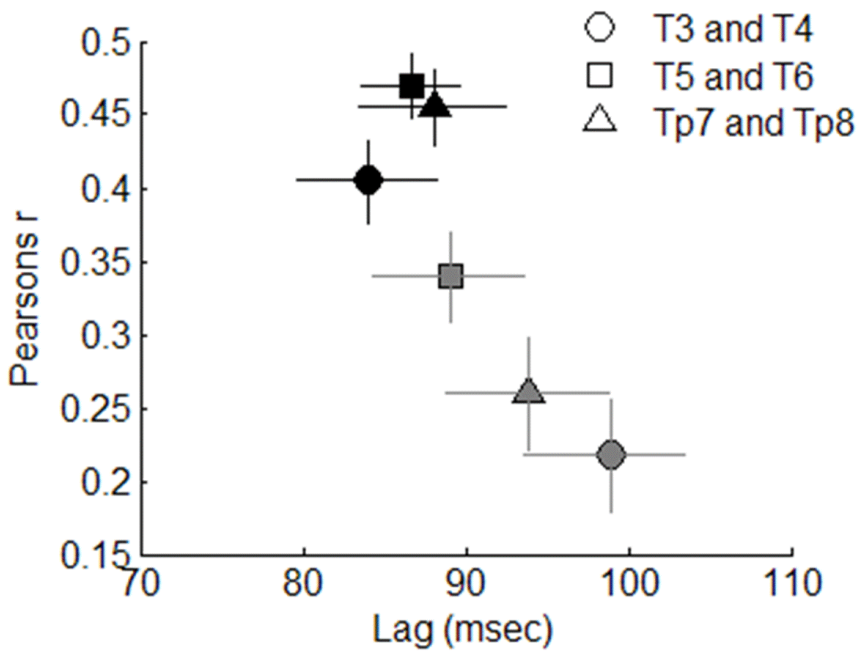


Fig. 5: Average cross-correlogram peaks. Values represent the average peak lag and r-value, collapsed across stimulus conditions, for each stimulus envelope – cortical response comparison at the 3 electrode pairs. Right-hemisphere electrodes are red and left-hemisphere electrodes are blue. Error bars represent 1 standard error of the mean.

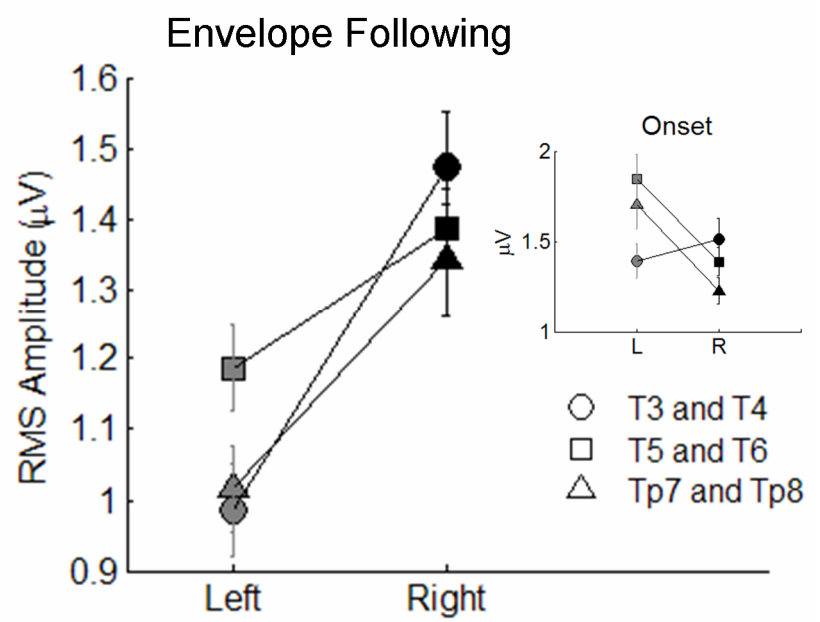


Fig. 6: Average RMS amplitudes for envelope-following and onset (inset) periods. Onset period was defined as 0-250 msec of the cortical response and envelope-following period was defined as 250-1500 or 250-750 msec. Error bars represent 1 standard error of the mean.

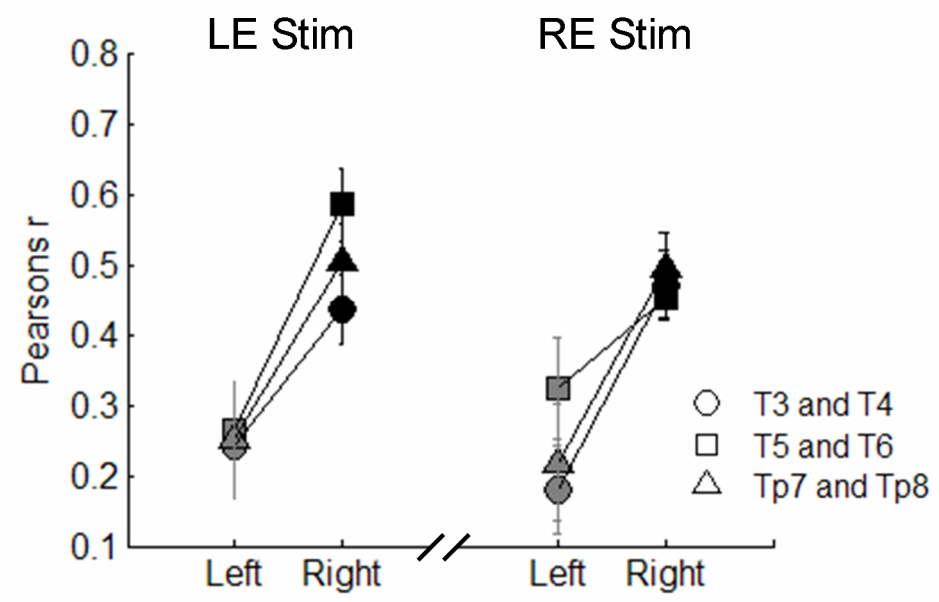


Fig. 7: Left-ear vs. right-ear stimulation comparison: precision of speech envelope phase-locking. Error bars represent 1 standard error of the mean.

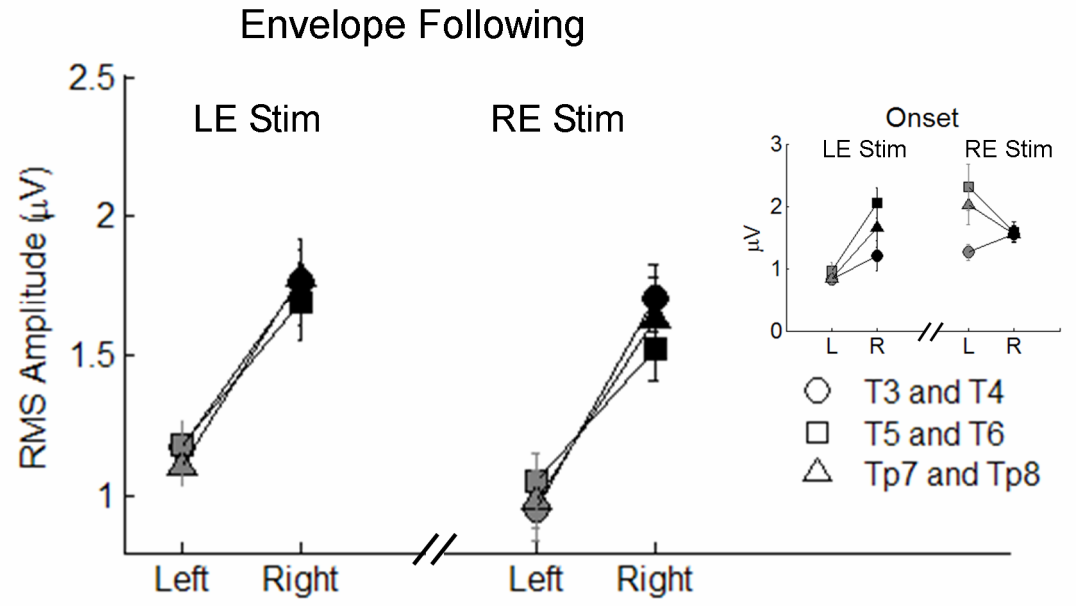


Fig. 8: Left-ear vs. right-ear stimulation comparison: RMS of the envelope-following period. Error bars represent 1 standard error of the mean. Figure Inset: RMS comparison of the onset period.

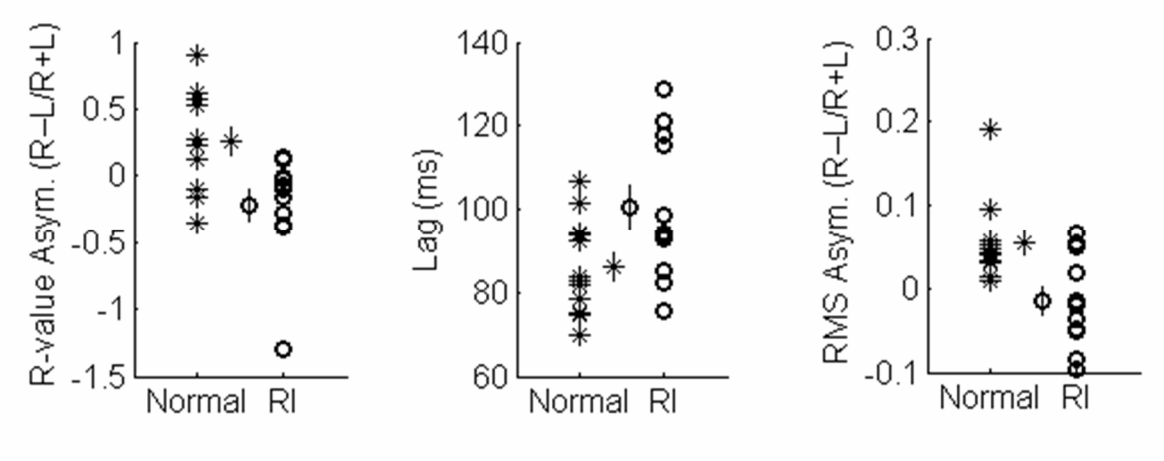


Fig. 9. Three measures of cortical speech envelope representation in all subjects. Measures consist of r-value asymmetry index in the compressed speech condition (left), right-hemisphere lag across stimulus conditions (middle), and RMS amplitude asymmetry index in the compressed speech condition (right). Mean and 1 standard error of the mean are represented in the middle of each plot.

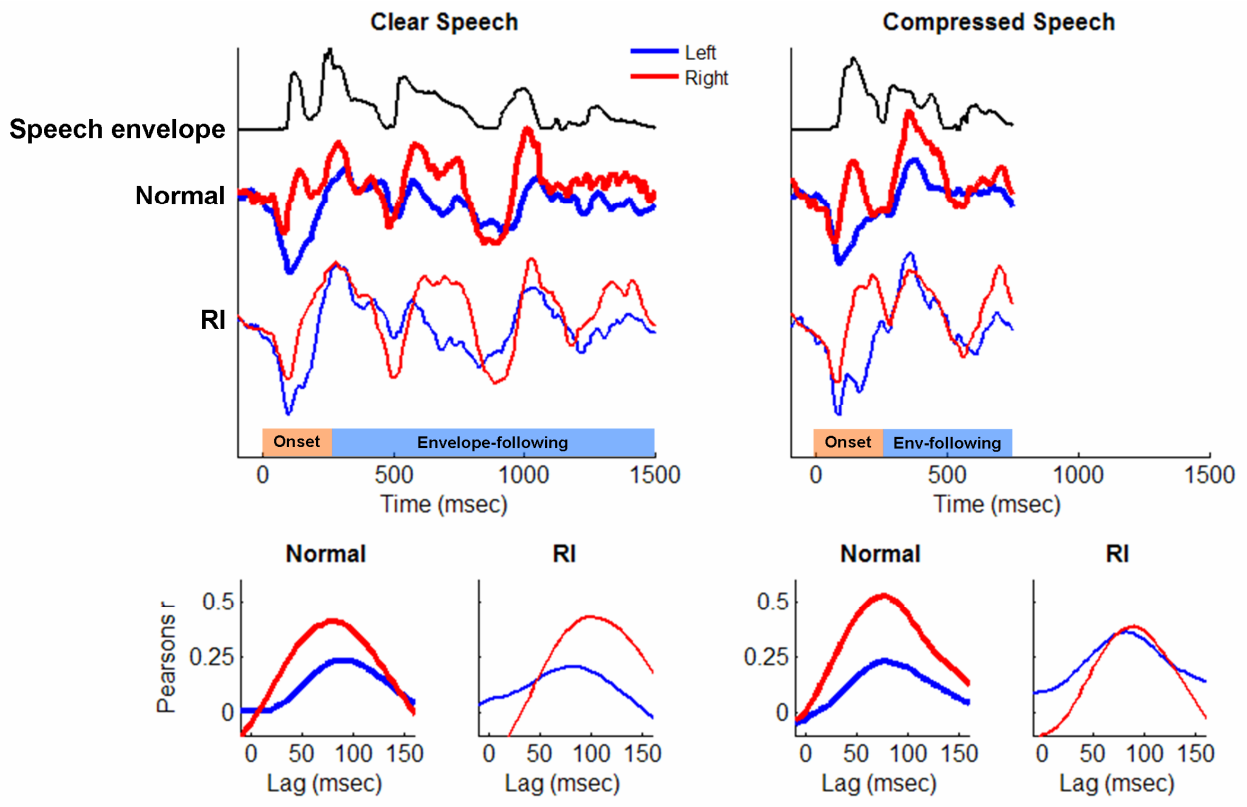


Fig. 10, top. Grand average cortical responses from left (Tp7) and right-hemisphere (Tp8) temporal lobe electrodes and the broadband speech envelope for "clear" (left) and "compressed" (right) stimulus conditions. The black lines represent the broadband speech envelope, red lines represent cortical activity measured at the right-hemisphere electrode and blue lines represent activity from the left-hemisphere electrode. 95 msec of the pre-stimulus period is plotted. The speech envelope was shifted forward in time 85 msec to enable comparison to cortical responses; this time shift is for display purposes only. Figure 10, bottom: mean cross-correlogram calculated between the speech envelope and individual subjects' cortical responses.

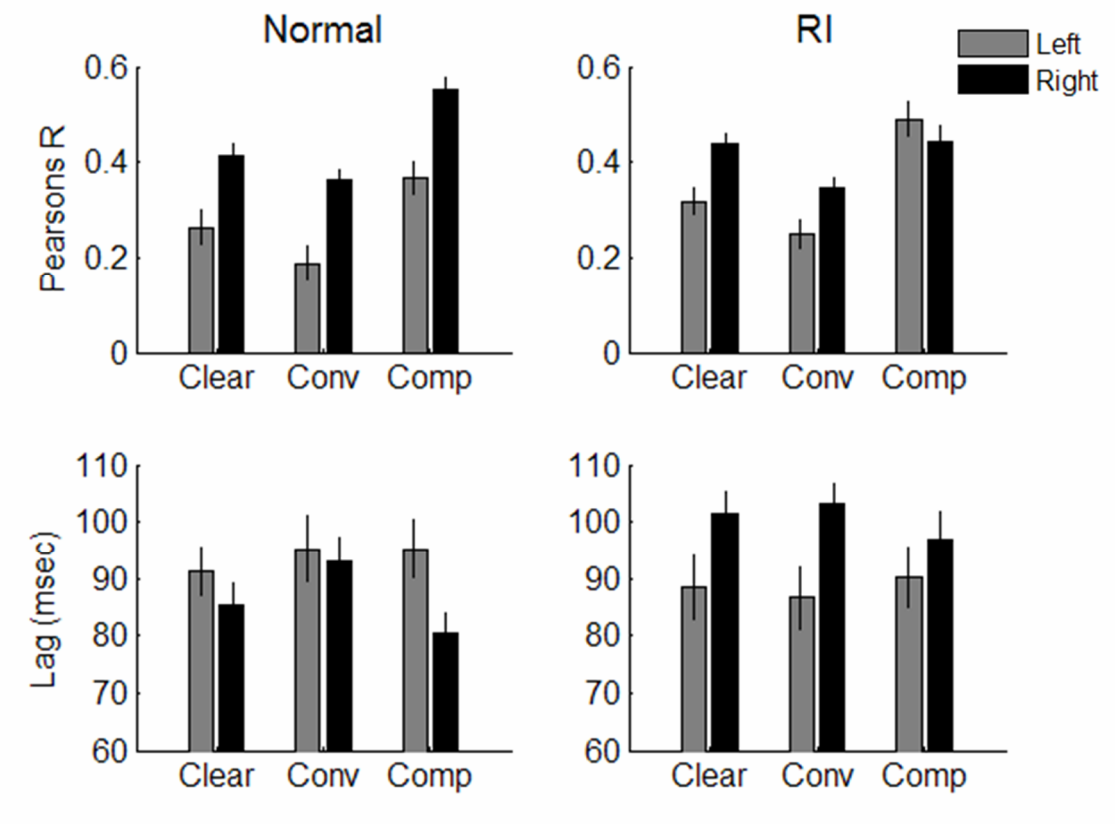


Fig. 11. Average phase-locking precision and timing values. Bars represent the mean peak Pearson's r-value (top row) and lag (bottom row) for each stimulus envelope – cortical response correlation. Values represent the average from left (gray; T3, T5, Tp7) and right-hemisphere (black; T4, T6, Tp8) electrodes. Error bars represent 1 standard error of the mean.

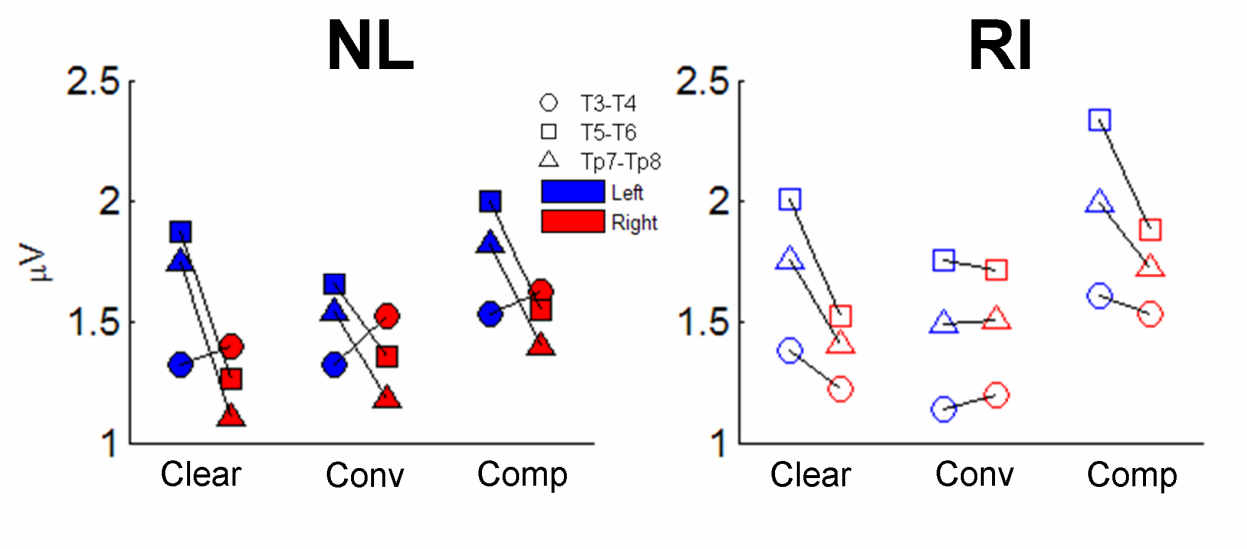


Fig. 12. Average RMS amplitudes for the onset period measures at six electrode locations. The onset period was defined as 0-250 msec for all stimulus conditions. Blue bars represent left-hemisphere electrodes and red bars represent right-hemisphere electrodes. Error bars represent 1 standard error of the mean.

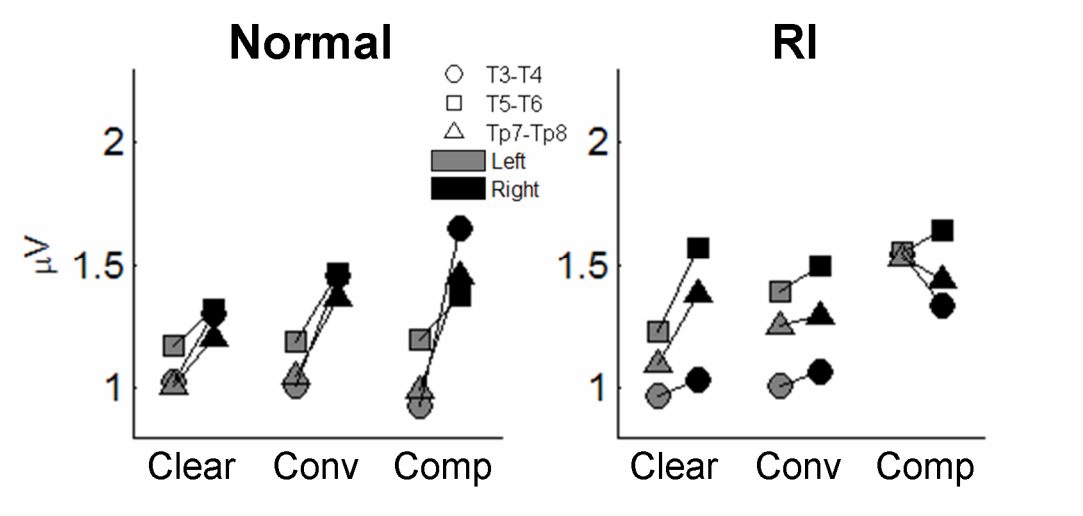


Fig. 13. Average RMS amplitude for the envelope-following period measured at six electrode locations. The envelope-following period was defined as 250-1500 msec (clear and compressed speech conditions) or 250-750 msec (compressed speech condition). Gray bars represent left-hemisphere electrodes and black bars represent right-hemisphere electrodes. Error bars represent 1 standard error of the mean.

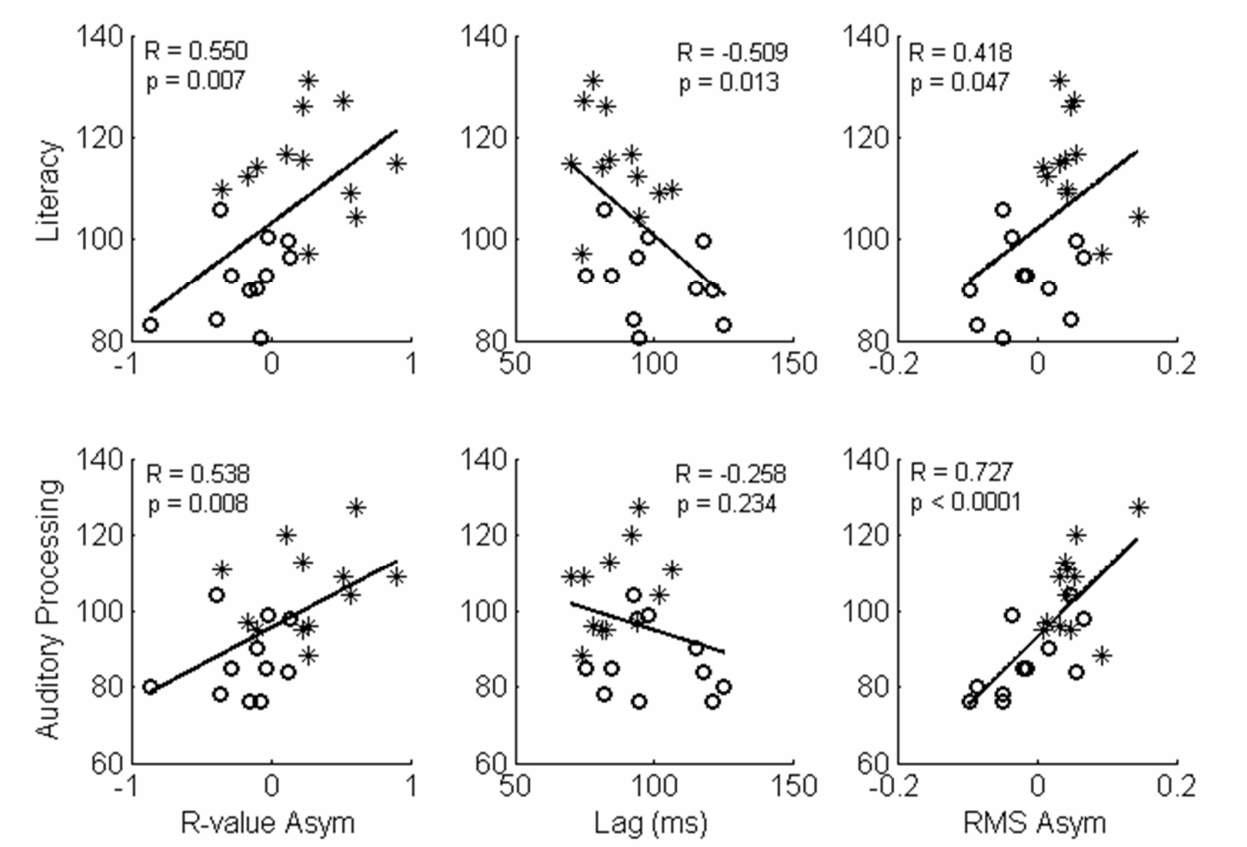


Fig. 14. Three measures of cortical speech envelope representation and measures of literacy (top row) and phonological processing (bottom row).

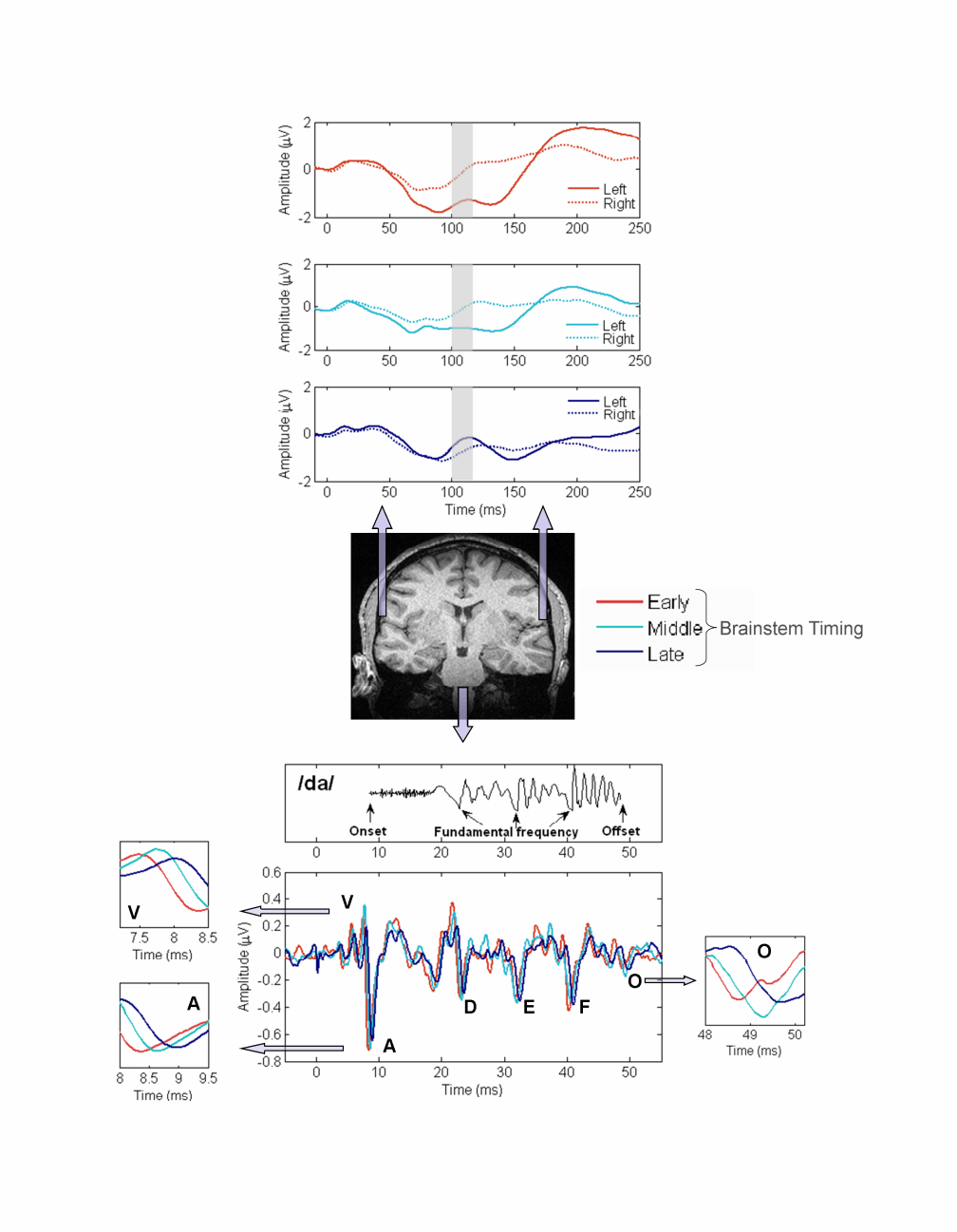


Fig. 15. Grand average neurophysiologic responses. (Bottom) Acoustic waveform of the synthesized speech stimulus /da/ (above) and grand average auditory brainstem responses to /da/ (below). The stimulus has been moved forward in time to the latency of onset responses (peak V) to enable direct comparisons with brainstem responses. Lower insets: 1st, 3rd and 5th quintile

responses for waves V, A and O. (Top) Grand average cortical responses measured from left and right hemisphere temporal electrodes, grouped by latency of ABR responses.

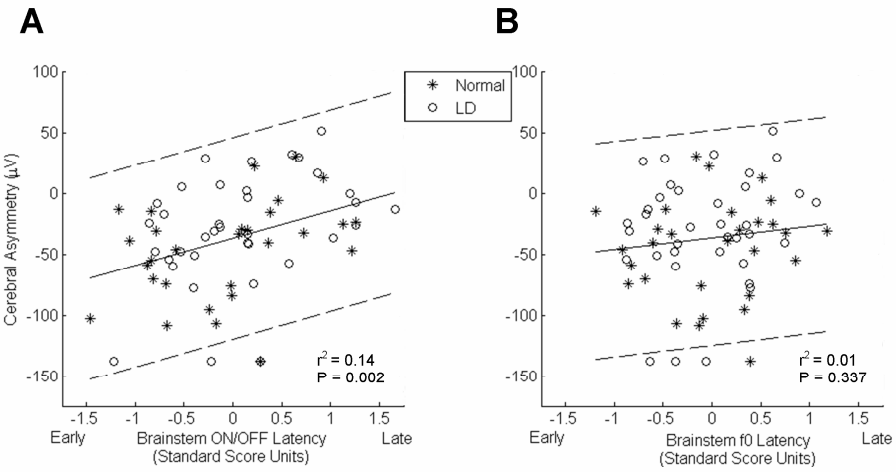


Fig. 16. Brainstem responses and cortical asymmetry. The linear fit for the data in both plots is indicated by the central line and is flanked by dashed lines indicating the limits of the range for prediction of individual data points with 95% certainty. (A) Brainstem onset/offset and cortical asymmetry. (B) Brainstem frequency following and cortical asymmetry.

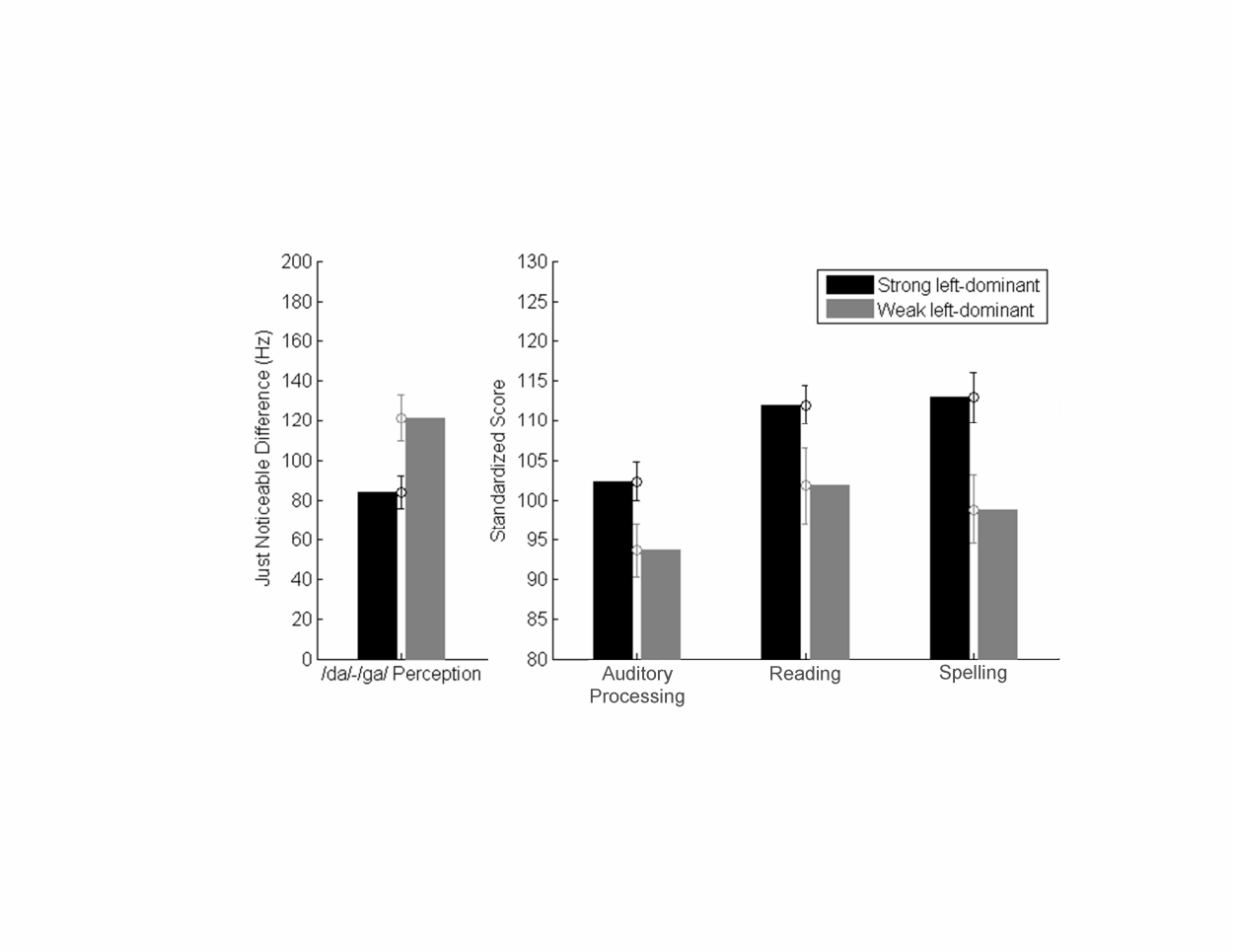


Fig. 17. Cortical asymmetry and measures of speech discrimination and academic achievement.

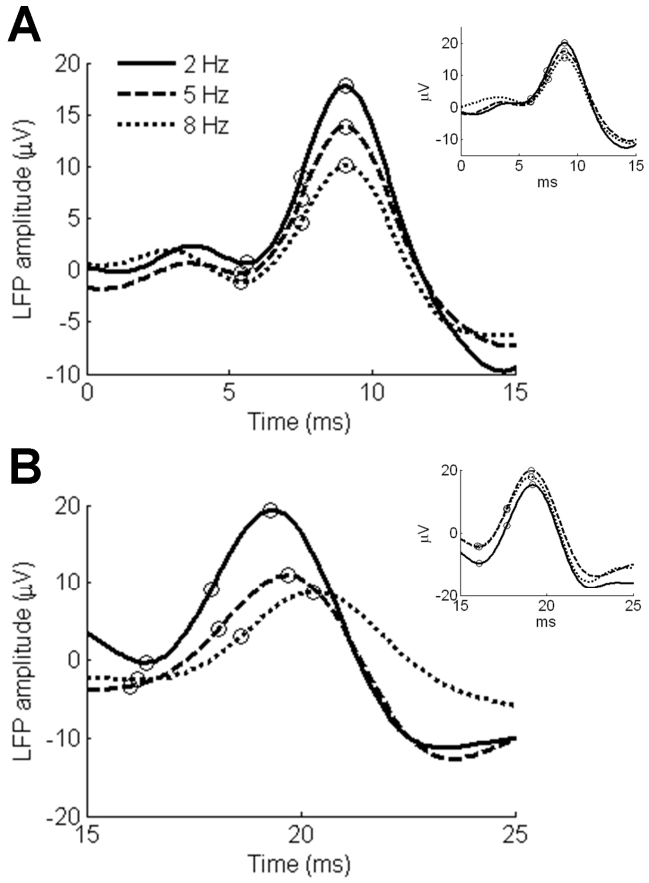


Fig. 18. Thalamic LFPS to click trains. (a) Representative lemniscal (MGv) and (b) paralemniscal thalamic (MGs) responses. Plots represent the average local field potential (LFP) in response to clicks 2 and 3 in a train. Figure 18 insets: plots represent LFPs in response to the first click in a train measured in MGv (Figure 18a inset) and MGs (Figure 18b inset). The circles on the plots indicate the peak, base and half-peak measures recorded for each LFP at each click rate.

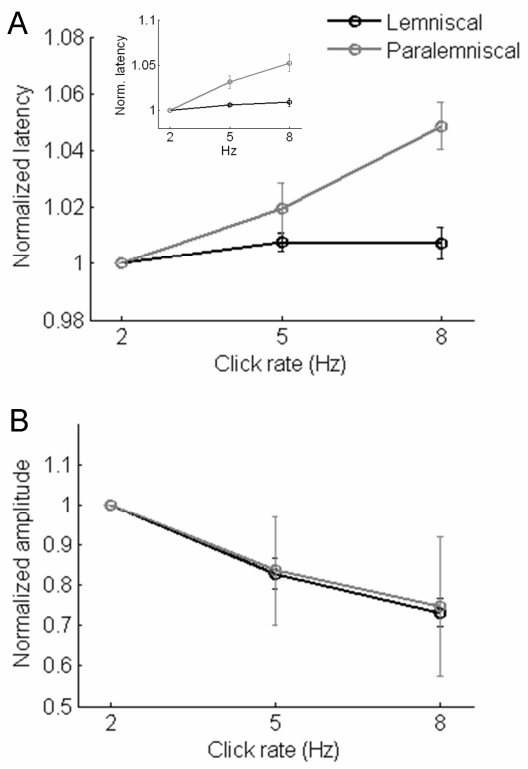


Fig. 19. Mean thalamic LFPs. Mean (± SE) local field potential (a) latency and (b) amplitude for lemniscal and paralemniscal thalamic responses. Figure 19a inset: mean "steady-state" thalamic LFP latencies.

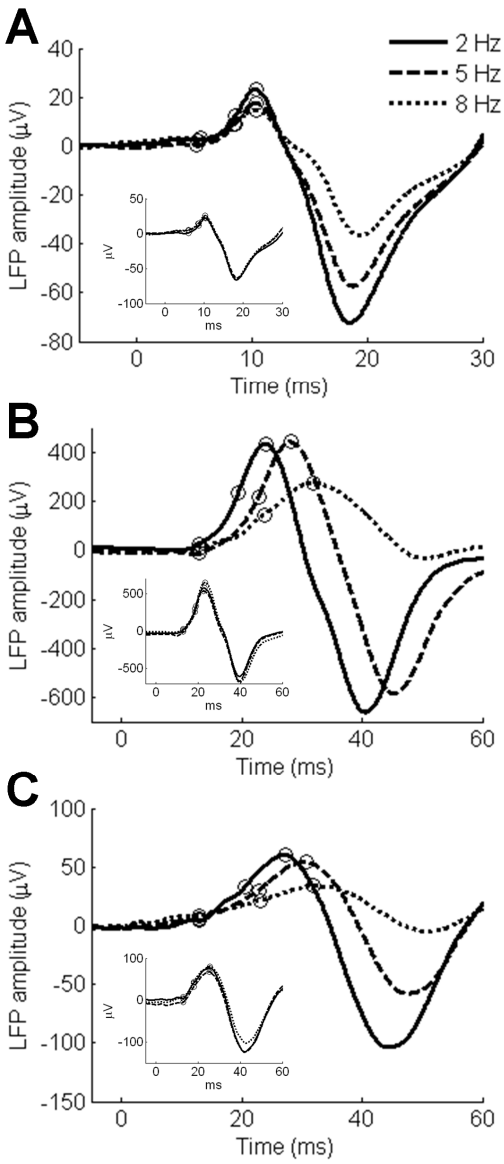


Fig. 20. Cortical LFPs to click trains**.** (a-b) Two representative lemniscal cortical responses (A1 and DC) and (c) a representative paralemniscal (VCB) cortical response. Plots represent the average local field potential in response to clicks 2 and 3 in a train. Figure 20 insets: plots represent LFPs in response to the first click in a train measured in lemniscal cortex (Figure 20a, 20b insets) and VCB (Figure 20c inset). The circles on the plots indicate the peak, base and half-peak measures recorded for each LFP at each click rate.

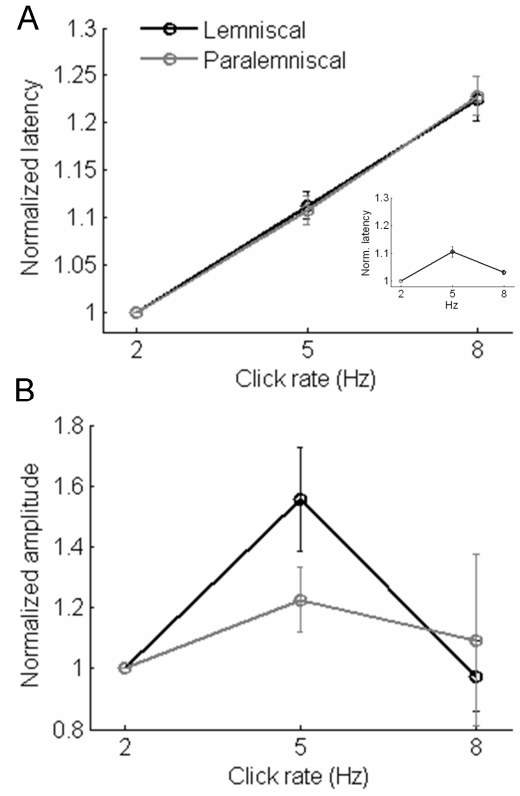


Fig. 21. Mean cortical LFPs. Mean (± SE) local field potential (a) latency and (b) amplitude for lemniscal and paralemniscal cortical responses. Figure 21a inset: mean "steady-state" cortical LFP latencies.

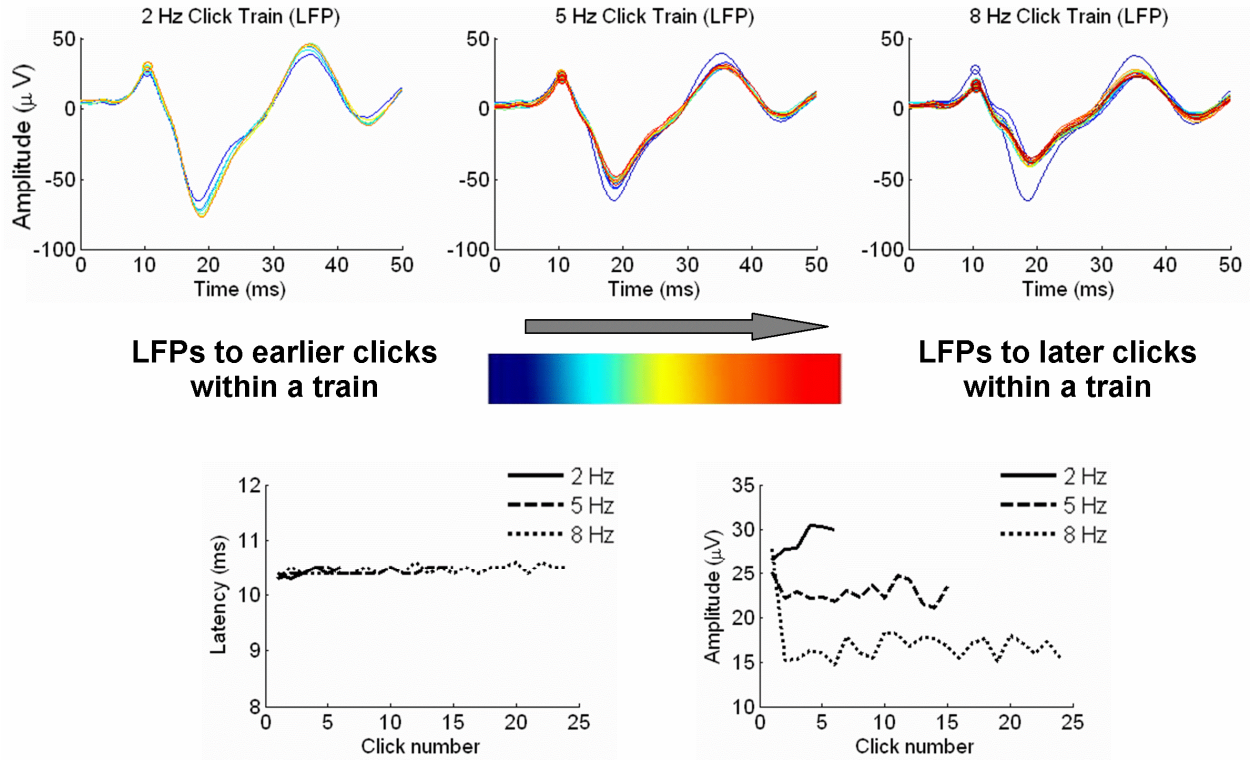


Fig. 22. Lemniscal cortical dynamics, pattern #1. Top: overlays of LFPs from a lemniscal recording site in response to each click in a click train for the 2 Hz condition (left), 5 Hz condition (center) and 8 Hz condition (right). LFPs in response to earlier clicks in the train are plotted in blue while responses to the last clicks in the train are plotted in red. Bottom left: peak LFP latency as a function of click number for the three rate conditions. Bottom right: peak LFP amplitude as a function of click number for the three rate conditions.

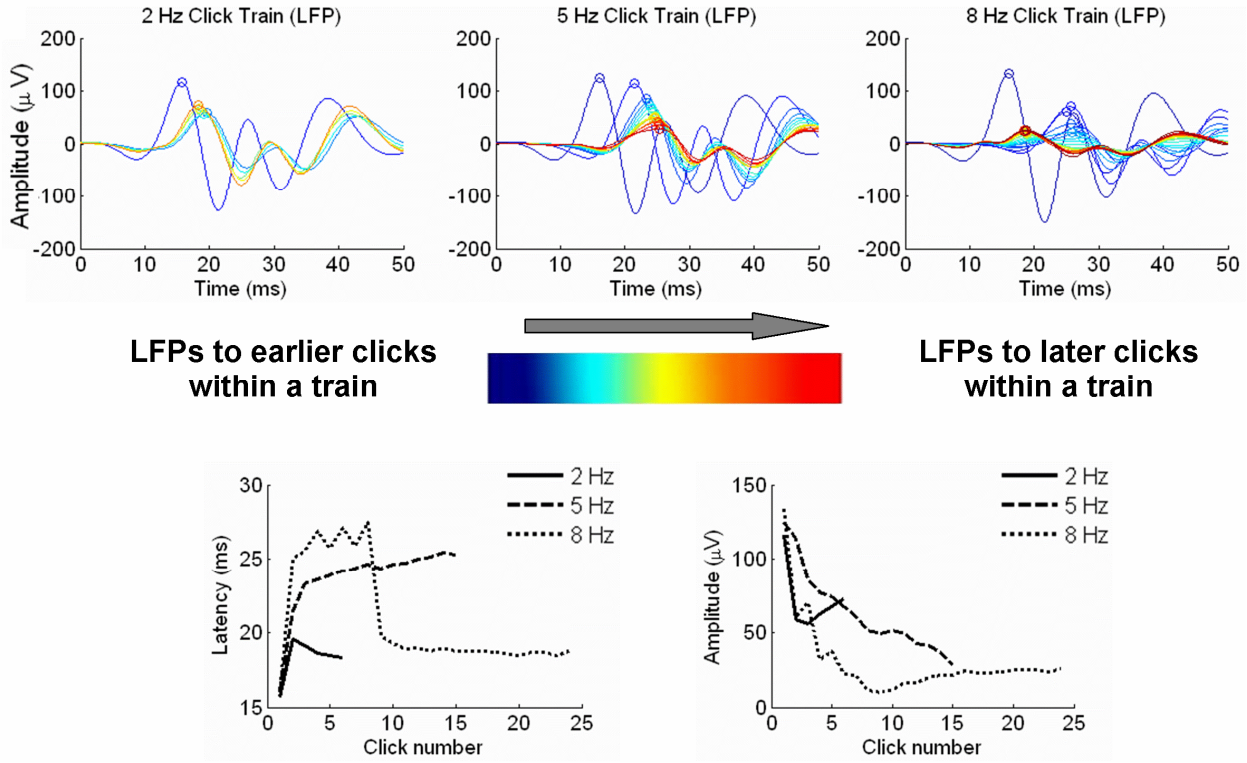


Fig. 23. Lemniscal cortical dynamics, pattern #2. Top: overlays of LFPs from a lemniscal recording site in response to each click in a click train for the 2 Hz condition (left), 5 Hz condition (center) and 8 Hz condition (right). LFPs in response to earlier clicks in the train are plotted in blue while responses to the last clicks in the train are plotted in red. Bottom left: peak LFP latency as a function of click number for the three rate conditions. Bottom right: peak LFP amplitude as a function of click number for the three rate conditions.

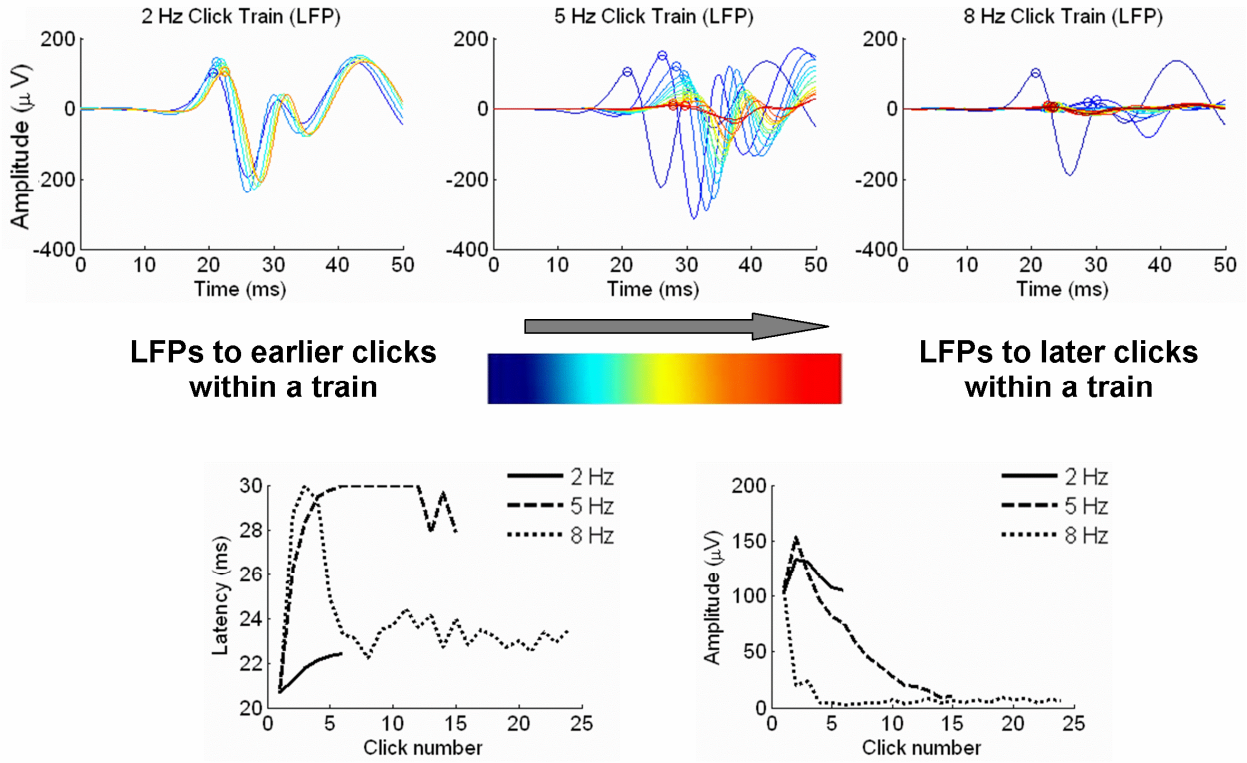


Fig. 24. Paralemniscal cortical dynamics. Top: overlays of LFPs from a paralemniscal recording site in response to each click in a click train for the 2 Hz condition (left), 5 Hz condition (center) and 8 Hz condition (right). LFPs in response to earlier clicks in the train are plotted in blue while responses to the last clicks in the train are plotted in red. Bottom left: peak LFP latency as a function of click number for the three rate conditions. Bottom right: peak LFP amplitude as a function of click number for the three rate conditions.

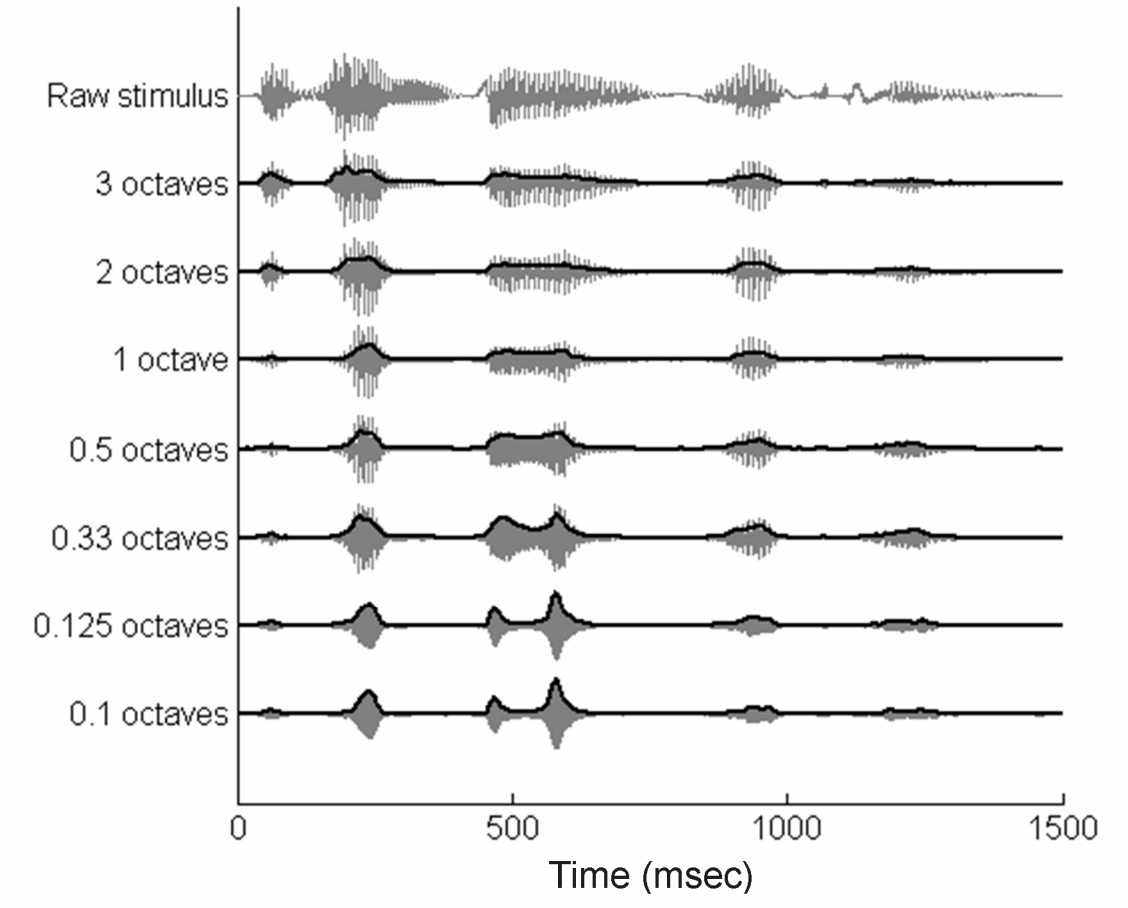


Fig. 25. Filtered stimuli and envelope extraction for the "clear" stimulus condition.

The center frequency for envelope extraction for this figure was 1000 Hz, but the best-frequency of cortical MUAs was used for the cross-correlation analysis. Since narrower pass-bands (i.e., plots at the bottom of the figure) are inherently smaller in amplitude, all plots were amplitude normalized for this figure to allow a more detailed view of each of the filtered stimuli.

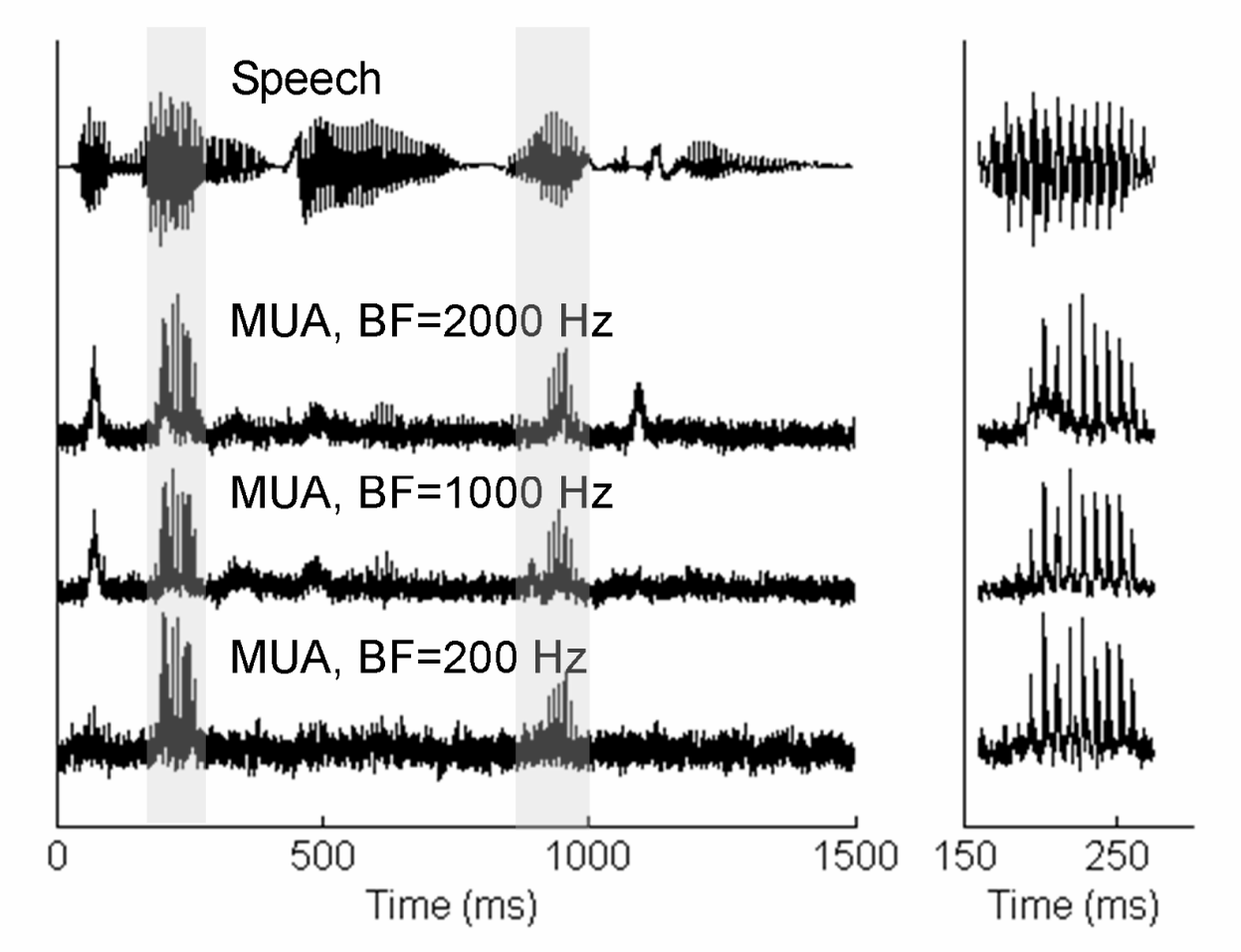


Fig. 26. Clear stimulus waveform and primary MUAs.

The highlighted region of the waveform in the main panel is enlarged in the right panel.

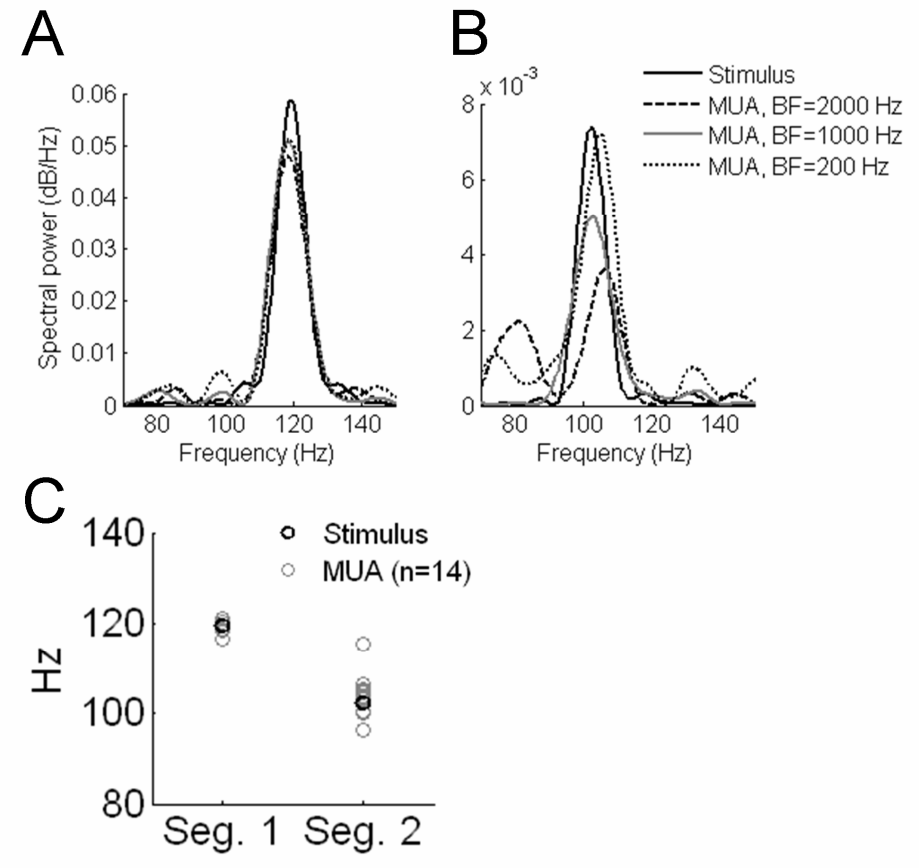


Fig. 27. Clear stimulus and primary MUS FFT for the latency ranges 160-260 (Segment 1, left) msec and 900-1000 msec (Segment 2, right). Figure 27c: Peak F0 frequency for stimulus and all primary MUA measured in Segments 1 and 2.

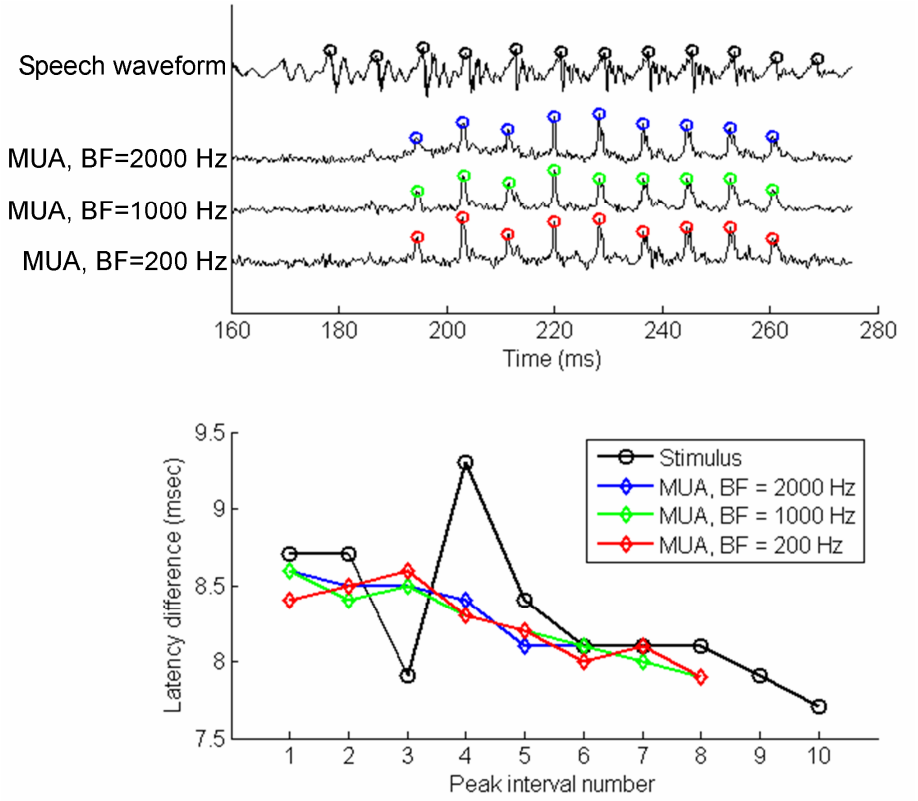


Fig. 28. Time-domain phase-locking analysis of the clear speech condition in three representative MUAs.

Fig. 28, top: peaks used in latency analysis in both the speech stimulus (top) and MUAs are identified with circles. Fig. 28, bottom: latency differences between peaks in the speech waveform and MUAs.

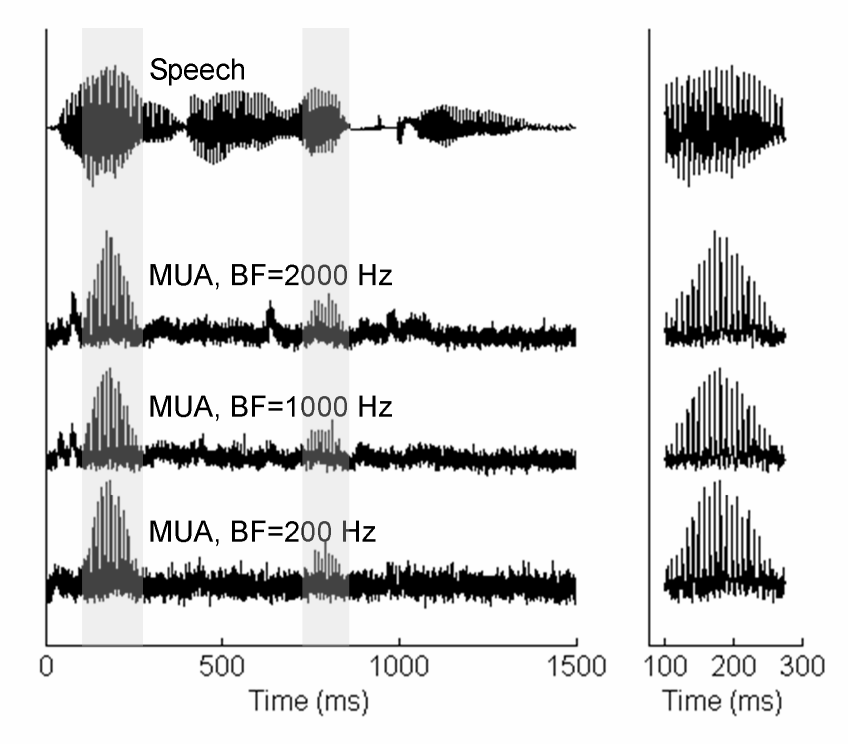


Fig. 29. Conversational stimulus waveform and primary MUAs.

The highlighted region of the waveform in the main panel is enlarged in the right panel.

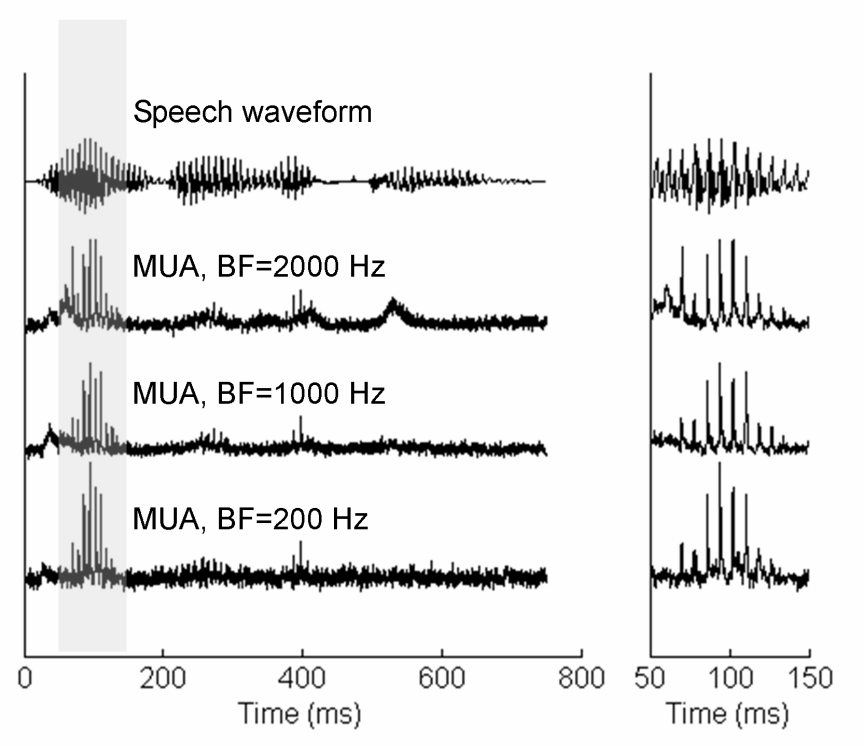


Fig. 30. Compressed stimulus waveform and primary MUAs.

The highlighted region of the waveform in the main panel is enlarged in the right panel.

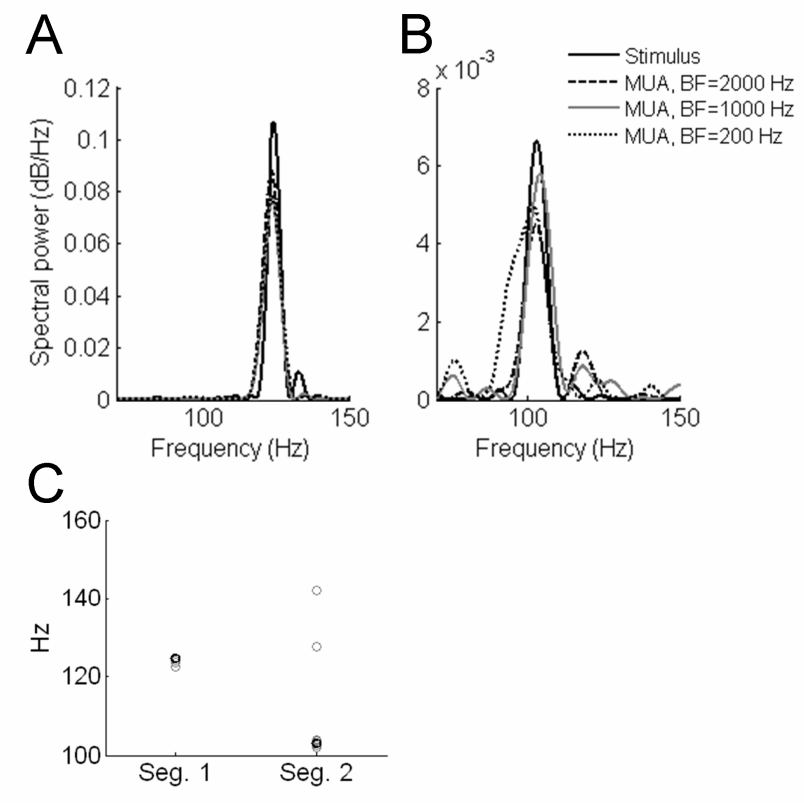


Fig. 31. Conversational stimulus and primary MUA FFT.

for the latency ranges 100-275 msec (a, left) and 725-850 msec (b, right). Figure 31c: Peak F0 frequency for stimulus and all primary MUA measured in Segments 1 and 2.

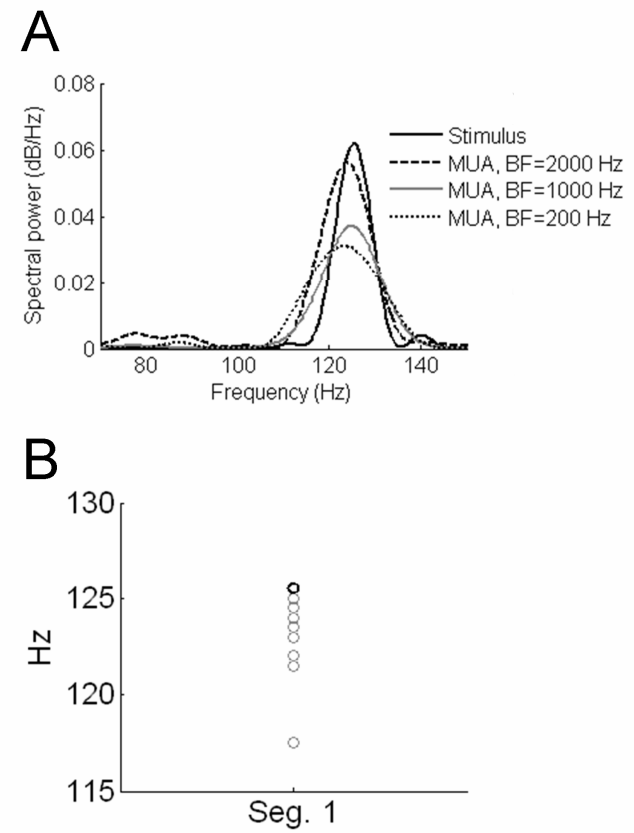


Fig. 32, top. Compressed stimulus and primary MUA FFT

for the latency ranges 50-150 msec. Fig. 32, bottom: Peak F0 frequency for stimulus and all primary MUAa.

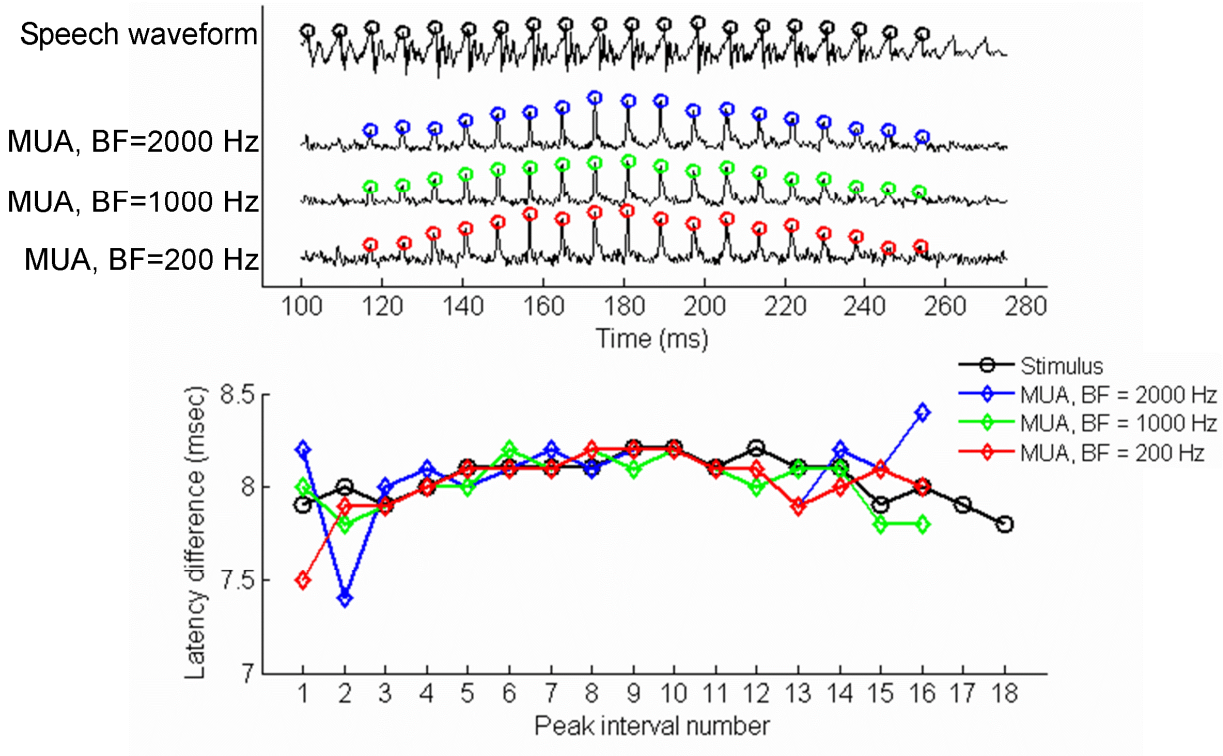


Fig. 33. Time-domain phase-locking analysis of the conversational speech condition in three representative MUAs.

Fig. 33, top: peaks used in latency analysis in both the speech stimulus (top) and MUAs are identified with circles. Fig. 33, bottom: latency differences between peaks in the speech waveform and MUAs.

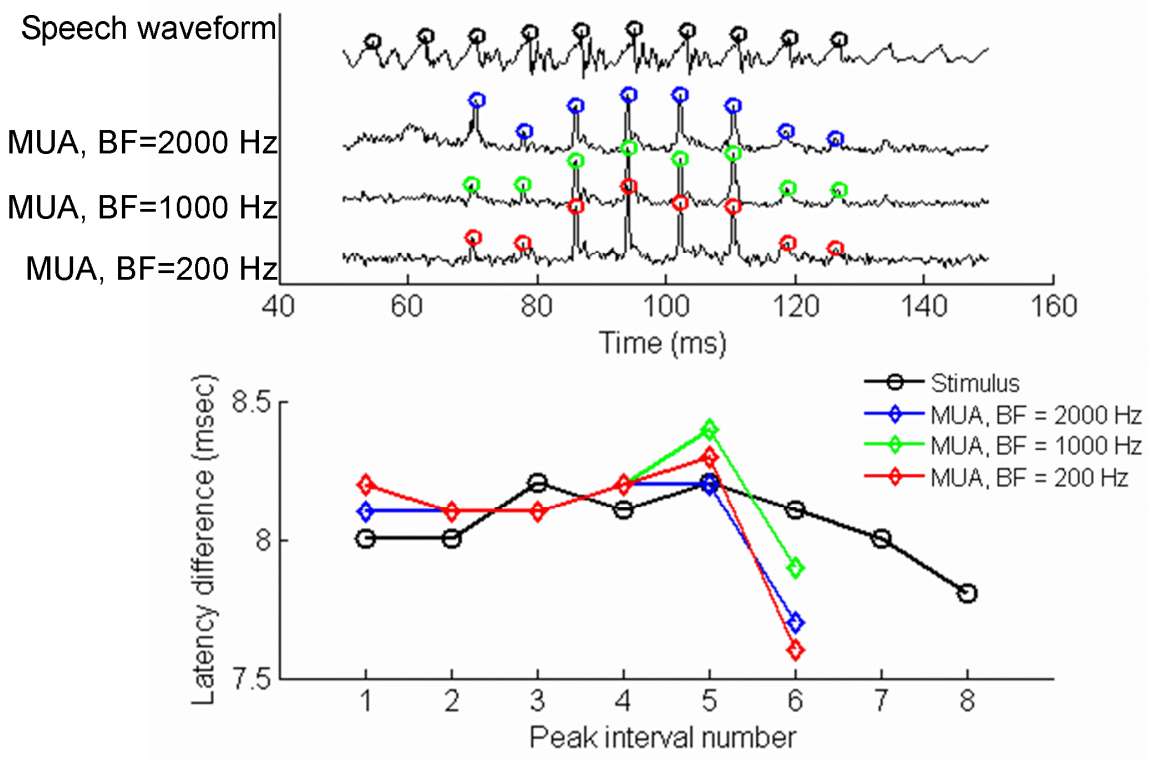


Fig. 34. Time-domain phase-locking analysis of the compressed speech condition in three representative MUAs.

Fig. 34, top: peaks used in latency analysis in both the speech stimulus (top) and MUAs are identified with circles. Fig. 34, bottom: latency differences between peaks in the speech waveform and MUAs.

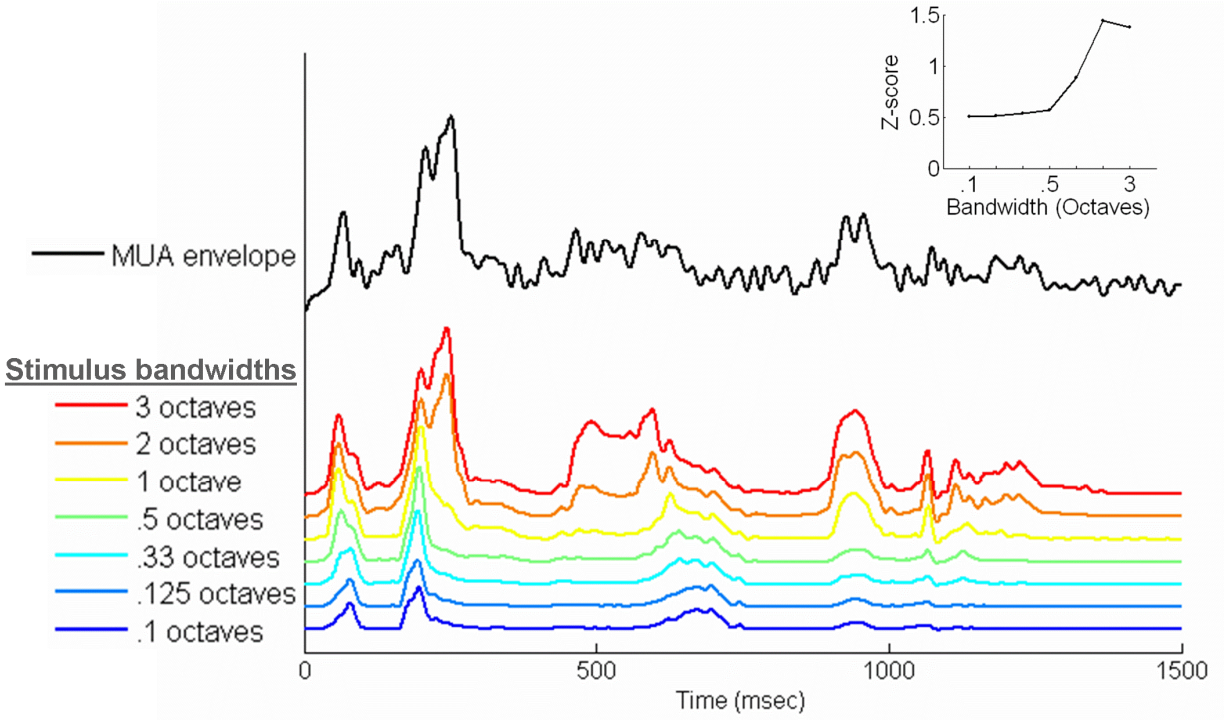


Fig. 35. Clear stimulus envelopes (bottom, in color) and raw primary MUA (top, black).

Inset: Bandwidth-correlation function for this MUA.

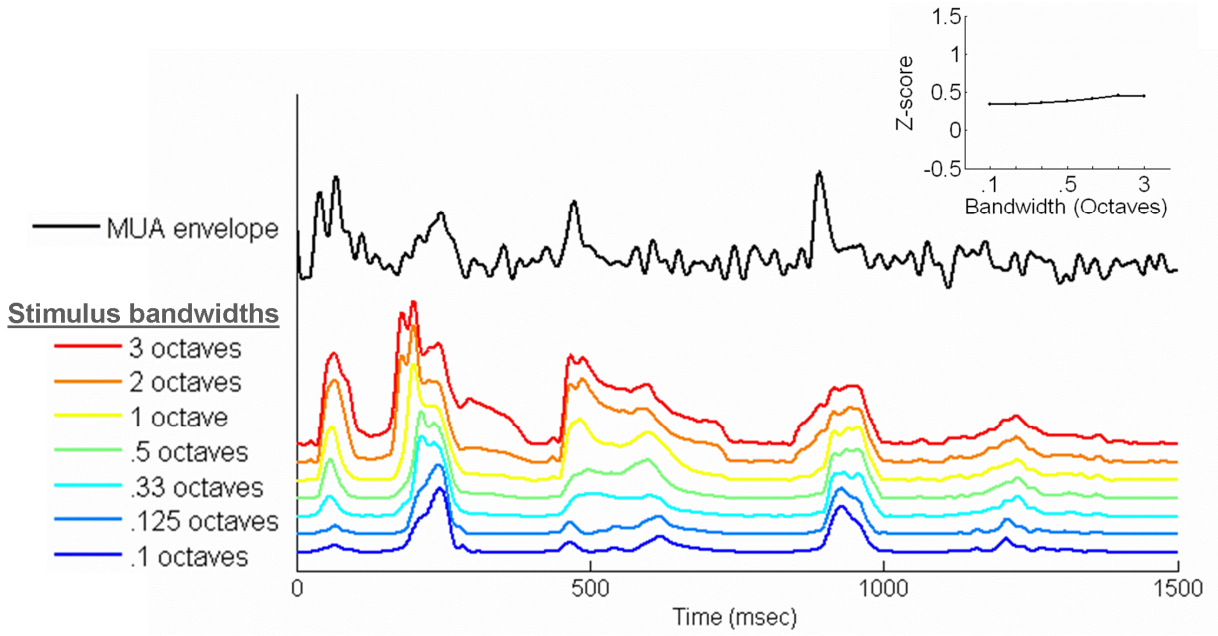


Fig. 36. Clear stimulus envelopes (bottom, in color) and raw primary MUA (top, black).

Inset: Bandwidth-correlation function for this MUA.

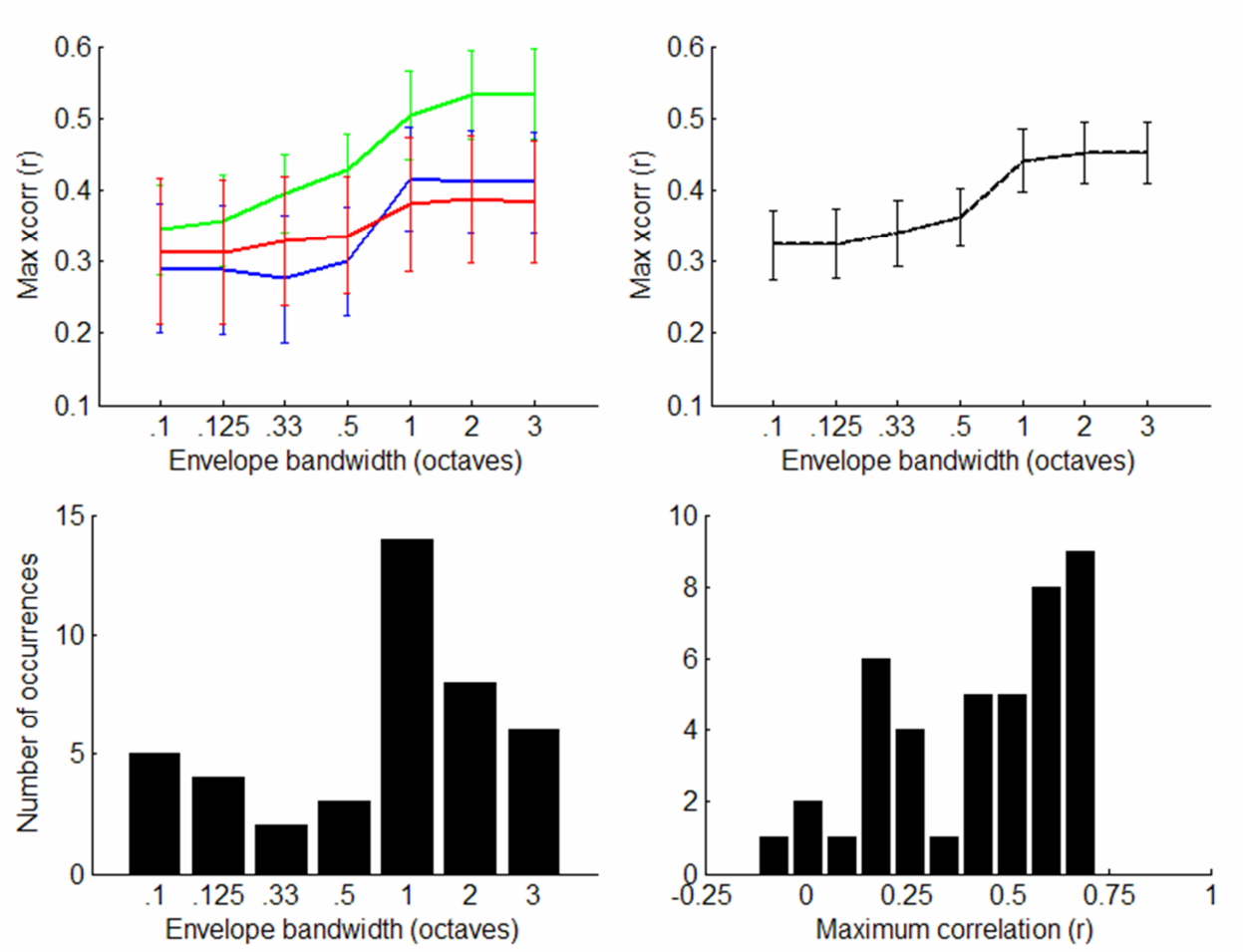


Fig. 37, top left. Mean bandwidth-correlation functions across all primary MUA envelopes in response to clear (green), conversational (blue) and compressed (red) stimuli.

Fig 37, top right: Mean bandwidth-correlation functions across all primary MUA envelopes in response to all stimuli. Fig. 37, bottom left: histogram showing envelope bandwidth preference for all primary MUA. Fig. 37, bottom right, histogram showing maximum envelope correlation coefiicient for all primary MUA and speech envelope correlations.

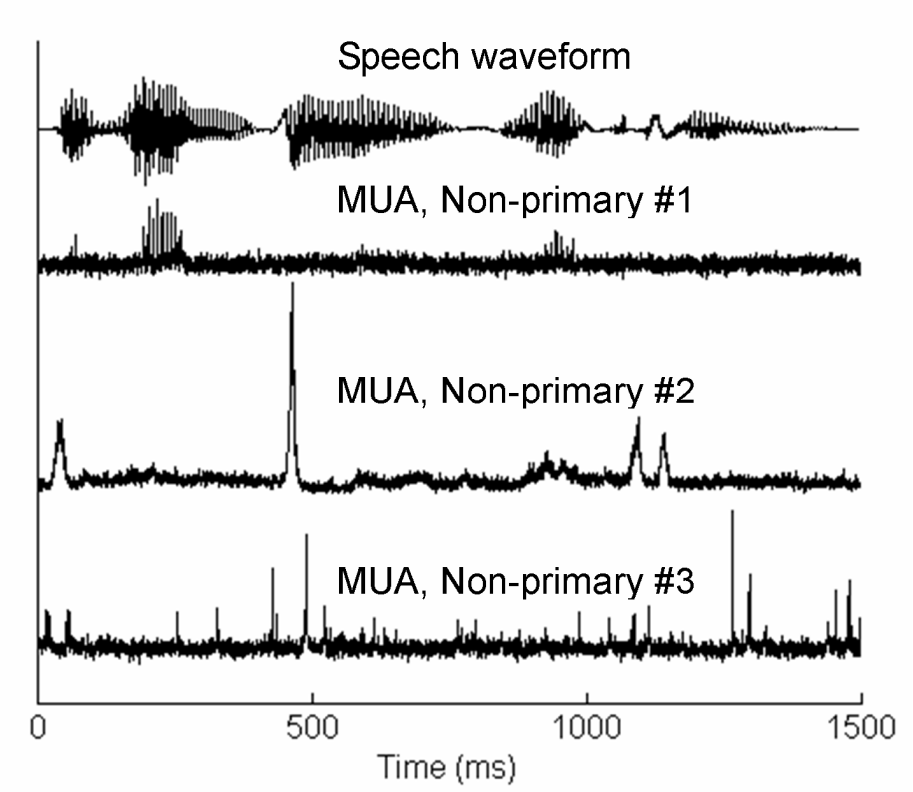


Fig. 38. Clear stimulus wavform and three representative non-primary MUAs.

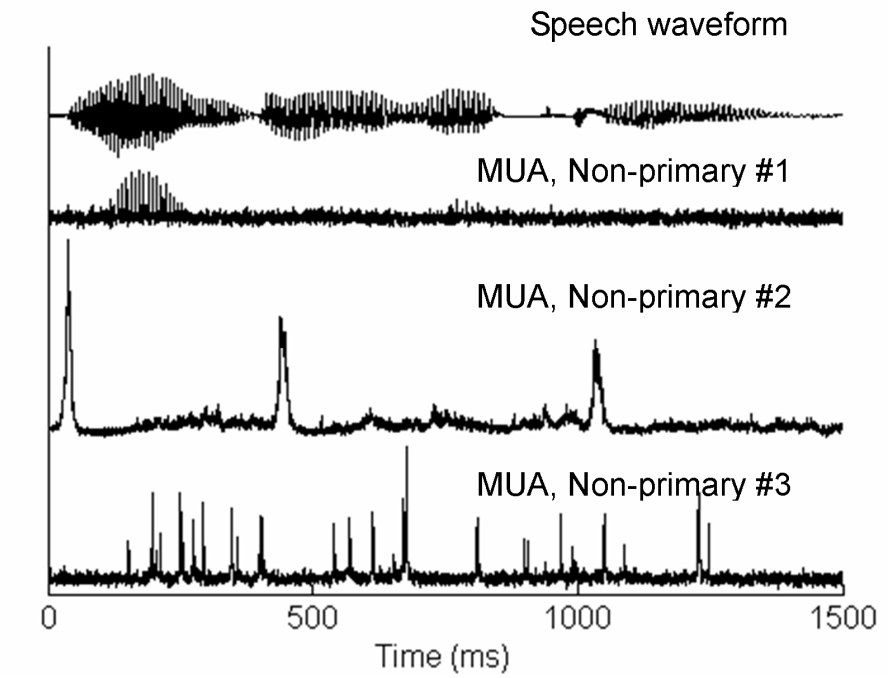


Fig. 39. Conversational stimulus wavform and three representative non-primary MUAs.

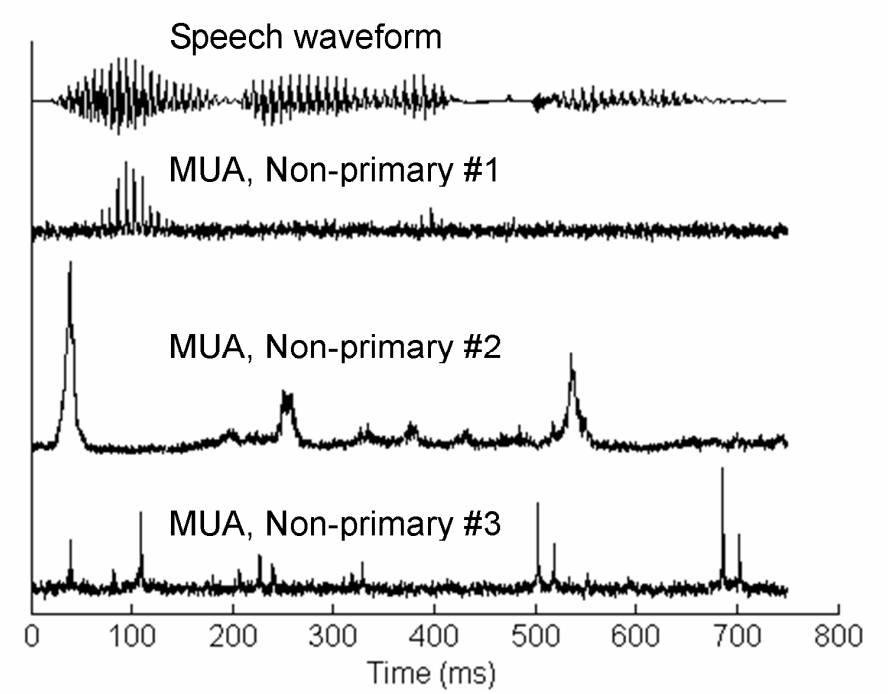


Fig. 40. Compressed stimulus wavform and three representative non-primary MUAs.

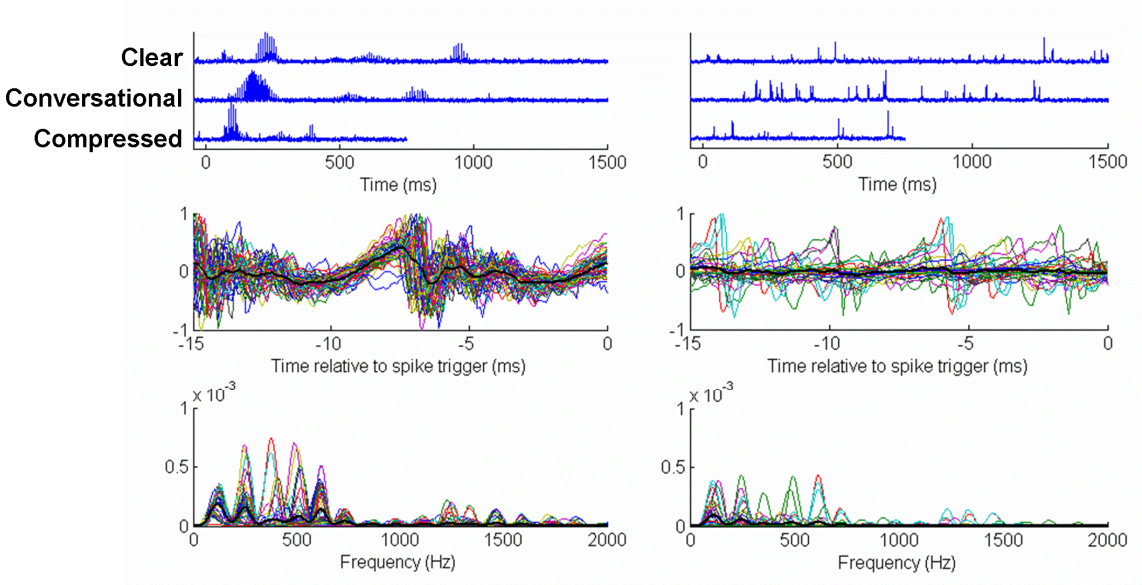


Fig. 41. Non-primary MUAs in response to clear, conversational and compressed stimuli (top), spike-triggered analysis (middle) and FFTs from spike-triggered analysis (bottom).

Fig. 41, left. Spike-triggered analysis for an ensemble that showed primary-like phase-locking to the stimulus f0 (non-primary population #1). Fig. 41, right. Spike-triggered analysis for an ensemble that showed activation that did not appear to be time-lokced to stimulus features. (non-primary population #3).

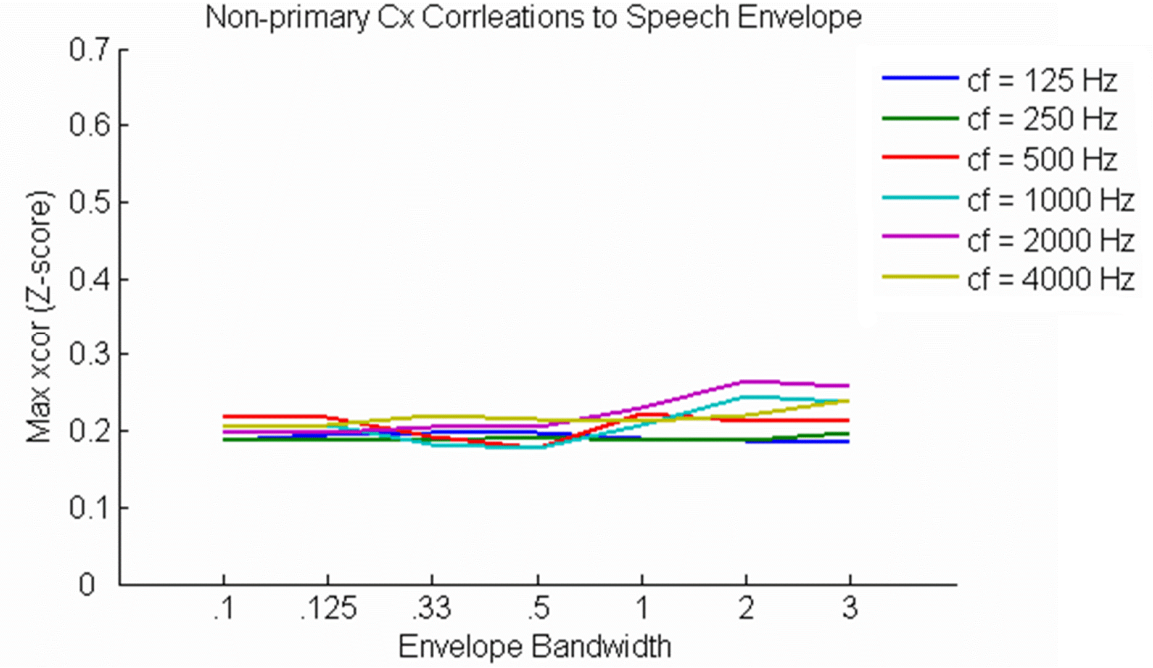


Fig. 42. Non-primary cortex MUA-envelope correlations as a function of speech envelope bandwidth.

REFERENCES

Abrams DA, Nicol T, Zecker SG, Kraus N (2006) Auditory brainstem timing predicts cerebral asymmetry for speech. J Neurosci 26:11131-11137.

Abrams DA, Nicol T, Zecker S, Kraus N (in press) Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. J Neurosci.

Abrams DA, Nicol T, Zecker S, Kraus N (submitted-a) Abnormal cortical processing of syllable rates in speech in reading-impaired children. Proc Natn Acad Sci USA.

Abrams DA, Nicol T, Zecker S, Kraus N (submitted-b) A role for the paralemniscal auditory pathway in the coding of slow temporal information. Exp Brain Res.

Ahissar E, Zacksenhouse M (2001) Temporal and spatial coding in the rat vibrissal system. Prog Brain Res 130:75-87.

Ahissar E, Ahissar M (2005) Processing of the Temporal Envelope of Speech. In: The Auditory Cortex: A Synthesis of Human and Animal Research (Konig R, Heil P, Budinger E, Scheich H, eds), pp 295-313. London: Lawrence Erlbaum Associates.

Ahissar E, Haidarliu S, Zacksenhouse M (1997) Decoding temporally encoded sensory input by cortical oscillations and thalamic phase comparators. Proc Natl Acad Sci U S A 94:11633-11638.

Ahissar E, Sosnik R, Haidarliu S (2000) Transformation from temporal to rate coding in a somatosensory thalamocortical pathway. Nature 406:302-306.

Ahissar E, Sosnik R, Bagdasarian K, Haidarliu S (2001a) Temporal frequency of whisker movement. II. Laminar organization of cortical representations. J Neurophysiol 86:354-367.

Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM (2001b) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc Natl Acad Sci U S A 98:13367-13372.

Ahissar M, Hochstein S (2004) The reverse hierarchy theory of visual perceptual learning. Trends Cogn Sci 8:457-464.

Anderson SE, Kilgard MP, Sloan AM, Rennaker RL (2006) Response to broadband repetitive stimuli in auditory cortex of the unanesthetized rat. Hear Res 213:107-117.

Banai K, Nicol T, Zecker SG, Kraus N (2005a) Brainstem timing: implications for cortical processing and literacy. J Neurosci 25:9850-9857.

Banai K, Russo NM, Nicol T, Zecker S, Kraus N (2005b) What can brainstem timing teach us about learning disabilities? In: The International Conference on the Neurophysiology of Developmental Disorders in Children. Freiburg.

Bartlett EL, Wang X (2007) Neural representations of temporally modulated signals in the auditory thalamus of awake primates. J Neurophysiol 97:1005-1017.

Beasley DS, Bratt GW, Rintelmann WF (1980) Intelligibility of time-compressed sentential stimuli. J Speech Hear Res 23:722-731.

Beitel RE, Schreiner CE, Cheung SW, Wang X, Merzenich MM (2003) Reward-dependent plasticity in the primary auditory cortex of adult monkeys trained to discriminate temporally modulated signals. Proc Natl Acad Sci U S A 100:11070-11075.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. Nature 403:309-312.

Belin P, Zilbovicius M, Crozier S, Thivard L, Fontaine A, Masure MC, Samson Y (1998) Lateralization of speech and auditory temporal processing. J Cogn Neurosci 10:536-540.

Bellis TJ, Nicol T, Kraus N (2000) Aging affects hemispheric asymmetry in the neural representation of speech sounds. J Neurosci 20:791-797.

Bieser A, Muller-Preuss P (1996) Auditory responsive cortex in the squirrel monkey: neural responses to amplitude-modulated sounds. Exp Brain Res 108:273-284.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex 10:512-528.

Bishop DV (2006) Dyslexia: what's the problem? Dev Sci 9:256-257; discussion 265-259.

Boemio A, Fromm S, Braun A, Poeppel D (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat Neurosci 8:389-395.

Bradlow AR, Kraus N, Hayes E (2003) Speaking clearly for children with learning disabilities: sentence perception in noise. J Speech Lang Hear Res 46:80-97.

Bureau I, von Saint Paul F, Svoboda K (2006) Interdigitated paralemniscal and lemniscal pathways in the mouse barrel cortex. PLoS Biol 4:e382.

Cariani PA, Delgutte B (1996) Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. J Neurophysiol 76:1698-1716.

Carrell TD, Bradlow AR, Nicol TG, Koch DB, Kraus N (1999) Interactive software for evaluating auditory discrimination. Ear Hear 20:175-176.

Castro-Alamancos MA (2002) Different temporal processing of sensory inputs in the rat thalamus during quiescent and information processing states in vivo. J Physiol 539:567-578.

Chait M, Eden G, Poeppel D, Simon JZ, Hill DF, Flowers DL (2007) Delayed detection of tonal targets in background noise in dyslexia. Brain Lang 102:80-90.

Creutzfeldt O, Hellweg FC, Schreiner C (1980) Thalamocortical transformation of responses to complex auditory stimuli. Exp Brain Res 39:87-104.

Cunningham J, Nicol T, Zecker S, Kraus N (2000) Speech-evoked neurophysiologic responses in children with learning problems: development and behavioral correlates of perception. Ear Hear 21:554-568.

Cunningham J, Nicol T, Zecker SG, Bradlow A, Kraus N (2001a) Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. Clin Neurophysiol 112:758-767.

Cunningham J, Nicol T, Zecker S, Bradlow A, Kraus N (2001b) Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. Clinical Neurophysiology 112:758-767.

Cunningham J, Nicol T, King C, Zecker SG, Kraus N (2002) Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex. Hear Res 169:97-111.

de la Mothe LA, Blumell S, Kajikawa Y, Hackett TA (2006a) Cortical connections of the auditory cortex in marmoset monkeys: core and medial belt regions. J Comp Neurol 496:27-71.

de la Mothe LA, Blumell S, Kajikawa Y, Hackett TA (2006b) Thalamic connections of the auditory cortex in marmoset monkeys: core and medial belt regions. J Comp Neurol 496:72-96.

Delgutte B, Cariani PA (1998) Neural coding of the temporal envelope of speech: relation to modulation transfer functions. In: Psychophysical and physiological advances in hearing (Palmer AR, Summerfield Q, Meddis R, eds), pp 595-603. London: Whurr Publishers Ltd.

Diamond ME, Armstrong-James M (1992) Role of parallel sensory pathways and cortical columns in learning. Concepts in Neuroscience 3:55-78.

Drullman R, Festen JM, Plomp R (1994a) Effect of temporal envelope smearing on speech reception. J Acoust Soc Am 95:1053-1064.

Drullman R, Festen JM, Plomp R (1994b) Effect of reducing slow temporal modulations on speech reception. J Acoust Soc Am 95:2670-2680.

Eggermont JJ (1991) Rate and synchronization measures of periodicity coding in cat primary auditory cortex. Hear Res 56:153-167.

Eggermont JJ (1992) Stimulus induced and spontaneous rhythmic firing of single units in cat primary auditory cortex. Hear Res 61:1-11.

Eggermont JJ (1998) Representation of spectral and temporal sound features in three cortical fields of the cat. Similarities outweigh differences. J Neurophysiol 80:2743-2764.

Eggermont JJ, Smith GM (1995) Synchrony between single-unit activity and local field potentials in relation to periodicity coding in primary auditory cortex. J Neurophysiol 73:227-245.

Everest FA (2001) The master handbook of acoustics 4th Edition. New York: McGraw-Hill.

Fishman YI, Reser DH, Arezzo JC, Steinschneider M (2000) Complex tone processing in primary auditory cortex of the awake monkey. I. Neural ensemble correlates of roughness. J Acoust Soc Am 108:235-246.

Gaese BH, Ostwald J (1995) Temporal coding of amplitude and frequency modulation in the rat auditory cortex. Eur J Neurosci 7:438-450.

Galaburda AM, Sherman GF, Rosen GD, Aboitiz F, Geschwind N (1985) Developmental dyslexia: four consecutive patients with cortical anomalies. Ann Neurol 18:222-233.

Gehr DD, Komiya H, Eggermont JJ (2000) Neuronal responses in cat primary auditory cortex to natural and altered species-specific calls. Hear Res 150:27-42.

Geschwind N, Galaburda AM (1985a) Cerebral lateralization. Biological mechanisms, associations, and pathology: II. A hypothesis and a program for research. Arch Neurol 42:521-552.

Geschwind N, Galaburda AM (1985b) Cerebral lateralization. Biological mechanisms, associations, and pathology: III. A hypothesis and a program for research. Arch Neurol 42:634-654.

Gleason HA (1961) An introduction to descriptive linguistics, revised edition. New York: Holt, Rinehart and Winston.

Goldstein JMH, Kiang NYS, Brown RM (1959) Responses of the Auditory Cortex to Repetitive Acoustic Stimuli. The Journal of the Acoustical Society of America 31:356-364.

Golomb D, Ahissar E, Kleinfeld D (2006) Coding of stimulus frequency by latency in thalamic networks through the interplay of GABAB-mediated feedback and stimulus shape. J Neurophysiol 95:1735-1750.

Goswami U (2002) Phonology, reading development and dyslexia: A cross-linguistic perspective. Ann Dyslexia 52:1-23.

Goswami U (2006) Sensorimotor impairments in dyslexia: getting the beat. Dev Sci 9:257-259; discussion 265-259.

Goswami U, Thomson J, Richardson U, Stainthorp R, Hughes D, Rosen S, Scott SK (2002) Amplitude envelope onsets and developmental dyslexia: A new hypothesis. Proc Natl Acad Sci U S A 99:10911-10916.

Gourevitch B, Eggermont JJ (2007) Spatial representation of neural responses to natural and altered conspecific vocalizations in cat auditory cortex. J Neurophysiol 97:144-158.

He J (2001) On and off pathways segregated at the auditory thalamus of the guinea pig. J Neurosci 21:8672-8679.

He J (2002) OFF responses in the auditory thalamus of the guinea pig. J Neurophysiol 88:2377-2386.

He J (2003a) Corticofugal modulation on both ON and OFF responses in the nonlemniscal auditory thalamus of the guinea pig. J Neurophysiol 89:367-381.

He J (2003b) Slow oscillation in non-lemniscal auditory thalamus. J Neurosci 23:8281-8290.

Hebb DO (1949) The organization of behavior: A neuropsychological theory. New York: Wiley.

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393-402.

Horst JW, Javel E, Farley GR (1986) Coding of spectral fine structure in the auditory nerve. I. Fourier analysis of period and interspike interval histograms. J Acoust Soc Am 79:398-416.

Inagaki M, Tomita Y, Takashima S, Ohtani K, Andoh G, Takeshita K (1987) Functional and morphometrical maturation of the brainstem auditory pathway. Brain Dev 9:597-601.

Jasper HH (1958) The ten-twenty electrode system of the international federation. Electroencephalogr Clin Neurophysiol 10:371-375.

Joanisse MF, Gati JS (2003) Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. Neuroimage 19:64-79.

Johnson KL, Nicol TG, Kraus N (2005) Brain Stem Response to Speech: A Biological Marker of Auditory Processing. Ear Hear 26:424-434.

Johnsrude IS, Zatorre RJ, Milner BA, Evans AC (1997) Left-hemisphere specialization for the processing of acoustic transients. Neuroreport 8:1761-1765.

Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. Physiol Rev 84:541-577.

Kaas JH, Hackett TA (2000) Subdivisions of auditory cortex and processing streams in primates. Proc Natl Acad Sci U S A 97:11793-11799.

Kaur S, Lazar R, Metherate R (2004) Intracortical pathways determine breadth of subthreshold frequency receptive fields in primary auditory cortex. J Neurophysiol 91:2551-2567.

Kilgard MP, Merzenich MM (1998a) Plasticity of temporal information processing in the primary auditory cortex. Nat Neurosci 1:727-731.

Kilgard MP, Merzenich MM (1998b) Cortical map reorganization enabled by nucleus basalis activity. Science 279:1714-1718.

King C, Nicol T, McGee T, Kraus N (1999) Thalamic asymmetry is related to acoustic signal complexity. Neurosci Lett 267:89-92.

King C, Warrier CM, Hayes E, Kraus N (2002) Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. Neurosci Lett 319:111-115.

Koch DB, McGee TJ, Bradlow AR, Kraus N (1999) Acoustic-phonetic approach toward understanding neural processes and speech perception. J Am Acad Audiol 10:304-318.

Kraus N, Nicol T (2005) Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. Trends Neurosci 28:176-181.

Kraus N, McGee TJ, Carrell TD, Zecker SG, Nicol TG, Koch DB (1996) Auditory neurophysiologic responses and discrimination deficits in children with learning problems. Science 273:971-973.

Krause JC, Braida LD (2004) Acoustic properties of naturally produced clear speech at normal speaking rates. J Acoust Soc Am 115:362-378.

Krishnan A, Xu Y, Gandour J, Cariani P (2005) Encoding of pitch in the human brainstem is sensitive to language experience. Brain Res Cogn Brain Res 25:161-168.

Ladefoged P (2001) A course in phonetics, 4th Edition. Fort Worth: Harcourt College Publishers.

Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J Neurophysiol 94:1904-1911.

Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. J Neurophysiol 60:1799-1822.

Lehmann D, Skrandies W (1980) Reference-free identification of components of checkerboard-evoked multichannel potential fields. Electroencephalogr Clin Neurophysiol 48:609-621.

Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA (2005) Neural substrates of phonemic perception. Cereb Cortex 15:1621-1631.

Liegeois-Chauvel C, de Graaf JB, Laguitton V, Chauvel P (1999) Specialization of left auditory cortex for speech perception in man depends on temporal coding. Cereb Cortex 9:484-496.

Liégeois-Chauvel C, Lorenzi C, Trébuchon A, Régis J, Chauvel P (2004) Temporal Envelope Processing in the Human Left and Right Auditory Cortices. Cerebral Cortex 14:731-740.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. Nature 412:150-157.

Lorenzi C, Dumont A, Fullgrabe C (2000) Use of temporal envelope cues by children with developmental dyslexia. J Speech Lang Hear Res 43:1367-1379.

Lu T, Wang X (2000) Temporal discharge patterns evoked by rapid sequences of wide- and narrowband clicks in the primary auditory cortex of cat. J Neurophysiol 84:236-246.

Lu T, Liang L, Wang X (2001) Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nat Neurosci 4:1131-1138.

Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54:1001-1010.

Ma X, Suga N (2001) Plasticity of bat's central auditory system evoked by focal electric stimulation of auditory and/or somatosensory cortices. J Neurophysiol 85:1078-1087.

McGee T, Kraus N, King C, Nicol T, Carrell TD (1996) Acoustic elements of speechlike stimuli are reflected in surface recorded responses over the guinea pig temporal lobe. J Acoust Soc Am 99:3606-3614.

Merzenich MM, Knight PL, Roth GL (1975) Representation of cochlea within primary auditory cortex in the cat. J Neurophysiol 38:231-249.

Merzenich MM, Jenkins WM, Johnston P, Schreiner C, Miller SL, Tallal P (1996) Temporal processing deficits of language-learning impaired children ameliorated by training. Science 271:77-81.

Meyer M, Zaehle T, Gountouna VE, Barron A, Jancke L, Turk A (2005) Spectro-temporal processing during speech perception involves left posterior auditory cortex. Neuroreport 16:1985-1989.

Miller LM, Escabi MA, Read HL, Schreiner CE (2002) Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. J Neurophysiol 87:516-527.

Moore DR, Rosenberg JF, Coleman JS (2005) Discrimination training of phonemic contrasts enhances phonological processing in mainstream school children. Brain Lang 94:72-85.

Morgan WP (1896) A case of congenital word-blindness. British Medical Journal 2:1378.

Naatanen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, Allik J, Sinkkonen J, Alho K (1997) Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature 385:432-434.

Nagarajan SS, Cheung SW, Bedenbaugh P, Beitel RE, Schreiner CE, Merzenich MM (2002) Representation of spectral and temporal envelope of twitter vocalizations in common marmoset primary auditory cortex. J Neurophysiol 87:1723-1737.

Narain C, Scott SK, Wise RJ, Rosen S, Leff A, Iversen SD, Matthews PM (2003) Defining a left-lateralized response specific to intelligible speech using fMRI. Cereb Cortex 13:1362-1368.

Nicholson RI, Fawcett AJ (1994) Reaction times and dyslexia. Q J Exp Physiol 47A:29–48.

Obleser J, Zimmermann J, Van Meter J, Rauschecker JP (2007) Multiple stages of auditory speech perception reflected in event-related FMRI. Cereb Cortex 17:2251-2257.

Perrot X, Ryvlin P, Isnard J, Guenot M, Catenoix H, Fischer C, Mauguiere F, Collet L (2005) Evidence for Corticofugal Modulation of Peripheral Auditory Activity in Humans. Cereb Cortex:bhj035.

Phillips DP, Farmer ME (1990) Acquired word deafness, and the temporal grain of sound representation in the primary auditory cortex. Behav Brain Res 40:85-94.

Phillips DP, Hall SE, Hollett JL (1989) Repetition rate and signal level effects on neuronal responses to brief tone pulses in cat auditory cortex. J Acoust Soc Am 85:2537-2549.

Poeppel D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Commun 41:245-255.

Ponton C, Eggermont JJ, Khosla D, Kwong B, Don M (2002) Maturation of human central auditory system activity: separating auditory evoked potentials by dipole source modeling. Clin Neurophysiol 113:407-420.

Ponton CW, Eggermont JJ, Kwong B, Don M (2000) Maturation of human central auditory system activity: evidence from multi-channel evoked potentials. Clin Neurophysiol 111:220-236.

Preuss A, Muller-Preuss P (1990) Processing of amplitude modulated sounds in the medial geniculate body of squirrel monkeys. Exp Brain Res 79:207-211.

Ramus F (2003) Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction? Curr Opin Neurobiol 13:212-218.

Ramus F, Rosen S, Dakin SC, Day BL, Castellote JM, White S, Frith U (2003) Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. Brain 126:841-865.

Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. Science 268:111-114.

Redies H, Brandner S (1991) Functional organization of the auditory thalamus in the guinea pig. Exp Brain Res 86:384-392.

Redies H, Sieben U, Creutzfeldt OD (1989a) Functional subdivisions in the auditory cortex of the guinea pig. J Comp Neurol 282:473-488.

Redies H, Brandner S, Creutzfeldt OD (1989b) Anatomy of the auditory thalamocortical system of the guinea pig. J Comp Neurol 282:489-511.

Rodd JM, Davis MH, Johnsrude IS (2005) The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. Cereb Cortex 15:1261-1269.

Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci 336:367-373.

Rosen S (2003) Auditory processing in dyslexia and specific language impairment: is there a deficit? What is its nature? Does it explain anything? J Phonetics 31:509-527.

Rouiller E, de Ribaupierre F (1982) Neurons sensitive to narrow ranges of repetitive acoustic transients in the medial geniculate body of the cat. Exp Brain Res 48:323-326.

Rouiller E, de Ribaupierre Y, Toros-Morel A, de Ribaupierre F (1981) Neural coding of repetitive clicks in the medial geniculate body of cat. Hear Res 5:81-100.

Russo N, Nicol T, Musacchia G, Kraus N (2004) Brainstem responses to speech syllables. Clinical Neurophysiology 115:2021-2030.

Russo NM, Nicol TG, Zecker SG, Hayes EA, Kraus N (2005) Auditory training improves neural timing in the human brainstem. Behav Brain Res 156:95-103.

Sachs MB (1984) Neural coding of complex sounds: speech. Annu Rev Physiol 46:261-273.

Sachs MB, Young ED (1979) Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate. J Acoust Soc Am 66:470-479.

Sachs MB, Voigt HF, Young ED (1983) Auditory nerve representation of vowels in background noise. J Neurophysiol 50:27-45.

Sakaguchi S, Arai T, Murahara Y (2000) The effect of polarity inversion of speech on human perception and data hiding as an application. International Conference on Acoustics, Speech, and Signal Processing 2:917-920.

Sakai M, Suga N (2001) Plasticity of the cochleotopic (frequency) map in specialized and nonspecialized auditory cortices. Proc Natl Acad Sci U S A 98:3507-3512.

Scherg M, Von Cramon D (1986) Evoked dipole source potentials of the human auditory cortex. Electroencephalogr Clin Neurophysiol 65:344-360.

Schonwiesner M, Rubsamen R, von Cramon DY (2005) Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. Eur J Neurosci 22:1521-1528.

Schwartz J, Tallal P (1980) Rate of acoustic change may underlie hemispheric specialization for speech perception. Science 207:1380-1381.

Scott SK, Blank CC, Rosen S, Wise RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123 Pt 12:2400-2406.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303-304.

Sharma A, Dorman M (2000) Neurophysiologic correlates of cross-language phonetic perception. J Acoust Soc Am 107:2697-2703.

Sharma A, Marsh C, Dorman M (2000) Relationship between N1 evoked potential morphology and the perception of voicing. J Acoust Soc Am 108:3030-3035.

Sosnik R, Haidarliu S, Ahissar E (2001) Temporal frequency of whisker movement. I. Representations in brain stem and thalamus. J Neurophysiol 86:339-353.

Stein J, Walsh V (1997) To see but not to read; the magnocellular theory of dyslexia. Trends Neurosci 20:147-152.

Steinschneider M, Arezzo J, Vaughan HG, Jr. (1980) Phase-locked cortical responses to a human speech sound and low-frequency tones in the monkey. Brain Res 198:75-84.

Steinschneider M, Arezzo J, Vaughan HG, Jr. (1982) Speech evoked activity in the auditory radiations and cortex of the awake monkey. Brain Res 252:353-365.

Steinschneider M, Arezzo JC, Vaughan HG, Jr. (1990) Tonotopic features of speech-evoked activity in primate auditory cortex. Brain Res 519:158-168.

Steinschneider M, Fishman YI, Arezzo JC (2003) Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey. J Acoust Soc Am 114:307-321.

Steinschneider M, Schroeder CE, Arezzo JC, Vaughan HG, Jr. (1994) Speech-evoked activity in primary auditory cortex: effects of voice onset time. Electroencephalogr Clin Neurophysiol 92:30-43.

Steinschneider M, Reser D, Schroeder CE, Arezzo JC (1995) Tonotopic organization of responses reflecting stop consonant place of articulation in primary auditory cortex (A1) of the monkey. Brain Res 674:147-152.

Steinschneider M, Reser DH, Fishman YI, Schroeder CE, Arezzo JC (1998) Click train encoding in primary auditory cortex of the awake monkey: evidence for two mechanisms subserving pitch perception. J Acoust Soc Am 104:2935-2955.

Tallal P (1980) Auditory temporal perception, phonics, and reading disabilities in children. Brain Lang 9:182-198.

Tallal P, Piercy M (1973) Defects of non-verbal auditory perception in children with developmental aphasia. Nature 241:468-469.

Tallal P, Piercy M (1974) Developmental aphasia: rate of auditory processing and selective impairment of consonant perception. Neuropsychologia 12:83-93.

Tallal P, Piercy M (1975) Developmental aphasia: the perception of brief vowels and extended stop consonants. Neuropsychologia 13:69-74.

Tallal P, Miller S, Fitch RH (1993) Neurobiological basis of speech: a case for the preeminence of temporal processing. Ann N Y Acad Sci 682:27-47.

Tallal P, Merzenich MM, Miller S, Jenkins W (1998) Language learning impairments: integrating basic science, technology, and remediation. Exp Brain Res 123:210-219.

Tonnquist-Uhlen I, Ponton CW, Eggermont JJ, Kwong B, Don M (2003) Maturation of human central auditory system activity: the T-complex. Clin Neurophysiol 114:685-701.

Uchanski RM (2005) Clear Speech. In: Handbook of Speech Perception (Pisoni DB, Remez RE, eds), pp 207-235. Malden, MA: Blackwell Publishers.

Vernier VG, Galambos R (1957) Response of single medial geniculate units to repetitive click stimuli. Am J Physiol 188:233-237.

Wagner AD, Pare-Blagoev EJ, Clark J, Poldrack RA (2001) Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. Neuron 31:329-338.

Wagner J, Manning P (1976) The biology of the guinea pig. New York: Academic Press.

Wallace MN, Rutkowski RG, Palmer AR (2000) Identification and localisation of auditory areas in guinea pig cortex. Exp Brain Res 132:445-456.

Wallace MN, Rutkowski RG, Palmer AR (2005a) Responses to the purr call in three areas of the guinea pig auditory cortex. Neuroreport 16:2001-2005.

Wallace MN, Shackleton TM, Anderson LA, Palmer AR (2005b) Representation of the purr call in the guinea pig primary auditory cortex. Hear Res 204:115-126.

Wang X, Kadia SC (2001) Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. J Neurophysiol 86:2616-2620.

Wang X, Lu T, Liang L (2003) Cortical processing of temporal modulations. Speech Commun 41:107-121.

Wang X, Merzenich MM, Beitel R, Schreiner CE (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. J Neurophysiol 74:2685-2706.

Wang XQ (2000) On cortical coding of vocal communication sounds in primates. Proceedings of the National Academy of Sciences of the United States of America 97:11843-11849.

Warrier CM, Johnson KL, Hayes EA, Nicol T, Kraus N (2004) Learning impaired children exhibit timing deficits and training-related improvements in auditory cortical responses to speech in noise. Exp Brain Res 157:431-441.

Watson M, Stewart M, Krause K, Rastatter M (1990) Identification of time-compressed sentential stimuli by good vs poor readers. Percept Mot Skills 71:107-114.

Wernicke C (1874) Der Aphasische Symptomencomplex. Breslau: Cohn and Weigert.

White S, Milne E, Rosen S, Hansen P, Swettenham J, Frith U, Ramus F (2006) The role of sensorimotor impairments in dyslexia: a multiple case study of dyslexic children. Dev Sci 9:237-255; discussion 265-239.

Wible B, Nicol T, Kraus N (2004) Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. Biol Psychol 67:299-317.

Wible B, Nicol T, Kraus N (2005) Correlation between brainstem and cortical auditory processes in normal and language-impaired children. Brain 128:417-423.

Wilkinson G (1993) Wide range Achievement Test-3 (WRAT-3). Wilmington, DE: Jastak Associates.

Witton C, Stein JF, Stoodley CJ, Rosner BS, Talcott JB (2002) Separate influences of acoustic AM and FM sensitivity on the phonological decoding skills of impaired and normal readers. J Cogn Neurosci 14:866-874.

Wolpaw JR, Penry JK (1975) A temporal component of the auditory evoked response. Electroencephalogr Clin Neurophysiol 39:609-620.

Wolpaw JR, Penry JK (1977) Hemispheric differences in the auditory evoked response. Electroencephalogr Clin Neurophysiol 43:99-102.

Woodcock R, Johnson M (1977) Woodcock-Johnson Psychoeducational Battery: Tests of Cognitive Ability. Allen, TX: DLM Teaching Resources.

Woodcock R, Johnson M (1989) Woodcock-Johnson Psychoeducational Battery Revised: Tests of Cognitive Ability. Allen, TX: DLM Teaching Resources.

Woolsey TA (1997) In: Encyclopedia of Neuroscience (Adelman G, Smith B, eds), pp 195-199. Amsterdam: Elsevier.

Xiao Z, Suga N (2002) Modulation of cochlear hair cells by the auditory cortex in the mustached bat. Nat Neurosci 5:57-63.

Young E (1998) Cochlear Nucleus. In: Synaptic Organization of the Brain, 4th Edition (Shepherd G, ed), pp 121-158. New York: Oxford Univ. Press.

Zaehle T, Wustenberg T, Meyer M, Jancke L (2004) Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. Eur J Neurosci 20:2447-2456.

Zatorre RJ, Belin P (2001) Spectral and temporal processing in human auditory cortex. Cereb Cortex 11:946-953.

Zatorre RJ, Belin P, Penhune VB (2002) Structure and function of auditory cortex: music and speech. Trends Cogn Sci 6:37-46.

Ziegler JC, Pech-Georgel C, George F, Alario FX, Lorenzi C (2005) Deficits in speech perception predict language learning impairment. Proc Natl Acad Sci U S A 102:14110-14115.

APPENDIX ITEMS

AUDITORY PATHWAY REPRESENTATIONS OF SPEECH SOUNDS IN HUMANS

Book chapter to be published in the Handbook of Clinical Audiology
(Katz J, Hood L, Burkard R, Medwetsky L, eds).

Daniel A. Abrams[1] and Nina Kraus[1,2]

[1]Auditory Neuroscience Laboratory, Department of Communication Sciences, and [2]Departments of Neurobiology and Physiology and Otolaryngology. Northwestern University, 2240 Campus Drive, Evanston, Illinois, 60208.

Introduction

An essential function of the central auditory system is the neural encoding of speech sounds. The ability of the brain to translate the acoustic events in the speech signal into meaningful linguistic constructs relies in part on the representation of the acoustic structure of speech by the central nervous system. Consequently, an understanding of how the nervous system accomplishes this task would provide important insight into the basis of language perception and cognitive function.

One of the challenges faced by researchers interested in this subject is that speech is a complex acoustic signal that is rich in both spectral and temporal features. In everyday listening situations,

the abundance of acoustical cues in the speech signal provides enormous perceptual benefits to listeners. For example, it has been shown that listeners are able to shift their attention between different acoustical cues when perceiving speech from different talkers to compensate for the inherent variability in the acoustical properties of speech between individuals (Nusbaum and Morin, 1992).

There are two basic approaches that researchers have adopted for conducting experiments on speech perception and underlying physiology. One approach uses "simple" acoustic stimuli, such as tones and clicks, as a means to control for the complexity of the speech signal. While simple stimuli enable researchers to reduce the acoustics of speech to its most basic elements, the auditory system is non-linear (Sachs and Young, 1979; Sachs et al., 1983; Rauschecker, 1997; Nagarajan et al., 2002), and, therefore responses to simple stimuli generally do not accurately predict responses to actual speech sounds. A second approach uses speech and speech-like stimuli (Song et al., 2006). There are many advantages to this approach. First, these stimuli are more ecologically valid than simple stimuli. Second, a complete description of how the auditory system responds to speech can only be obtained by using speech stimuli, given the non-linearity of the auditory system. Third, long-term exposure to speech sounds and the subsequent use of these speech sounds in linguistic contexts induces plastic changes in the auditory pathway which may alter neural representation of speech in a manner that cannot be predicted by simple stimuli. Fourth, when speech stimuli are chosen carefully, the acoustic properties of the signal can still be well controlled.

This chapter reviews the literature that has begun to elucidate how the human auditory system encodes acoustic features of speech. This chapter is organized into five sections, with each section describing what is currently known about how the brain represents a particular acoustic feature present in speech (see Table 1). These acoustic features of speech were chosen because of their essential roles in normal speech perception. Each section contains a description of the acoustical feature, an elaboration of its relevance to speech perception, followed by a review and assessment of the data for that acoustic feature.

An important consideration is that the acoustical features described in this chapter are not mutually exclusive. For example, one section of this chapter describes the neural encoding of "periodicity," which refers to acoustical events that occur at regular time intervals. Many features in the speech signal are periodic, however describing the neurophysiologic encoding of all of the periodic features that are processed simultaneously in the speech stimulus in a study of the auditory system would be experimentally unwieldy. Consequently, for the sake of simplicity, and to reflect the manner in which they have been investigated in the auditory neuroscience literature, some related acoustical features will be discussed in separate sections. Efforts will be made throughout the chapter to identify when there is overlap among acoustical features.

The signal: basic speech acoustics

The speech signal can be described according to a number of basic physical attributes (Johnson, 1997). An understanding of these acoustic attributes is essential to any discussion of how the auditory system encodes speech. The linguistic roles of these acoustic features are described separately within each section of the chapter.

Fundamental frequency. The fundamental frequency is a low frequency component of speech that results from the periodic beating of the vocal folds. In Figure 1a, the frequency content of the naturally produced speech sentence "The Young Boy Left Home" is plotted as a function of time: greater amounts of energy at a given frequency are represented with red lines while smaller amounts of energy are depicted in blue. The fundamental frequency can be seen as the horizontal band of energy in Figure 1a that is closest to the x-axis (i.e., lowest in frequency). The fundamental frequency is notated F0 and provides the perceived pitch of an individuals voice.

Harmonic structure. An acoustical phenomenon that is related to the fundamental frequency of speech is known as the harmonic structure of speech. Speech harmonics, which are integer multiples of the fundamental frequency, are present in ongoing speech. The harmonic structure of speech is displayed in Figure 1a as the regularly spaced horizontal bands of energy seen throughout the sentence.

Formant structure. Another essential acoustical feature of speech is the formant structure. Formant structure describes a series of discrete peaks in the frequency spectrum of speech that are the result of an interaction between the frequency of vibration of the vocal folds and the resonances within a speaker's vocal tract. The frequency of these peaks, as well as the relative frequency between peaks, varies for different speech sounds. The formant structure of speech interacts with the harmonic structure of speech: the harmonic structure is represented by integer multiples of the fundamental frequency, and harmonics that are close to a resonant frequency of the vocal tract are formants. In Figure 1, the formant structure of speech is represented by the series of horizontal, and occasionally diagonal, red lines that run through most of the speech utterance. The word "left" has been enlarged in Figure 1b to better illustrate this phenomenon.

The broad and dark red patches seen in this figure represent the peaks in the frequency spectrum of speech that are the result of an interaction between the frequency of vibration of the vocal folds and the resonances of a speaker's vocal tract. The frequency of these peaks, as well as the relative frequency between peaks, varies for different speech sounds within the sentence. The lowest frequency formant is known as the first formant and is notated F1, while subsequent formants are notated F2, F3, etc.

The measures of brain activity

We begin by describing the neurophysiologic measures that have been used to probe auditory responses to speech and speech-like stimuli (comprehensive descriptions of these measures can be found elsewhere (Sato, 1990; Hall, 1992; Jezzard et al., 2001) as well as in chapters in this text. Historically, the basic research on the neurophysiology of speech perception has borrowed a number of clinical tools to assess auditory system function.

Brainstem responses

The auditory brainstem response (ABR) consists of small voltages originating from auditory structures in the brainstem in response to sound. While these responses do not pinpoint the specific origin of auditory activity among the auditory brainstem nuclei, the great strength of the ABR (and auditory potentials in general) is that they precisely reflect the time-course of neural activity at the microsecond level. The ABR is typically measured with a single active electrode referenced to the earlobe or nose. Clinical evaluations using the ABR typically use brief acoustic stimuli, such as clicks and tones, to elicit brainstem activity. The auditory brainstem response is unique among the AEPs because of the remarkable reliability of this response, both within and

across subjects. In the clinic, the ABR is used to assess the integrity of the auditory periphery and lower brainstem (Hall, 1992). The response consists of a number of peaks, with wave V being the most clinically reliable. Deviations on the order of microseconds are deemed "abnormal" in the clinic, and are associated with some form of peripheral hearing damage or with retrocochlear pathologies. Research using the ABR to probe acoustic processing of speech utilizes similar recording procedures, but different acoustic stimuli.

Cortical responses

Cortical evoked potentials and fields

Cortical evoked responses are used as a research tool to probe auditory function in normal and clinical populations. Cortical evoked potentials are small voltages originating from auditory structures in the cortex in response to sound. These potentials are typically measured with multiple electrodes, often referenced to a "common reference," which is the average response measured across all electrodes. Cortical evoked "fields" are the magnetic counterpart to cortical evoked potentials; however instead of measuring voltage across the scalp, the magnetic fields produced by brain activity are measured. Electroencephalography is the technique by which evoked potentials are measured and magnetoencephalography (MEG) is the technique by which evoked fields are measured. Similar to the ABR, the strength of assessing cortical evoked potentials and fields is that they provide detailed information about the time-course of activation and how sound is encoded by temporal response properties of large populations of auditory neurons, although this technique is limited in its spatial resolution. Due to large inter- and intra-subject variability in cortical responses, they are not generally used clinically. Results from these two methodologies are generally compatible, despite some differences in the neural

generators that contribute to each of these responses. Studies using both EEG and MEG are described interchangeably throughout this chapter despite the subtle differences between the measures. The nomenclature of waveform peaks is similar for EEG and MEG: typically an N or P, depicting a negative or positive deflection, followed by a number indicating the approximate latency of the peak. Finally, the letter "m" follows the latency for MEG results. For example, N100/N100m are the labels for a negative deflection at 100 msec as measured by EEG and MEG, respectively.

Functional imaging

Functional imaging of the auditory system is another often-used technique to quantify auditory activity in the brain. The technology that is used to measure these responses, as well as the results they yield, is considerably different from the previously described techniques. The primary difference is that functional imaging is an indirect measure of neural activity; that is, instead of measuring voltages or fields resulting from activity in auditory neurons, functional imaging measures hemodynamics, a term used to describe changes in metabolism as a result of changes in brain activity. The data produced by these measures is a three-dimensional map of activity within the brain as a result of a given stimulus. The strong correlation between actual neural activity and blood flow to the same areas of the brain (Smith et al., 2002a) has made functional imaging a valuable investigative tool to measure auditory activity in the brain. The two methods of functional imaging described here are functional magnetic resonance imaging (fMRI) and positron emission tomography (PET). The difference between these two techniques is that fMRI measures natural levels of oxygen in the brain, as oxygen is consumed by neurons

when they become active. PET, however, requires the injection of a radioactive isotope into a subject. The isotope emits positrons, which can be detected by a scanner, as it circulates in the subject's bloodstream. Increases in neural activity draws more blood, and consequently more of the radioactive isotope, to a given region of the brain. The main advantage that functional imaging offers relative to evoked potentials and evoked fields is that it provides extremely accurate spatial information regarding the origin of neural activity in the brain. A disadvantage is the poor resolution in the temporal domain: neural activity is often integrated over the course of seconds, which is considered extremely slow given that speech token are as brief as 30 msec. Although recent work using functional imaging has begun describing activity in subcortical regions, the work described here will only cover studies of temporal cortex.

Acoustic features of speech

Periodicity

Definition and role in the perception of speech

Periodicity refers to regular temporal fluctuations in the speech signal between 50-500 Hz (Rosen, 1992). Important aspects of the speech signal that contain periodic acoustic information include the fundamental frequency and all components of the formant structure (note that encoding of the formant structure of speech is covered in a later section). The acoustic information provided by periodicity conveys both phonetic information as well as prosodic cues, such as intonation and stress, in the speech signal. As stated in Rosen's paper, this category of temporal information represents both the periodic features in speech, as well as the distinction between periodic and aperiodic portions of the signal, which fluctuate at much faster rates.

This section will review studies describing the neural representation of relatively stationary periodic components in the speech signal, most notably the fundamental frequency. An understanding of the mechanism for encoding a simple periodic feature of the speech signal, the F0, will facilitate descriptions of complex periodic features of the speech signal, such as the formant structure and frequency modulations.

Physiologic representation of periodicity in the human brain

Auditory brainstem

The short-latency frequency-following response (FFR) is an electrophysiological measure of phase-locked neural activity originating from brainstem nuclei that represents responses to periodic acoustic stimuli up to approximately 1000 Hz (Smith et al., 1975; Stillman et al., 1978; Gardi et al., 1979; Galbraith et al., 2000). Based on the frequency range that can be measured with the FFR, a representation of the fundamental frequency can be measured using this methodology (Cunningham et al., 2001; King et al., 2002; Krishnan et al., 2004; Russo et al., 2004; Wible et al., 2004; Johnson et al., 2005; Krishnan et al., 2005; Russo et al., 2005), as well as the F1 in some instances (encoding of F1 is discussed in detail in the Formant Structure section).

A number of studies have shown that F0 is represented within the steady-state portion of the brainstem response (i.e., FFR) according to a series of negative peaks which are temporally spaced in correspondence to the wavelength of the fundamental frequency. An example of F0 representation in the FFR can be seen in Figure 2, which shows the waveform of the speech stimulus /da/ (top), an experimental stimulus that has been studied in great detail, as well as the

brainstem response to this speech sound (bottom). A cursory inspection of this figure shows that the primary periodic features of the speech waveform provided by the F0 are clearly represented in peaks D, E, and F of the FFR brainstem response. Importantly, it has been shown that the FFR is highly sensitive to F0 frequency; this aspect of the brainstem response accurately "tracks" modulations in frequency (Krishnan et al., 2004), a topic which is discussed in depth in the Frequency Transitions section of this chapter.

An hypothesis regarding the brainstem's encoding of different aspects of the speech signal has been proposed in a recent paper (Kraus and Nicol, 2005). Specifically, it is proposed that the source (referring to vocal-fold vibration) and filter aspects (vocal musculature in the production of speech) of a speech signal, show dissociation in their acoustical representation in the auditory brainstem. The source portion of the brainstem's response to speech is the representation of the F0, while the filter refers to all other features, including speech onset, offset and the representation of formant frequencies. For example, it has been demonstrated that brainstem responses are correlated within source and filter classes but are not correlated between classes (Russo et al., 2004). Moreover, in a study of children with language-learning disabilities, whose behavioral deficits may be attributable to central auditory processing disorders, it has been shown that source representation in the auditory brainstem is normal while filter class representation is impaired (Cunningham et al., 2001; King et al., 2002; Hayes et al., 2003; Wible et al., 2004, , 2005). These data suggest that the acoustical representations of source and filter aspects of a given speech signal are differentially processed, and provide evidence for neural specialization at the level of the brainstem. Additionally, it is proposed that this scheme may constitute brainstem origins for cortical 'what' 'where' pathways (Kraus and Nicol, 2005).

Cortex

It has been shown that neurons in the auditory cortex respond robustly with time-locked responses to slow rates of stimulation (< ~25 Hz), and generally do not phase-lock to frequencies greater than approximately 100 Hz (Creutzfeldt et al., 1980; Eggermont, 1991; Steinschneider et al., 1998; Lu et al., 2001). Therefore, cortical phase-locking to the fundamental frequency of speech, which is greater than 100 Hz, is poor, and it is generally thought that the brainstem's phase-locked (i.e., linear) representation of F0 is transformed at the level of cortex to a more abstract representation. For example, it has been shown that cortical neurons produce sustained, non-synchronized discharges throughout a high frequency (>50 Hz) stimulus (Lu et al., 2001), which is a more abstract representation of the stimulus frequency compared to time-locked neural activation.

An important aspect of F0 perception is that listeners native to a particular language are able to perceive a given speech sound as invariant regardless of the speaker's F0, which varies considerably among men (F0 ~100 Hz), women (F0 ~200 Hz) and children (F0 up to 400 Hz). For example, the speech sound "dog" is categorized by a listener to mean the exact same thing regardless of whether an adult or a child produces the vocalization, even though there is a considerable difference in the acoustic properties of the adult's and child's vocalization with respect to the fundamental frequency. To address how auditory cortical responses reflect relatively large variations in F0 between listeners, N100m cortical responses were measured with MEG for a set of Finnish vowel and vowel-like stimuli that varied in F0 while keeping all other formant information (F1-F4) constant (Makela et al., 2002). Results indicated that N100m

responses were extremely similar in spatial activation pattern and amplitude for all vowel and vowel-like stimuli, irrespective of the F0. This is a particularly intriguing finding given that N100m responses differed when 100 Hz, 200 Hz, and 400 Hz pure-tone stimuli were presented to the same subjects in a control condition. The similarity of the speech-evoked brain responses, which were independent of the F0 frequency, suggests that variances in F0 may be filtered out of the neural representation by the time it reaches the cortex. The authors suggest that the insensitivity of cortical responses to variations in the F0 may facilitate the semantic categorization of the speech sound. In other words, since the F0 does not provide essential acoustic information relevant to the semantic meaning of the speech sound, it may the case that the cortex does not respond to this aspect of the stimulus in favor of other acoustic features that are essential for decoding word meaning.

In summary, periodicity of the fundamental frequency is robustly represented in the FFR of the auditory brainstem response. Moreover, the representation of the fundamental frequency is normal in learning disabled children despite the abnormal representations of speech-sound onset and first formant frequency. This disparity in the learning disabled auditory system provides evidence that different features of speech sounds may be served by different neural mechanisms and/or populations. In the cortex, MEG results show that cortical responses are relatively insensitive to changes in the fundamental frequency of speech sounds, suggesting that the differences between F0s between speakers is filtered out by the time it reaches the level of auditory cortex.

Formant structure

Role in the perception of speech

Formant structure describes a series of discrete peaks in the frequency spectrum of speech that are the result of an interaction between the frequency of vibration of the vocal folds and the resonances within a speaker's vocal tract (see Introduction for a more complete acoustical description of the formant structure). The formant structure is a dominant acoustic feature of sonorants, a class of speech sounds that includes vowels, approximants and nasals. The formant structure has a special role in the perception of vowels in that formant frequencies, particularly the relationship between F1 and F2 (Peterson and Barney, 1952), are the primary phonetic determinants of vowels. For example, the essential acoustic difference between /u/ and /i/ is a positive shift in F2 frequency (Peterson and Barney, 1952). Due to the special role of formants for vowel perception, much of the research regarding the formant structure of speech uses vowel stimuli.

Physiologic representation of formant structure in the human brain

Auditory brainstem

The question of how the human auditory brainstem represents important components of the formant structure was addressed in a study by Krishnan (Krishnan, 2002). In this study, brainstem (FFR) responses to three steady-state vowels were measured and the spectral content of the responses were compared to that of the vowel stimuli. All three of the stimuli had approximately the same fundamental frequency, however the first two formant frequencies were different in each of the vowel stimuli. Results indicate that at higher stimulus intensities the brainstem FFR accurately represents F1 and F2, however the representation of F1 has an increased representation relative to F2. The author indicates the similarity between this finding

and a similar result in a classic study of vowel representation in the auditory nerve of anesthetized cats (Sachs and Young, 1979) in which the predominance of the representation to F1 was also demonstrated. These data provide evidence that phase-locking serves as a mechanism for encoding critical components of the formant structure not only in the auditory nerve, but also in the auditory brainstem.

Auditory cortex

A number of studies have described the representation of formant structure in the human cortex as a means of investigating whether a cortical map of phonemes, termed the "phonemotopic" map, exists in the human brain. Specifically, researchers want to know if the phonemotopic map is independent of the tonotopic map, or alternatively whether phonemes are more simply represented according to their frequency content along the tonotopic gradient in auditory cortex. To this end, investigators have measured cortical responses to vowel stimuli, a class of speech sounds that differ acoustically from one another according to the distribution of F1-F2 formant frequencies. Vowel stimuli also offer the advantage of exhibiting no temporal structure beyond the periodicity of the formants.

The method that has been used to investigate the relationship between the tonotopic map in human auditory cortex and the representation of formant structure has been to compare cortical source locations for tones and for specific speech sounds with similar frequency components. For example, in one study (Diesch and Luce, 1997) N100m source location was measured in response to separately presented 600 Hz and 2100 Hz pure tones as well as a two tone composite signal comprising the component pure tones (i.e., simultaneous presentation of the 600 Hz and

2100 Hz pure tones). These responses were compared to isolated formants, defined as the first and second formant frequencies of a vowel stimulus, complete with their harmonic structure, separated from the rest of the frequency components of the stimulus (i.e., F0, higher formant frequencies). These isolated formants had the same frequency as the tonal stimuli (i.e., 600 Hz and 2100 Hz). Finally, a two formant composite signal, which constituted a vowel, was also presented. Results indicated that the N100m source in response to the vowel stimulus was different in location from that predicted by both the pure-tone responses and by the superposition of responses to the component single formant stimuli. These data indicate that formant structure is spatially represented in human cortex differently than the linear sum of responses to the component formant stimuli, and suggest that formant structure is represented orthogonal to the tonotopic map. The authors of this work hypothesize that the different spatial representation of the vowel stimuli reflects the additional acoustic components of the vowel stimuli, including the harmonic and formant structures. The authors of this work refrain from a potentially more intriguing conclusion; that is, does the spatial representation of the vowel stimuli in some way reflect the behavioral experience of the subjects with these speech sounds. For example, it is possible that a larger, or different, population of cortical neurons is recruited for sounds that are familiar, or have significant ecological importance, relative to the population recruited for pure tones or single formant frequencies, and that the source location for the vowels reflects this phenomenon.

Additional studies have attempted to better describe the acoustic representation of vowels in the human brain. In one study, Obleser and colleagues (Obleser et al., 2003) addressed the neurophysiology underlying a classic study of speech acoustics in which it was shown that the

distinction of vowels is largely carried by the frequency relationship of F1 and F2 (Peterson and Barney, 1952). To this end, cortical source locations were measured in response to German vowels that naturally differ in F1-F2 relationships. Results indicated that the location of the N100m source reflects the frequency relationship of the F1-F2 formant components. This finding was replicated in a second study using 450 natural speech exemplars of three Russian vowels; again, the spectral distance between F1 and F2 was reflected in the dipole location of N100m responses (Shestakova et al., 2004). In both studies, the authors suggest that cortical sensitivity to F1-F2 differences can be explained by inhibitory response patterns in the auditory cortex: the closer the F1 frequency is to F2, the greater the reciprocal neural inhibition, which, in turn, could influence the location of the dipole source as measured by MEG (Obleser et al., 2003).

While these studies provide evidence that the cortex represents the formant structure of vowels in a manner that is (a) unrelated to the tonotopic map and (b) organized according to the perceptually essential formant frequencies, these findings require a number of caveats. First, the source locations described in these studies represent the center of gravity, as a single point in three dimensional space in the cortex, of the neural contributors to a given N100m response (Naatanen and Picton, 1987). As it is known that the N100 response has as many as 6 separate cortical generators, the N100m sources for even a simple cortical map (i.e., the tonotopic map), let alone a complex map such as the putative phonemotopic map, represent at least a partial abstraction of the underlying anatomy and should not be viewed as an exact representation of well-described auditory maps in animal models (Schreiner, 1998). This is particularly relevant given that the clear tonotopic gradient in auditory cortex is no longer apparent when pure-tone

stimuli are presented above 50 dB SPL (Schreiner, 1998), such as the levels used in the MEG experiments described in this section. In addition, it has not yet been definitively shown that the neural representations of phonemes described in these studies truly constitute a "phonemotopic" map. The presence of a phonemotopic map suggests behavioral relevance of phoneme stimuli beyond their acoustic attributes. None of the studies described here have tested if cortical responses to the F1-F2 components for non-native vowel sounds show similar sensitivity as native phonemes. Despite these limitations, these studies provide consistent evidence that a perceptually-critical aspect of the formant structure of vowels, the F1-F2 relationship, is represented in a spatial map in auditory cortex as early as ~100 msec post-stimulus onset.

Another line of evidence has used functional imaging to show the particular regions of the temporal cortex that are sensitive to the formant structure of speech sounds relative to other natural and vocally generated (i.e., laughs, coughs) sounds (Belin et al., 2000). Cortical responses to natural vocal stimuli were compared to vocal stimuli in which the formant structure of speech was replaced by white noise and scrambled vocal sounds. All stimuli were matched for overall RMS energy. In both of these experimental conditions, the altered spectral information was modulated by the original amplitude envelope of the speech signal. Results from this experiment indicated that all stimuli activated regions along the superior temporal sulcus (STS), a cortical region consisting of unimodal auditory and multimodal areas that is hypothesized to be a critical speech-processing center subsequent to more rudimentary acoustic processing in the superior temporal gyrus. However, responses to the natural vocal stimuli were significantly larger and more widespread throughout the STS, particularly in the right hemisphere, than for the spectrally manipulated vocal stimuli. These data indicate that the

formant structure of speech deeply affects activity patterns in the STS, a speech-selective region of temporal cortex, even when the temporal components of the signals are held constant. Moreover, these data suggest a right-hemisphere bias for processing the formant structure, which supports the more general hypothesis that the right-hemisphere is dominant for resolving spectral components in acoustic signals (Zatorre and Belin, 2001; Zatorre et al., 2002).

An interesting consideration is how cortical asymmetries in response to the acoustic features of speech relate to well-established cerebral asymmetries for higher-order language processing, such as phonemic and semantic processing (Geschwind and Galaburda, 1985; Binder et al., 1997; Hickok and Poeppel, 2004), which are strongly lateralized to the left hemisphere . While a direct link between these forms of asymmetry has not been established, a plausible scenario is that the acoustic-level asymmetries precede, and serve as the input to, phonemic and semantic processing in left-hemisphere language regions. If this is the case, it remains to be seen what physiologic advantage a right-hemisphere preference for formant structure processing (Belin et al., 2000) might offer given that phonemic and semantic processing of speech stimuli takes place in the opposite hemisphere, thereby requiring transmission through the corpus callosum. Future studies investigating acoustic-level asymmetries and their interface with higher-order language asymmetries would provide essential information regarding the functional neuroanatomy of speech perception.

In summary, the brainstem encodes lower formant frequencies, which are critical to vowel perception, with phase-locked responses. Converging evidence indicates that the cortex encodes a perceptually essential aspect of the formant structure of speech. Specifically, the F1-F2

relationship is spatially mapped in the cortex at ~100 msec post-stimulus onset as measured by N100m source location. In addition, functional imaging data provide evidence that the STS, a non-primary region of temporal cortex, is more responsive to speech stimuli that contain formant structure than speech in which the formant structure has been replaced with other sounds. Together, these results suggest that both primary and non-primary regions of temporal cortex are sensitive to aspects of the formant structure that are essential for normal perception.

Frequency transitions

Acoustic description and role in the perception of speech

Frequency transitions of the fundamental and formant frequencies are ubiquitous in ongoing speech. In English, modulation of the fundamental frequency typically does not provide segmental cues, rather it provides suprasegmental cues such as the intent (e.g., question or statement) and emotional state of the speaker. In other languages, such as Mandarin and Thai, modulations to the fundamental frequency provide phonetic cues. Formant transitions on the other hand are critical to speech perception of English in that they serve as a cue for consonant identification and signal the presence of diphthongs and glides (Lehiste and Peterson, 1961). Moreover, formant transitions also have been shown to play a role in vowel identification (Nearey and Assmann, 1986). The movements of formant frequencies can be distilled to three basic forms that occur during an ongoing sequence of phonemes, (taken from Lehiste and Peterson, 1961): (a) the movement of a formant from the initiation of the consonant until the beginning of the vowel in a consonant-vowel combination, (b) the movement of a formant from one vowel to another vowel (i.e., in a diphthong), and (c) formant movement from a vowel until vowel termination for a vowel-consonant combination. The frequency modulations that occur

during formant transitions can occur at relatively fast rates (~40 msec) while spanning large frequency ranges (>2000 Hz in F2 transitions).

Physiologic representation of frequency transitions in the human brain

Auditory brainstem

The short-latency FFR is able to "track," or follow, frequency changes in speech. This phenomenon was demonstrated in a study of FFR tracking of the fundamental frequency (F0) in Mandarin speech sounds (Krishnan et al., 2004). In this study, FFR to four different tonal permutations of the Mandarin word "yi" were measured in a group of native Mandarin speakers. Specifically, synthetic stimuli consisted of "yi" pronounced with (1) a flat F0 contour, (2) a rising F0 contour, (3) a falling F0 contour, and (4) a concave F0 contour that fell then rose in frequency. In Mandarin, which is a "tonal" language, these four stimuli are different words: the F0 contour provides the only acoustic cue to differentiate them. Results indicated that the FFR represented the fundamental frequency modulations for all of the stimulus conditions irrespective of the form of the frequency contour. These data indicate that the FFR represents phase-locked activity in the brainstem for rapidly changing frequency components in speech, an essential acoustical cue for consonant identification.

A similar methodology was used in another study by Krishnan and colleagues to investigate the role of language experience on auditory brainstem encoding of pitch (Krishnan et al., 2005). FFRs to the "yi" stimuli described above were measured in native Mandarin speakers as well as native speakers of American English, to whom the stimuli bear no linguistic value. Results from this study indicate greater FFR pitch strength and pitch tracking in the Chinese subjects

compared to the native English speakers across all four of the Mandarin tones. The FFR of the Chinese subjects also indicated increased harmonic representation of the fundamental frequency (i.e., larger neural representation of the harmonic content of the F0) compared to the English speakers. These data indicate that responses from the auditory brainstem reflect the behavioral experience of a listener by enhancing the neural representation of linguistically relevant acoustic features.

An hypothesis proposed by Ahissar and Hochstein (Ahissar and Hochstein, 2004) may explain how experience engenders plasticity at low levels of sensory systems. Their "Reverse Hierarchy" theory proposes that when a naïve subject attempts to perform a perceptual task, the performance on that task is governed by the "top" of a sensory hierarchy. As this "top" level of the system masters performance of the task, over time, lower levels of the system are modified and refined to provide more precise encoding of sensory information. This can be thought of as efferent pathway-mediated tuning of efferent sensory input. While the reverse hierarchy theory does not explicitly discuss plasticity of the brainstem, this theory could account for the findings of Krishnan. Specifically, due to the importance of extracting lexical information present in pitch contours, native Mandarin speakers are "expert" at encoding this acoustic feature, which is accomplished, at least in part, by extreme precision and robustness of sensory encoding in low levels of the auditory system such as the brainstem. Native English speakers, who are not required to extract lexical meaning from pitch contours, are relative novices at this form of pitch tracking, and consequently their brainstems have not required this level of modification.

An interesting question that was not addressed in this study, but was proposed as a discussion item, is whether native Mandarin speakers are better than English speakers at pitch tracking the F0 exclusively for familiar speech sounds or whether Mandarin speakers' superior performance would extend to all periodic acoustic signals, including non-native speech sounds. This question would address whether a lifetime of experience using F0 to extract linguistic meaning generally improves the auditory system's ability to track all types of pitches, or alternatively if this phenomenon is exclusive to pitches present in familiar speech sounds. Data from our lab suggests that another form of long-term auditory experience, musicianship, contributes to enhanced neural encoding of speech sounds in the auditory brainstem relative to non-musicians (Wong et al., 2004). This finding provides evidence that expertise associated with one type of acoustic signal (i.e., music) provides a general augmentation of the auditory system that is manifested in brain responses to another type of acoustic signal (i.e., speech), and indicates that auditory experience can modify basic sensory encoding.

Auditory cortex

Similar to Krishnan's work involving the brainstem, multiple studies have investigated cortical processing of F0 pitch contours and its relationship to language experience (Gandour et al., 1998; Klein et al., 2001; Wang et al., 2001). The most convincing of these studies is that by Wong et al. (2004). In this study, native Mandarin and native English speakers underwent PET scanning during passive listening and while performing a pitch discrimination task. Stimuli consisted of (a) Mandarin speech sounds that contained modulations of the fundamental frequency which signal lexical meaning and (b) English speech sounds which also contained modulations to the fundamental frequency, however, F0 modulations never provide lexical information in English.

Imaging results indicated that native Mandarin speakers showed significant activation of the left anterior insular cortex, adjacent to Broca's area, only when discriminating Mandarin speech sounds; the homologous right anterior insula was activated when this group discriminated English speech sounds, as well as when native English speakers discriminated both Mandarin and English speech sounds. These data suggest that the left anterior insula is involved in auditory processing of modulations to the fundamental frequency only when those modulations are associated with lexical processing. Moreover, these data suggest that the neural processing of acoustic signals is context dependent and is not solely based on the acoustical attributes of the stimuli.

In addition to studies of the neural representation of F0 modulations, a number of studies have also addressed the cortical representation of formant frequency modulation in humans. As it is known that neurons in auditory cortex do not phase-lock to frequencies greater than approximately 100 Hz (Creutzfeldt et al., 1980; Eggermont, 1991; Steinschneider et al., 1998; Lu et al., 2001), and the formant structure of speech consists of frequencies almost exclusively above 100 Hz, the cortical representation of frequency modulation as measured by evoked potentials is abstract (i.e., not represented with time-locked responses) relative to those described for the auditory brainstem. One cortical mechanism that has received considerable attention for the processing of rapid formant modulations is that of asymmetric processing in the left hemisphere auditory cortex. A more general hypothesis proposes that left-hemisphere auditory cortex is specialized for all forms of rapid acoustic stimuli and serves as an early acoustic analysis stage at the level of the cortex. A significant piece of evidence in support of this hypothesis was provided in a study of cortical activation patterns for rapid and slow formant

frequency modulations (Belin et al., 1998). In this study, non-speech sounds containing temporal and spectral characteristics similar to speech sounds were presented to listeners as they were PET-scanned. Non-speech sounds were used so that any cortical asymmetry could not be associated with well-known asymmetries for language processing. Results indicated that the left superior temporal gyrus (STG), including primary auditory cortex, showed greater activation than the right STG for rapid (40 msec) formant frequency transitions but not for slow (200 msec) transitions. In addition, a left-hemisphere region of prefrontal cortex was asymmetrically activated for the rapid formant transition, which was corroborated in a separate fMRI study that used nearly identical acoustic stimuli (Temple et al., 2000). These data suggest that left-hemisphere auditory regions preferentially process rapid formant modulations present in ongoing speech.

In summary, results measured from the auditory brainstem indicate that modulations in the fundamental frequency of speech are faithfully encoded in the FFR. Moreover, these particular brainstem responses appear to be shaped by linguistic experience, a remarkable finding which indicates that cognitive processes (e.g., language) influence basic sensory processing. In the cortex, a mechanism for encoding frequency modulation is the specialization of left hemisphere auditory regions, and results indicate that rapid frequency changes in speech-like stimuli preferentially activate the left hemisphere relative to slower frequency changes. In addition, the anterior insular cortex is activated for the processing of F0 modulations: the left hemisphere insula is specifically activated when F0 modulations provide lexical information to a native speaker, while the right hemisphere insula is activated when F0 modulations do not provide lexical information. These cortical findings would appear to be contradictory: the former

indicates asymmetric activation by left-hemisphere structures is based on physical parameters of the speech signal, irrespective of linguistic content, while the latter suggests that linguistic context is essential for left-asymmetric insular processing of F0 modulations. However, Wong et al. (2005) stated that these results can be reconciled if the insular activity shown in their study occurs after the "acoustically specialized" cortical activity described by Belin et al. (1998) and Temple et al. (2000). If this were true, it would indicate two independent levels of cortical asymmetry: one based on the acoustic attributes of the signal and one based on the linguistic relevance to the listener. This hypothesis needs to be tested in future studies.

Acoustic onsets

Acoustic description and role in the perception of speech

Acoustic onsets are defined here as the spectral and temporal features present at the beginning (the initial ~40 msec) of speech sounds. While the acoustics of phonemes are only slightly altered- based on their location in a word (i.e., beginning, middle or end of a word), an emphasis has been put on acoustic onsets in the neurophysiologic literature. Consequently, acoustic onsets are discussed here separately, despite some overlap with acoustic features (i.e., frequency transitions) discussed previously.

Onset acoustics of speech sounds vary considerably in both their spectral and temporal attributes. In some cases, the spectral features of the onset are essential for perception (e.g., the onset frequency of F3 for discriminating /da/ vs. /ga/), while in other cases temporal attributes of onsets are the critical feature for perception. A frequently studied acoustic phenomenon associated with the temporal attributes of speech-sound onset is that of the voice onset time

(VOT), which is present in stop consonants. The VOT is defined as the duration of time between the release of a stop consonant by speech articulators and the beginning of vocal fold vibration. The duration of the VOT is the acoustic cue that enables for differentiation between consonants that are otherwise extremely similar (e.g., /da/ vs. /ta/, /ba/ vs. /pa/, /ga/ vs. /ka/).

Physiologic representation of acoustic onsets in the human brain

Auditory brainstem

The brainstem response to speech-sound onset have been studied extensively (Cunningham et al., 2001; King et al., 2002; Russo et al., 2004; Wible et al., 2004; Banai et al., 2005; Johnson et al., 2005; Kraus and Nicol, 2005; Russo et al., 2005; Wible et al., 2005). The first components of the speech-evoked ABR reflect the onset of the brainstem to the stimulus (Figure 2). Speech onset is represented in the brainstem response at approximately 7 msec in the form of two peaks, positive peak V and negative peak A.

Findings from a number of studies have demonstrated that the brainstem's response to acoustic transients is closely linked to auditory perception and to language-based cortical function such as literacy. These studies have investigated brainstem responses to speech in normal children and children with language-based learning disabilities (LD), a population that has consistently demonstrated perceptual deficits in auditory tasks using both simple (Tallal and Piercy, 1973; Reed, 1989; Hari and Kiesila, 1996; Wright et al., 1997; Hari et al., 1999; Nagarajan et al., 1999; Ahissar et al., 2000; Benasich and Tallal, 2002; Witton et al., 2002) and complex (Tallal and Piercy, 1975; Kraus et al., 1996; Bradlow et al., 1999; Bradlow et al., 2003; Ramus et al., 2003) acoustic stimuli. A general hypothesis proposes a causal link between basic auditory perceptual

deficits in LDs and higher-level language skills, such as reading and phonological tasks (Tallal et al., 1993), although this relationship has been debated (Mody et al., 1997; Schulte-Korne et al., 1998; Bishop et al., 1999; Ramus et al., 2003). In support of a hypothesis linking basic auditory function and language skills, studies of the auditory brainstem indicate a fundamental deficiency in the synchrony of auditory neurons in the brainstem for a significant proportion of language disabled subjects.

The brainstem's response to acoustic transients in speech, features prominently in distinguishing LD from normal (control) subjects. A number of studies have provided compelling evidence that the representation of speech onset (Cunningham et al., 2000; King et al., 2002; Wible et al., 2004; Banai et al., 2005; Wible et al., 2005) is abnormal in a significant proportion of LD subjects. For example, brainstem responses to the speech syllable /da/ were measured for a group of 33 normal and 54 LD children, and a "normal range" was established from the results of the normal subjects (King et al., 2002). Results indicated that 20 LD subjects (37%) showed abnormally late responses to onset peak A. Another study showed a significant difference between normal and LD subjects based on another measure of the brainstem's representation of acoustic transients (Wible et al., 2004). Specifically, it was shown that the slope between onset peaks V and A to the /da/ syllable was significantly smaller in LD subjects compared to normal subjects. The authors of this study indicate that diminished V/A slope demonstrated by LDs is a measure of abnormal synchrony to the onset transients of the stimulus and could be the result of abnormal neural conduction by brainstem generators. The suggestion of abnormal neural conduction is consistent with anatomical findings of deficient axonal myelination in the temporal cortex of LD subjects (Klingberg et al., 2000). In another study (Banai et al., 2005), LD

subjects with abnormal brainstem timing for acoustic transients were more likely to have a more severe form of learning disability, manifested in poorer scores on measures of literacy, compared to LD subjects with normal brainstem responses.

Taken together, these data suggest that the brainstem responses to acoustic transients can not only differentiate a sub-population of LDs from normal subjects, but can also differentiate the LD population in terms of the severity of the disability. Findings from the brainstem measures also indicate a link between sensory encoding and cognitive processes such as literacy. An important question is whether the link between sensory encoding and cognition is a causal one, and if so, whether brainstem deficits are responsible for cortical deficits (or vise versa). Alternatively, these two abnormalities may be merely coincident. Nevertheless, the consistent findings of brainstem abnormalities in a certain portion of the LD population have led to the incorporation of this experimental paradigm into the clinical evaluation of LD and central auditory processing disorders. The "BioMAP" (Biological Marker of Auditory Processing, Biologic Systems Corp., Mundelein, IL) measures and analyzes the brainstem response to speech and has been shown to be a reliable measure for the objective evaluation of children with learning and listening disorders.

Auditory cortex

Cortical encoding of spectral features of speech sound onsets has been reported in the literature, most recently in a paper by Obleser and colleagues (Obleser et al., 2005). In this paper, it was shown that a spectral contrast at speech onset, resulting from consonant place of articulation (i.e., front produced consonant /d/ or /t/ vs. back produced consonant /g/ or /k/), is mapped along the

anterior-posterior axis in auditory cortex as measured by N100m source location. This is significant because it indicates that phonemes differentially activate regions of auditory cortex according to their spectral characteristics at speech onset. It was also shown that the discrete mapping of consonants according to onset acoustics is effectively erased when the speech stimuli are manipulated to become unintelligible despite keeping the spectral complexity of the stimuli largely the same. This stimulus manipulation was accomplished by altering the spectral distribution of the stimuli. The authors argue that this latter finding indicates that the cortex is spatially mapping only those sounds that are intelligible to listeners. These data provide important evidence that cortical spatial representations may serve as an important mechanism for the encoding of spectral characteristics in speech-sound onsets. In addition to differences in spatial representations for place of articulation contrast, cortical responses also showed latency differences for these contrasts. Specifically, it was shown that front consonants, which have higher frequency onsets, elicited earlier N100m responses than back consonants. This finding is consistent with near-field recordings measured from animal models indicating earlier response latencies for speech onsets with higher frequency formants (McGee et al., 1996).

Cortical responses to temporal features of speech sound onsets have also been reported in the literature, all of which have utilized VOT contrasts as stimuli. These studies were performed by measuring obligatory evoked potentials (N100 responses) to continua of consonant-vowel speech sounds that varied gradually according to VOT (Sharma and Dorman, 1999; Sharma and Dorman, 2000),. Additionally, perception of these phonetic contrasts was also measured using the same continua as a means of addressing whether cortical responses reflected categorical perception of the phonemes. Neurophysiologic results indicated that for both /ba/-/pa/ and /ga/-/ka/ phonetic

contrasts, one large negative peak was evident at approximately 100 ms in the response waveform for stimulus VOTs < 40 ms. Importantly, a second negative peak in the response waveform emerged for stimulus VOTs of 40 ms, and this second peak occurred approximately 40 ms after the first peak and was thought to represent the onset of voicing in the stimulus. Moreover, as the VOT of the stimulus increased in duration, the lag between the second peak relative to the first increased proportionally, resulting in a strong correlation between VOT and the latency between the successive peaks (r = ~0.80). The onset of double peaks in cortical responses with a VOT of 40 msec is consistent with neurophysiologic responses measured directly from the auditory cortex of humans (Steinschneider et al., 1999), and an important consideration is that the onset of the double-peak occurred at 40 msec for both /ba/-/pa/ and /ga/-/ka/ phonetic contrasts. In contrast, behavioral results require different VOTs to distinguish the /ba/-/pa/ and /ga/-/ka/ phonetic contrasts. Specifically, a VOT of ~40 ms was required for listeners to correctly identify /pa/ from /ba/, while a VOT of ~60 ms was required for correct identification of /ga/ from /ka/. Taken together, these data indicate that cortical responses reflect the actual VOT at 40 msec irrespective of the categorical perception of the phonetic contrasts, which in the case of the /ga/ - /ka/ contrast requires 60 msec.

Brainstem-cortex relationships

In addition to linking precise brainstem timing of acoustic transients to linguistic function, it has also been shown that abnormal encoding of acoustic transients in the brainstem is related to abnormal auditory responses measured at the level of cortex. In addition to their imprecise representation of sounds at the auditory brainstem, a significant proportion of LDs have also consistently demonstrated abnormal representations of simple (Menell et al., 1999; Ahissar et al.,

2000) and complex (Kraus et al., 1996; Bradlow et al., 1999; Ahissar et al., 2001; Wible et al., 2002; Banai et al., 2005; Wible et al., 2005) acoustic stimuli at the level of the auditory cortex. Three recent studies linked abnormal neural synchrony for acoustic transients at the auditory brainstem to abnormal representations of sounds in the cortex. In one study (Wible et al., 2005), it was shown that a brainstem measure of the encoding of acoustic transients, the duration of time between onset peaks V and A, was positively correlated to auditory cortex's susceptibility to background noise in both normal and LD subjects. Specifically, the longer the duration between onset peaks V and A, the more degraded cortical responses became in the presence of background noise. In another study, it was shown that individuals with abnormal brainstem timing to acoustic transients were more likely to indicate reduced cortical sensitivity to acoustic change, as measured by the mismatch negativity response (MMN) (Banai et al., 2005). Finally, a third study showed that brainstem timing for speech sound onset and offset predicts the degree of cortical asymmetry for speech sounds measured across a group of children with a wide range of reading skills (Abrams et al., in press). Results from these studies indicate that abnormal encoding of acoustic onsets at the brainstem may be a critical marker for systemic auditory deficits manifested at multiple levels of the auditory system, including the cortex.

In summary, evidence from examining the auditory brainstem response indicates that acoustic transients are encoded in a relatively simple fashion in the brainstem, yet they represent a complex phenomenon that is related to linguistic ability and cortical function. In the cortex, results indicate that spectral contrasts of speech onsets are mapped along the anterior-posterior axis in the auditory cortex, while temporal attributes of speech onsets, as manifested by the VOT, are precisely encoded with double peaked N100 responses.

The speech envelope

Definition and role in the perception of speech

The speech envelope refers to the temporal fluctuations in the speech signal between 2-50 Hz. The dominant frequency of the speech envelope is at ~4 Hz, which reflects the average syllabic rate of speech (Steeneken and Houtgast, 1980). Envelope frequencies in normal speech are generally below 8 Hz (Houtgast and Steeneken, 1985), and the perceptually essential frequencies of the speech envelope are between 4-16 Hz (Drullman et al., 1994; van der Horst et al., 1999), although frequencies above 16 Hz contribute slightly to speech recognition (Shannon et al., 1995). The speech envelope provides phonetic and prosodic cues to the duration of speech segments, manner of articulation, the presence (or absence) of voicing, syllabication, and stress (van der Horst et al., 1999). The perceptual significance of the speech envelope has been investigated using a number of methodologies (Drullman et al., 1994; Shannon et al., 1995; Smith et al., 2002b) and, taken together, these data indicate that the speech envelope is both necessary and sufficient for normal speech recognition.

Physiologic representation of the speech envelope in auditory cortex

Only a few studies have investigated how the human brain represents the slow temporal information of the speech envelope. It should be noted that the representation of the speech envelope in humans has only been studied at the level of the cortex, since measuring ABRs typically involves filtering out the neurophysiologic responses below ~100 Hz (Hall, 1992). Since speech envelope frequencies are between 2-50 Hz, any linear representation of the speech envelope in brainstem responses is removed with brainstem filtering.

In one MEG study, responses from the auditory cortex to natural and time-compressed (i.e. rapid) speech sentences were measured while subjects listened for semantic incongruities in the experimental sentences (Ahissar et al., 2001). Results indicate that human cortex synchronizes its response to the contours of the speech envelope, a phenomenon known as "phase-locking," and mimics the frequency content of the speech envelope, which they called "frequency matching." Moreover, it was shown that these two neurophysiologic measures correlate with subjects' ability to perceive the speech sentences: as speech sentences become more difficult to perceive due to increased time compression, the ability of the cortex to phase-lock and frequency match is more impaired. These results are in concert with results from the animal literature, which show that cortical neurons of primary auditory cortex represent the temporal envelope of complex acoustic stimuli (i.e., animal communication calls) by phase-locking to this temporal feature of the stimulus (Wang et al., 1995; Gehr et al., 2000; Nagarajan et al., 2002).

A second line of inquiry into the cortical representation of speech envelope cues was described previously in this chapter in the discussion of cortical responses to voice onset time (VOT) (Sharma and Dorman, 1999; Sharma and Dorman, 2000; Sharma et al., 2000). Acoustically, VOT is a slow temporal cue in speech (40-60 ms; 17-25 Hz.) that falls within the range of speech envelope frequencies. Briefly, neurophysiologic results indicated that for both /ba/-/pa/ and /ga/-/ka/ phonetic contrasts, cortical N100 responses precisely represented the acoustic attributes of the VOT. In addition, it was shown that neural responses were independent of the categorical perception of these phonetic contrasts (see the Acoustic Onsets section for a more detailed description of this study).

On the surface it may appear that the findings from these experiments contradict one another since cortical phase-locking to the speech envelope correlates with perception in one study (Ahissar et al., 2001) while phase-locking fails to correlate with perception in the other study (Sharma and Dorman, 1999; Sharma and Dorman, 2000; Sharma et al., 2000). These data are not, however, in contradiction to one another. In both cases, an a priori requirement for perception is phase-locking to the speech envelope; there is no evidence for perception in the absence of accurate phase-locking to the temporal envelope in either study. The primary difference between the studies is that despite phase-locking to the temporal envelope in the /ka/ stimulus condition at a VOT of ~40 msec, reliable perception of /ka/ occurs at ~ 60 msec. This suggests that accurate phase-locking is required for perception, however perception cannot be predicted by phase-locking alone. Presumably, in the case of the /ka/ VOT stimulus, there is another processing stage that uses the phase-locked temporal information in conjunction with additional auditory-linguistic information (e.g., repeated exposure to /ka/ stimuli with 60 msec VOT) as a means to form phonetic category boundaries. The question of if and how category boundaries are established irrespective of auditory phase-locking requires additional investigation.

Conclusions

Speech is a highly complex signal composed of a variety of acoustic features, all of which are important for normal speech perception. Normal perception of these acoustic features certainly relies on their neural encoding, which has been the subject of this review. An obvious conclusion from these studies is that the central auditory system is a remarkable machine, able to

simultaneously process the multiple acoustic cues of ongoing speech in order to decode a linguistic message. Furthermore, how the human brain is innately and dynamically programmed to utilize any number of these acoustic cues for the purpose of language, given the appropriate degree and type of stimulus exposure, further underscores the magnificence of this system.

A limitation of this chapter is that it has adopted a largely "bottom-up" approach to acoustic encoding of speech sounds: neural encoding of acoustic signals is generally discussed as an afferent phenomenon with minimal consideration for the dynamic interactions provided by top-down connections in the auditory system (Xiao and Suga, 2002; Perrot et al., 2005). A notable exception to this includes work by Krishnan (Krishnan et al., 2004), described in the frequency modulation section in which the role of language experience was shown to affect sensory encoding in the auditory brainstem. Another limitation to this chapter is that it has also ignored the influence of other systems of the central nervous system, such as cognitive and emotional effects on auditory processing of speech, which most certainly have a role in shaping auditory activity.

To garner a greater understanding of how the central auditory system processes speech, it is important to consider both sub-cortical and cortical auditory regions. Across the acoustic features described in this review, the brainstem appears to represent acoustic events in a relatively linear fashion: the fundamental frequency and its modulation are represented with highly synchronized activity as reflected by the FFR; speech sound onset is represented with highly predictable neural activation patterns that vary within fractions of milliseconds. Alternatively, the cortex appears to transform many of these acoustic cues, resulting in more

complex representations of acoustic features of speech. For example, many of the cortical findings described here are based on the spatial representation of acoustic features (i.e., the relationship between F1-F2 required for vowel identification; the differentiation of speech transients; the encoding of periodicity). Because cortical neurons are not able to phase-lock to high frequency events, it is tempting to propose that cortex has found an alternative method for encoding these features based on the activity of spatially distributed neural populations. The extent to which these acoustic features are truly represented via a spatial organization in cortex is a future challenge that will be likely achieved using high resolution imaging technologies in concert with EEG and MEG technologies.

Table 1

Major Sections: Acoustic Features in Speech	Feature's Role in the Speech Signal	Brainstem measure	Cortical measure
1. Formant structure	Ubiquitous in vowels, approximants and nasals; essential for vowel perception.	Frequency Following Response	N100m source location; STS activity (fMRI)
2. Periodicity	Temporal cue for the fundamental frequency and low formant frequencies (50-500 Hz)	Frequency Following Response	N100m source location and amplitude; non-primary auditory cortex activity patterns (fMRI)
3. Frequency transitions	Consonant identification; signal the presence of diphthongs and glides; linguistic pitch	Frequency Following Response	Left vs. right STG activity (fMRI)
4. Acoustic onsets	Phoneme identification	ABR onset complex	N100m source location; N100 latency
5. Speech envelope	Syllable and low frequency (<50 Hz) patterns in speech	N/A	N100m phase-locking

Figure 1

Figure 1a

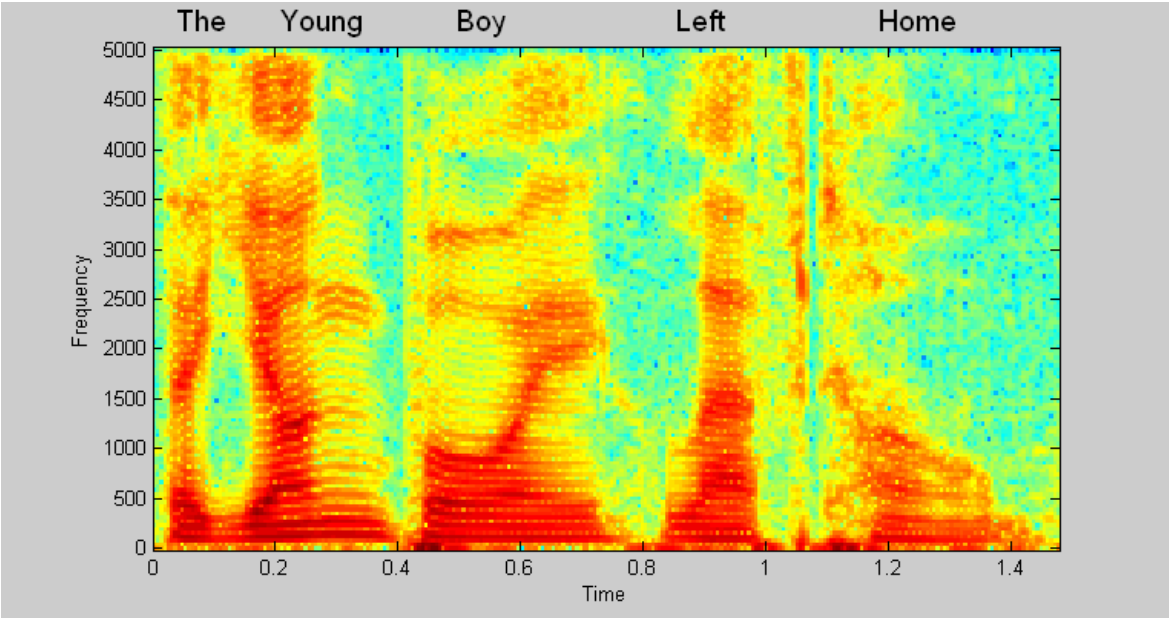


Figure 1b

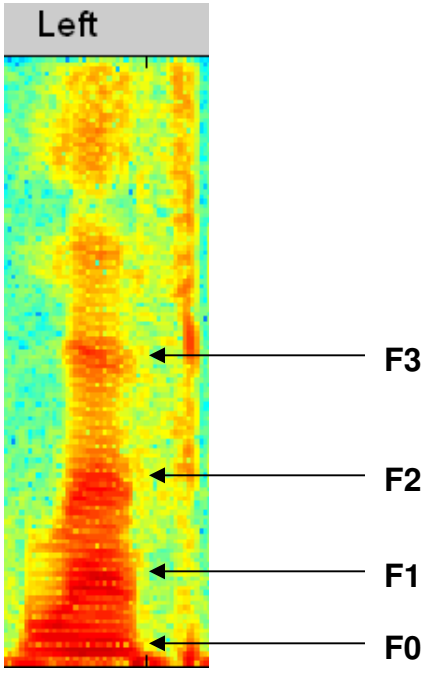


Figure 2

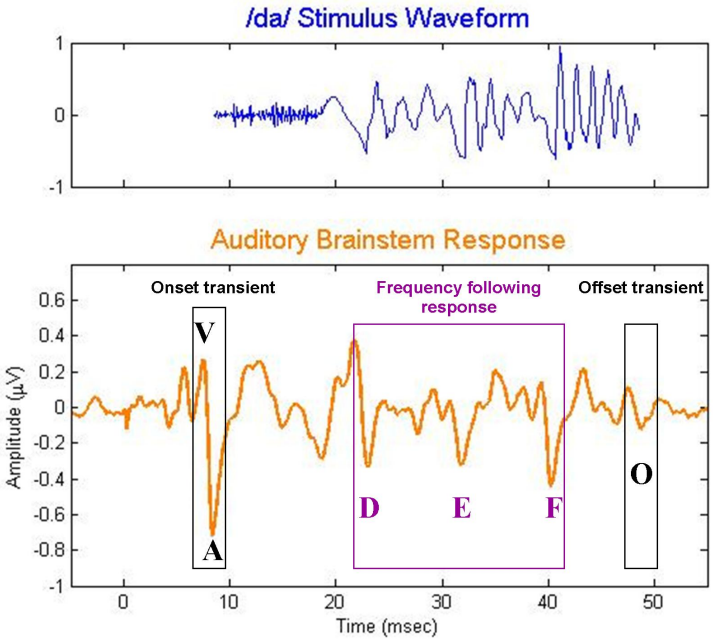

/da/ Stimulus Waveform
1
0
-1
Auditory Brainstem Response
Onset transient
Frequency following response
Offset transient
V
A
D
E
F
O
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
Amplitude (μV)
0
10
20
30
40
50
Time (msec)

Figure Captions:

Figure 1: Spectrogram for the naturally produced speech sentence "The young boy left home." (A) the complete sentence; (B) the word "left" is enlarged to illustrate the frequency structure: the fundamental frequency (F0) and formants (F1-F3) are represented in the spectrogram by broad red lines of energy.

Figure 2: Acoustic waveform of the synthesized speech stimulus /da/ (above) and grand average auditory brainstem responses to /da/ (below). The stimulus has been moved forward in time to the latency of onset responses (peak V) to enable direct comparisons with brainstem responses. Peaks V and A reflect the onset of the speech sound and peak O reflects stimulus offset. Peaks D, E and F represent a phase-locked representation to the fundamental frequency of the speech stimulus, and the peaks between D, E and F occur at the F1 frequency.

References:

Abrams DA, Nicol T, Zecker S, Kraus N (in press) Auditory brainstem timing predicts cerebral asymmetry for speech. J Neurosci.

Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM (2001) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc Natl Acad Sci U S A 98:13367-13372.

Ahissar M, Hochstein S (2004) The reverse hierarchy theory of visual perceptual learning. Trends Cogn Sci 8:457-464.

Ahissar M, Protopapas A, Reid M, Merzenich MM (2000) Auditory processing parallels reading abilities in adults. Proc Natl Acad Sci U S A 97:6832-6837.

Banai K, Nicol T, Zecker SG, Kraus N (2005) Brainstem timing: implications for cortical processing and literacy. J Neurosci 25:9850-9857.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. Nature 403:309-312.

Belin P, Zilbovicius M, Crozier S, Thivard L, Fontaine A, Masure MC, Samson Y (1998) Lateralization of speech and auditory temporal processing. J Cogn Neurosci 10:536-540.

Benasich AA, Tallal P (2002) Infant discrimination of rapid auditory cues predicts later language impairment. Behav Brain Res 136:31-49.

Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T (1997) Human brain language areas identified by functional magnetic resonance imaging. J Neurosci 17:353-362.

Bishop DV, Bishop SJ, Bright P, James C, Delaney T, Tallal P (1999) Different origin of auditory and phonological processing problems in children with language impairment: evidence from a twin study. J Speech Lang Hear Res 42:155-168.

Bradlow AR, Kraus N, Hayes E (2003) Speaking clearly for children with learning disabilities: sentence perception in noise. J Speech Lang Hear Res 46:80-97.

Bradlow AR, Kraus N, Nicol TG, McGee TJ, Cunningham J, Zecker SG, Carrell TD (1999) Effects of lengthened formant transition duration on discrimination and neural representation of synthetic CV syllables by normal and learning-disabled children. J Acoust Soc Am 106:2086-2096.

Creutzfeldt O, Hellweg FC, Schreiner C (1980) Thalamocortical transformation of responses to complex auditory stimuli. Exp Brain Res 39:87-104.

Cunningham J, Nicol T, Zecker S, Kraus N (2000) Speech-evoked neurophysiologic responses in children with learning problems: development and behavioral correlates of perception. Ear Hear 21:554-568.

Cunningham J, Nicol T, Zecker S, Bradlow A, Kraus N (2001) Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. Clinical Neurophysiology 112:758-767.

Diesch E, Luce T (1997) Magnetic fields elicited by tones and vowel formants reveal tonotopy and nonlinear summation of cortical activation. Psychophysiology 34:501-510.

Drullman R, Festen JM, Plomp R (1994) Effect of temporal envelope smearing on speech reception. J Acoust Soc Am 95:1053-1064.

Eggermont JJ (1991) Rate and synchronization measures of periodicity coding in cat primary auditory cortex. Hear Res 56:153-167.

Galbraith GC, Threadgill MR, Hemsley J, Salour K, Songdej N, Ton J, Cheung L (2000) Putative measure of peripheral and brainstem frequency-following in humans. Neurosci Lett 292:123-127.

Gandour J, Wong D, Hutchins G (1998) Pitch processing in the human brain is influenced by language experience. Neuroreport 9:2115-2119.

Gardi J, Merzenich M, McKean C (1979) Origins of the scalp recorded frequency-following response in the cat. Audiology 18:358-381.

Gehr DD, Komiya H, Eggermont JJ (2000) Neuronal responses in cat primary auditory cortex to natural and altered species-specific calls. Hear Res 150:27-42.

Geschwind N, Galaburda AM (1985) Cerebral lateralization. Biological mechanisms, associations, and pathology: I. A hypothesis and a program for research. Arch Neurol 42:428-459.

Hall JH (1992) Handbook of auditory evoked responses. Boston: Allyn and Bacon.

Hari R, Kiesila P (1996) Deficit of temporal auditory processing in dyslexic adults. Neurosci Lett 205:138-140.

Hari R, Saaskilahti A, Helenius P, Uutela K (1999) Non-impaired auditory phase locking in dyslexic adults. Neuroreport 10:2347-2348.

Hayes EA, Warrier CM, Nicol TG, Zecker SG, Kraus N (2003) Neural plasticity following auditory training in children with learning problems. Clin Neurophysiol 114:673-684.

Hickok G, Poeppel D (2004) Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92:67-99.

Houtgast T, Steeneken HJM (1985) A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. The Journal of the Acoustical Society of America 77:1069-1077.

Jezzard P, Matthews PM, Smith SM (2001) Functional MRI: an introduction to methods. Oxford: Oxford University Press.

Johnson K (1997) Acoustic and auditory phonetics. Cambridge: Blackwell Publishers Inc.

Johnson KL, Nicol TG, Kraus N (2005) Brain Stem Response to Speech: A Biological Marker of Auditory Processing. Ear Hear 26:424-434.

King C, Warrier CM, Hayes E, Kraus N (2002) Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. Neurosci Lett 319:111-115.

Klein D, Zatorre RJ, Milner B, Zhao V (2001) A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. Neuroimage 13:646-653.

Klingberg T, Hedehus M, Temple E, Salz T, Gabrieli JD, Moseley ME, Poldrack RA (2000) Microstructure of temporo-parietal white matter as a basis for reading ability: evidence from diffusion tensor magnetic resonance imaging. Neuron 25:493-500.

Kraus N, Nicol T (2005) Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. Trends Neurosci 28:176-181.

Kraus N, McGee TJ, Carrell TD, Zecker SG, Nicol TG, Koch DB (1996) Auditory neurophysiologic responses and discrimination deficits in children with learning problems. Science 273:971-973.

Krishnan A (2002) Human frequency-following responses: representation of steady-state synthetic vowels. Hear Res 166:192-201.

Krishnan A, Xu Y, Gandour JT, Cariani PA (2004) Human frequency-following response: representation of pitch contours in Chinese tones. Hear Res 189:1-12.

Krishnan A, Xu Y, Gandour J, Cariani P (2005) Encoding of pitch in the human brainstem is sensitive to language experience. Brain Res Cogn Brain Res 25:161-168.

Lehiste I, Peterson GE (1961) Transitions, Glides, and Diphthongs. The Journal of the Acoustical Society of America 33:268-277.

Lu T, Liang L, Wang X (2001) Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nat Neurosci 4:1131-1138.

Makela AM, Alku P, Makinen V, Valtonen J, May P, Tiitinen H (2002) Human cortical dynamics determined by speech fundamental frequency. Neuroimage 17:1300-1305.

McGee T, Kraus N, King C, Nicol T, Carrell TD (1996) Acoustic elements of speechlike stimuli are reflected in surface recorded responses over the guinea pig temporal lobe. J Acoust Soc Am 99:3606-3614.

Menell P, McAnally KI, Stein JF (1999) Psychophysical sensitivity and physiological response to amplitude modulation in adult dyslexic listeners. J Speech Lang Hear Res 42:797-803.

Mody M, Studdert-Kennedy M, Brady S (1997) Speech perception deficits in poor readers: auditory processing or phonological coding? J Exp Child Psychol 64:199-231.

Naatanen R, Picton T (1987) The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 24:375-425.

Nagarajan S, Mahncke H, Salz T, Tallal P, Roberts T, Merzenich MM (1999) Cortical auditory signal processing in poor readers. Proc Natl Acad Sci U S A 96:6483-6488.

Nagarajan SS, Cheung SW, Bedenbaugh P, Beitel RE, Schreiner CE, Merzenich MM (2002) Representation of spectral and temporal envelope of twitter vocalizations in common marmoset primary auditory cortex. J Neurophysiol 87:1723-1737.

Nearey TM, Assmann PF (1986) Modeling the role of inherent spectral change in vowel identification. J Acoust Soc Am 80:1297-1308.

Nusbaum HC, Morin TM (1992) Paying attention to differences among talkers. In: Speech Perception, Production, and Linguistic Structure (Tohkura Y, Sagisaka Y, Vatikiotis-Bateson E, eds), pp 113-134. Tokyo: Ohmasha Publishing.

Obleser J, Scott SK, Eulitz C (2005) Now You Hear It, Now You Don't: Transient Traces of Consonants and their Nonspeech Analogues in the Human Brain. Cereb Cortex.

Obleser J, Elbert T, Lahiri A, Eulitz C (2003) Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. Brain Res Cogn Brain Res 15:207-213.

Perrot X, Ryvlin P, Isnard J, Guenot M, Catenoix H, Fischer C, Mauguiere F, Collet L (2005) Evidence for Corticofugal Modulation of Peripheral Auditory Activity in Humans. Cereb Cortex:bhj035.

Peterson GE, Barney HL (1952) Control Methods Used in a Study of the Vowels. The Journal of the Acoustical Society of America 24:175-184.

Ramus F, Rosen S, Dakin SC, Day BL, Castellote JM, White S, Frith U (2003) Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. Brain 126:841-865.

Rauschecker JP (1997) Processing of complex sounds in the auditory cortex of cat, monkey, and man. Acta Otolaryngol Suppl 532:34-38.

Reed MA (1989) Speech perception and the discrimination of brief auditory cues in reading disabled children. J Exp Child Psychol 48:270-292.

Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci 336:367-373.

Russo N, Nicol T, Musacchia G, Kraus N (2004) Brainstem responses to speech syllables. Clinical Neurophysiology 115:2021-2030.

Russo NM, Nicol TG, Zecker SG, Hayes EA, Kraus N (2005) Auditory training improves neural timing in the human brainstem. Behav Brain Res 156:95-103.

Sachs MB, Young ED (1979) Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate. J Acoust Soc Am 66:470-479.

Sachs MB, Voigt HF, Young ED (1983) Auditory nerve representation of vowels in background noise. J Neurophysiol 50:27-45.

Sato S (1990) Magnetoencephalography New York: Raven Press.

Schreiner CE (1998) Spatial distribution of responses to simple and complex sounds in the primary auditory cortex. Audiol Neurootol 3:104-122.

Schulte-Korne G, Deimel W, Bartling J, Remschmidt H (1998) Auditory processing and dyslexia: evidence for a specific speech processing deficit. Neuroreport 9:337-340.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303-304.

Sharma A, Dorman MF (1999) Cortical auditory evoked potential correlates of categorical perception of voice-onset time. J Acoust Soc Am 106:1078-1083.

Sharma A, Dorman M (2000) Neurophysiologic correlates of cross-language phonetic perception. J Acoust Soc Am 107:2697-2703.

Sharma A, Marsh C, Dorman M (2000) Relationship between N1 evoked potential morphology and the perception of voicing. J Acoust Soc Am 108:3030-3035.

Shestakova A, Brattico E, Soloviev A, Klucharev V, Huotilainen M (2004) Orderly cortical representation of vowel categories presented by multiple exemplars. Brain Res Cogn Brain Res 21:342-350.

Smith AJ, Blumenfeld H, Behar KL, Rothman DL, Shulman RG, Hyder F (2002a) Cerebral energetics and spiking frequency: the neurophysiological basis of fMRI. Proc Natl Acad Sci U S A 99:10765-10770.

Smith JC, Marsh JT, Brown WS (1975) Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. Electroencephalogr Clin Neurophysiol 39:465-472.

Smith ZM, Delgutte B, Oxenham AJ (2002b) Chimaeric sounds reveal dichotomies in auditory perception. Nature 416:87-90.

Song JH, Banai K, Russo NM, Kraus N (2006) On the relationship between speech- and nonspeech-evoked auditory brainstem responses. Audiol Neurootol 11:233-241.

Steeneken HJ, Houtgast T (1980) A physical method for measuring speech-transmission quality. J Acoust Soc Am 67:318-326.

Steinschneider M, Reser DH, Fishman YI, Schroeder CE, Arezzo JC (1998) Click train encoding in primary auditory cortex of the awake monkey: evidence for two mechanisms subserving pitch perception. J Acoust Soc Am 104:2935-2955.

Steinschneider M, Volkov IO, Noh MD, Garell PC, Howard MA, 3rd (1999) Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. J Neurophysiol 82:2346-2357.

Stillman RD, Crow G, Moushegian G (1978) Components of the frequency-following potential in man. Electroencephalogr Clin Neurophysiol 44:438-446.

Tallal P, Piercy M (1973) Defects of non-verbal auditory perception in children with developmental aphasia. Nature 241:468-469.

Tallal P, Piercy M (1975) Developmental aphasia: the perception of brief vowels and extended stop consonants. Neuropsychologia 13:69-74.

Tallal P, Miller S, Fitch RH (1993) Neurobiological basis of speech: a case for the preeminence of temporal processing. Ann N Y Acad Sci 682:27-47.

Temple E, Poldrack RA, Protopapas A, Nagarajan S, Salz T, Tallal P, Merzenich MM, Gabrieli JD (2000) Disruption of the neural response to rapid acoustic stimuli in dyslexia: evidence from functional MRI. Proc Natl Acad Sci U S A 97:13907-13912.

van der Horst R, Leeuw AR, Dreschler WA (1999) Importance of temporal-envelope cues in consonant recognition. J Acoust Soc Am 105:1801-1809.

Wang X, Merzenich MM, Beitel R, Schreiner CE (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. J Neurophysiol 74:2685-2706.

Wang Y, Jongman A, Sereno JA (2001) Dichotic perception of Mandarin tones by Chinese and American listeners. Brain Lang 78:332-348.

Wible B, Nicol T, Kraus N (2002) Abnormal neural encoding of repeated speech stimuli in noise in children with learning problems. Clin Neurophysiol 113:485-494.

Wible B, Nicol T, Kraus N (2004) Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. Biol Psychol 67:299-317.

Wible B, Nicol T, Kraus N (2005) Correlation between brainstem and cortical auditory processes in normal and language-impaired children. Brain 128:417-423.

Witton C, Stein JF, Stoodley CJ, Rosner BS, Talcott JB (2002) Separate Influences of Acoustic AM and FM Sensitivity on the Phonological Decoding Skills of Impaired and Normal Readers. J Cogn Neurosci 14:866-874.

Wong PC, Parsons LM, Martinez M, Diehl RL (2004) The role of the insular cortex in pitch pattern perception: the effect of linguistic contexts. J Neurosci 24:9153-9160.

Wright BA, Lombardino LJ, King WM, Puranik CS, Leonard CM, Merzenich MM (1997) Deficits in auditory temporal and spectral resolution in language-impaired children. Nature 387:176-178.

Xiao Z, Suga N (2002) Modulation of cochlear hair cells by the auditory cortex in the mustached bat. Nat Neurosci 5:57-63.

Zatorre RJ, Belin P (2001) Spectral and temporal processing in human auditory cortex. Cereb Cortex 11:946-953.

Zatorre RJ, Belin P, Penhune VB (2002) Structure and function of auditory cortex: music and speech. Trends Cogn Sci 6:37-46.

Acknowledgements:

We would like to thank Karen Banai and Trent Nicol for their comments on a previous draft of this chapter. This work is supported by the National Institutes of Health grant R01 DC01510-10 and National Organization for Hearing Research grant 340-B208.

SENSORY-BASED LEARNING DISABILITY: INSIGHT FROM BRAINSTEM PROCESSING OF SPEECH SOUNDS

Published in *International Journal of Audiology*, 46:9, 524 - 532

Karen Banai[1,3], Daniel Abrams[1] and Nina Kraus[1-3]

[1]Auditory Neuroscience Laboratory; Departments of Communication Sciences, [2]Neurobiology and Physiology, Otolaryngology and [3]Northwestern Institute of Neuroscience, Northwestern University, Evanston, IL USA.

Abstract

Speech-evoked auditory brainstem responses (speech-ABR) provide a reliable marker of learning disability in a substantial subgroup of individuals with language-based learning problems (LDs). Here we review work describing the properties of the speech-ABR in typically developing children and in children with LD. We also review studies on the relationships between speech-ABR and the commonly used click-ABR and between speech-ABR and auditory processing at the level of the cortex. In a critical examination of previously published data, we conclude that as many as 40% of LDs have abnormal speech-ABRs and that these individuals are also likely to exhibit abnormal cortical processing. Yet, the profile of learning problems these individuals exhibit is unspecific. Leaving open the question of causality, these data suggest that speech-ABR can be used to identify a large sub-population of LDs, those with abnormal auditory physiological function. Further studies are required to determine the functional relationships

among abnormal speech-ABR, speech perception, and the pattern of literacy-related and cognitive deficits in LD.

Key words: AEP, ABR, auditory processing, dyslexia, learning disability, speech encoding.

Introduction

The auditory system is extremely sensitive to the temporal characteristics of sound (see Frisina, 2001; Oertel, 1997 for reviews) and auditory evoked potentials (AEPs) are commonly used to characterize these temporal properties in a non-invasive fashion. Furthermore, AEPs have long been recognized as a reliable vehicle for providing objective information about the structural and functional integrity of the central auditory system (Hall, 1992; Kraus and McGee, 1992). AEPs provide an important tool not only in auditory neuroscience laboratories but also in the audiologist's clinic (Hood, 1998) and the operating room (Martin and Mishler, 2002).

Brief and rapid acoustic events (i.e clicks) result in a synchronized pattern of neural activity in nuclei along the auditory brainstem. When recorded from the scalp, this activity results in a series of voltage fluctuations known as the click-ABR. This response provides information about brainstem nuclei along the ascending auditory pathway (Hood, 1998; Jacobsen, 1985; Møller, 1999). Fractions of a millisecond deviations from the normal pattern are clinically important in the diagnosis of hearing loss (Hood, 1998) and pathologies such as brainstem tumors (Musiek and Gollegly, 1985) and multiple sclerosis (Keith and Jacobson, 1985).

Known temporal properties of brainstem neurons, which can phase lock up to ~1000 Hz, as well as the remarkable temporal precision of the scalp recorded response they evoke implies that the brainstem is likely to also faithfully encode many of the acoustic properties of speech and other complex auditory signals. Evidence that AEPs may be used to study various aspects of this complex speech/acoustical encoding in humans has been obtained in several laboratories (Galbraith et al., 1995; Krishnan, 2002; Russo et al., 2004). Here we review work on the normal subcortical encoding of one of the building blocks of speech – consonant vowel (CV) syllables and the disruption of this normal process in the learning impaired population.

The speech evoked brainstem response

Speech is a complex signal whose acoustic properties change continuously over time and whose processing extends from the cochlea to the cortex. Work in animal models has shown that neurons in the auditory nerve and the cochlear nucleus are sensitive to various properties of speech-like stimuli such as formant structure (Delgutte, 1980; Delgutte and Kiang, 1984a), formant transitions (Delgutte and Kiang, 1984b) and voice onset time (Clarey et al., 2004). Relatively little is known about the encoding of speech or speech-like stimuli in higher areas of the brainstem, where the majority of animal studies focused on simpler stimuli such as amplitude modulated noise bursts to study coding properties at both the single cell and multi-unit levels (e.g. Langner and Schreiner, 1988; Schreiner and Langner, 1988).

Nonetheless, clinical evidence indicates that higher brainstem nuclei such as the inferior colliculus (IC) play an important role in auditory processing in humans (Johkura et al., 1998; Musiek et al., 2004). For example, Johkura et al. (1998) report the case of a patient with bilateral

IC lesions who showed symptoms of auditory agnosia in the absence of a cortical temporal lobe lesion. Indeed, the response generators of both the late waves of the ABR (V and Vn here called A) and the FFR (frequency following response) have been localized to the upper brainstem (lateral lemniscus, IC), (Boston and Møller, 1985; Møller, 1999). Corroborating evidence from animal models supports the idea that these regions of the brainstem are sensitive to complex spectral and temporal properties of complex stimuli (Eggermont and Ponton, 2002; Irvine, 1992; Langner and Schreiner, 1988; Schreiner and Langner, 1988; Sinex and Chen, 2000) and are therefore likely to have a role in speech processing in humans.

Encoding of speech and speech-like signals at the level of the brainstem (lateral lemniscus, IC) has been studied in humans using AEPs (Galbraith et al., 1995; Krishnan, 2002; Plyler and Ananthanarayan, 2001; Russo et al., 2004). In particular, studies focusing on the FFR demonstrated its role in encoding speech and speech-like sounds (Galbraith et al., 2004; Galbraith et al., 1995; Krishnan, 2002; Krishnan et al., 2004). Understanding how complex acoustic stimuli are encoded in the brainstem, and how this processing is related to processes taking place in lower (e.g. the auditory nerve) and higher (e.g. the auditory cortex) areas of the auditory pathway, should lead to a better understanding of processes underlying normal and abnormal human communication.

Description of the normal speech-ABR

Brainstem responses elicited by speech stimuli can provide clues about encoding of the sound structure of speech syllables by the CNS. In recent years it has been demonstrated that the neural code indeed reflects specific features of the acoustic signal (e.g. formants, VOT). Thus, the

morphology of the brainstem response elicited by a speech syllable can be described in terms similar to those used to describe the physical stimulus itself. As shown in Figure 1, the brainstem response can be divided into two components: an onset response and the frequency following response (FFR).

Together, the onset and the FFR components of the speech-ABR roughly reflect the acoustic parameters of the CV stimulus used to evoke the response. The onset component arises as a response to the onset of sound. In the case of a CV stimulus the onset represents the initiation of the consonant and contains aperiodic information. Its initial waves are similar to those observed in response to click stimuli (waves I, III and the VA complex) whereas wave C possibly reflects the onset of voicing. The FFR reflects phase locking to the fundamental frequency of the stimulus. It arises in response to the periodic information present in the vowel at the frequency of the sound source (i.e. the glottal pulse). Thus the period between peaks D, E and F of the FFR corresponds to the fundamental frequency of the stimulus (F0), whereas the peaks between waves D, E and F represent phase locking at the frequencies of the first formant (F1). The parallels between the morphology of the syllable /da/ and the ABR it evokes have been recently reviewed in detail by Johnson, Nicol and Kraus (2005) and by Russo et al. (2004).

In the following paragraphs the characteristics of the speech-ABR evoked by the syllable /da/ will be described in some detail. The Kraus laboratory has been studying this response intensively in both typically developing children and children with learning problems 8-12 years of age. We will first describe the characteristics of the normal response and then examine the abnormal response as measured in a large group of children with language-based learning

problems (LD). The stimulus and recording parameters have been described in detail in previous publications (Johnson et al., 2005; Russo et al., 2004; Song et al., 2006).

The speech-evoked brainstem response is a complex pattern of voltage fluctuations. As can be seen in Figure 1, the physiological response includes an orderly series of peaks and troughs. In analyzing the response both timing (peak latencies) and magnitude measures (peak amplitudes, RMS) are used. The first positive peaks (labeled I and III in Figure 1) are similar to waves I and III generated by click stimulation and likely originate at low levels of the auditory system (the 8th nerve and the low brainstem respectively) (Boston and Møller, 1985). Similarly, wave V represents the onset of the speech stimulus at the upper brainstem, followed by a large negative deflection (wave A). Characteristic latency and amplitude values are shown in Table 1 (top part). The brainstem response evoked by the /da/ syllable is reliable at the individual level. First, between different individuals the speech-evoked brainstem response is consistent, with the same morphological and spectral features identifiable for the large majority of individuals (see Table 1 for mean and standard deviation values in the normal population). Second, within an individual the evoked responses measured on different occasions are highly replicable (Russo et al., 2004).

In addition to the peak latency/amplitude analysis, yielding information about transient events within the response, sustained aspects of the response can be analyzed as well. A sustained magnitude measure describing total response energy over different time windows is the RMS. Additional sustained measures can be obtained using an analysis in the frequency domain (Fast Fourier Transform, FFT), providing information about the presence of specific frequencies in the response. As shown in Figure 2, an FFT over the periodic portion of the response (23-44 ms) reveals that the bulk of physiological energy is distributed in frequency ranges roughly

corresponding to the F0 and F1 formants of the /da/ syllable. Characteristic magnitude values are shown in table 1 (bottom). While the formants are prominent in the /da/ signal, by definition formant frequencies always correspond to harmonics of the fundamental. Thus while the spectral peaks observed in Figure 2 around 220 Hz and 450 Hz are larger than those roughly corresponding to the other harmonics it could still be claimed that the response is encoding, at least in part, the harmonics and not F1. Also it should be noted that because both F0 and F1 change over time, the FFT, which is calculated over time, provides only an approximation for the spectral shape of the response in any given point in time.

In the normal population, significant correlations exist between the latencies of the onset measures, but not between the latencies of the onset and the FFR waves (see Russo et al., 2004 for details). Russo et al. (2004) have further found that significant correlations also exist between the latencies of the onset measures and the spectral magnitude of F1, indicating a relationship between precision of temporal and spectral aspects of the response. It has been suggested that the pattern of correlation between the onset peaks, and the lack of correlation between the onset and the FFR peaks reflects dissociation between these two classes of response – filter and source classes respectively, representing the building blocks of the message (i.e. the content) vs. talkers' identity (see Kraus and Nicol, 2005). Taken together with the pattern of brainstem abnormalities observed in children with LD (reviewed below), Kraus and Nicol (2005) proposed that the separate encoding of these response classes at the brainstem may be a precursor for the cortical 'what/where' pathways (Rauschecker and Tian, 2000; Romanski et al., 1999).

Finally, in the presence of background noise, brainstem encoding of speech is disrupted. In particular, noise interferes with the onset response. In the majority of normal subjects the onset

response is severely degraded while in 40% of subjects it is completely abolished. On the other hand, the FFR portion of the response is less susceptible to noise and the FFR peaks are identifiable in cases where the onset has disappeared (Russo et al., 2004).

Abnormal speech-ABR and learning disability

The focus here is on children with language-based learning problems. Previous work indicated that some children with LDs exhibit abnormal encoding of sound at the cortical level (see Heim and Keil, 2004; Lyytinen et al., 2005 for recent reviews), and our studies (Banai et al., 2005b; King et al., 2002; Wible et al., 2004, 2005) have also revealed abnormal encoding at the brainstem level. The first studies looking into brainstem encoding in the LD population compared learning disabled children to typically developing ones at the group level. Thus, Cunningham et al. (2001) found that wave V latency was delayed in a group of LDs in noise, but not in quiet. They further demonstrated that the magnitude of the spectral content of the response in children with LD during the FFR period was reduced in background noise, especially in the frequency range corresponding to F1. Using a slightly different version of the /da/ stimulus (described in Russo et al., 2004; Wible et al., 2004), subsequent studies found that LDs had delayed waves A, C and F (King et al., 2002), a less synchronized onset of the speech-ABR as measured by the VA complex (Wible et al., 2004) and, consistent with the findings of Cunningham et al. (2001), reduced spectral representation in the F1 range (Wible et al., 2004). Wible et al. (2004) also established that the slope of the VA complex (i.e. the inter-peak amplitude divided by the inter-peak duration) provides a useful metric to describe the abnormal response by capturing both the duration of the V-to-A transition and its amplitude in a single number and indeed this measure has been found useful in later studies with larger samples. These

findings were in contrast to the normal click-evoked ABRs typically reported in earlier studies in individuals with LD (see "The relationship between speech- and click-ABR" below).

A careful examination of the data however, reveals that in the group with LD responses are abnormal due to the contribution of a subgroup of the LD population, whereas many children with LD exhibit a normal response. For example, King et al. (2002) have observed for wave A that 20/54 children with LD had responses deviating from the mean normal response latency by 1 standard deviation or more. Thus the question arises – how to define and characterize the abnormal response at the individual level?

Definition of the abnormal speech-ABR

Like most biological signals the speech-ABR is a continuous response. Consequently each parameter of the response may span a wide range of values, even in the normal population. Determining whether an individual response is abnormal presents a challenge. From a purely scientific standpoint it may be advantageous to look at the entire response continuum, but from a clinical perspective it is important to be able to easily distinguish a normal from an abnormal response based on some classification rule. In order for the definition of an abnormal response to be meaningful, the criteria chosen should be sensitive to the clinical population at hand, but also have a low rate of false positives. Yet, it should be noted that every criterion chosen will be arbitrary to some extent.

Indeed, King et al. (2002) used a 1 SD criterion for wave A latency, whereas Banai et al. (2005b) used 1.5-2 SD over a wider range of response parameters that included all onset measures. Based

on these two criteria, 30-40% of LD responses were classified as abnormal, but each yielded a somewhat different grouping of LDs to those with 'normal' vs. 'abnormal' responses. In ongoing work, Abrams et al. (work in progress) are looking at different classification criteria. Our goal is to optimize grouping by using a criterion that will be sensitive to the presence of LD while keeping the false positive rate (i.e. the number of typically developing children whose responses are classified as abnormal) under 10%. This effort is geared towards a sub-group within the LD population and is not necessarily related to the debate surrounding causal role of abnormal speech-ABR in LD. Indeed, several scenarios can account for the presence of abnormal speech-ABR (and other auditory functions) in some but not all persons with LD. First, abnormal brainstem function could be a cause of LD in some individuals whereas LD is caused by other reasons in other individuals. Alternatively, abnormal brainstem function could be a risk factor that contributes to the learning problem only when present with other genetic and environmental risk factor (see Bishop, 2006). This scenario can explain both why there is a high incidence of auditory processing deficits among individuals with LD and why some people have abnormal auditory processing but do not develop LD.

In order for the speech-ABR to be clinically useful, it is important to establish its test re-test reliability not only in the general population, but also among individuals with learning problems. Ten of the children with LD identified by Banai et al. (2005b) as having abnormal speech-ABR were retested using a clinical system (BioMAPTM, see below). Since the number of sampling points differs between the BioMAPTM and the original laboratory measurement it was not possible to directly calculate a correlation score between the two measurements for each individual. However, using the norms collected for the clinical system and the same

classification criteria used in the original study, all 10 individuals were, again, classified as having an abnormal response indicating that abnormal responses identified in the first measurement were not of transient nature.

The relationship between click- and speech-ABR

Numerous studies in the LD population have shown that individuals with LD have normal click-ABRs (Grontved et al., 1988a, b; Jerger et al., 1987; Lauter and Wood, 1993; Mason and Mellor, 1984; McAnally and Stein, 1997; Purdy et al., 2002). A prerequisite to participation in our speech-ABR studies is a clinically normal click-evoked wave V, to rule out peripheral hearing loss as a cause of abnormal speech-ABRs. In the general population, the early waves of the speech-ABR are similar to the waves evoked by click stimuli. Furthermore, in the normal population, significant correlations exist between the latency of wave V evoked by a click and the latencies of waves V and A evoked by speech (Song et al., 2006), suggesting that processing of these two types of stimuli is (at least to some extent) shared. This pattern of correlation is maintained among children with LD. On the other hand, this normal pattern of correlation is disrupted when speech-ABR is delayed, such that in children with abnormal onset of speech-ABR the correlation between the latencies of the speech- and click- evoked measures is significantly reduced. These findings indicate that these two processes do not always overlap.

Song et al. (2006) further noted that among children with abnormal speech-ABR, click-ABR latencies were delayed compared to children with normal speech-ABR, even if latencies were still within the normal range. This provides further support for the notion that speech and click stimuli are not independently encoded, even if deficits can not be observed using common clinical procedures.

Song et al. (2006) suggested that while the encoding of speech and click stimuli shares some common characteristics, the ABRs they evoke differ based on the acoustic characteristics of the evoking stimuli. Thus, the acoustic characteristics of the speech syllable /da/ used to measure the speech-ABR may be more challenging to the auditory system of persons with LD since the periodic portion of the vowel may mask the abrupt onset of the consonant (backward masking). This idea received support in a recent study by Marler and Champlin (2005) demonstrating, in a group of children with language disorder, a significant delay in wave V latency under backward masking conditions. Alternatively, the slower rise time of the speech-stimulus compared to the click could potentially enhance the effects of neural desynchronization in the population with LD. The findings of slightly delayed (yet within normal) click-ABR is consistent with recent findings in an animal model. Strata et al. (2005) have shown that experimentally induced perinatal anoxia in rats results in progressively delayed auditory processing from the brainstem to the auditory cortex. Taken together with our own findings regarding the relationships between click- and speech- ABR, these findings raise the possibility that abnormal speech-ABR may be a manifestation of a broader 'problem' in the central auditory system not detected by pure tone audiometry or supra-threshold click-ABR. This hypothesis should be tested in further studies.

Early waves of the speech-ABR

Song et al. (manuscript in preparation) are looking at the early waves (I, III) of the speech-ABR, aiming to characterize those waves, similarly to the work of Russo et al., (2004) for the later waves (wave V and later). Preliminary findings indicate the timing of the early waves appears normal in the majority of individuals with abnormal late waves (V and A) suggesting that for the

most part, the origins of the speech encoding deficits documented using the speech-ABR are retrocochlear.

The relationship between brainstem and cortical processing

Deficient brainstem timing has been linked to several manifestations of abnormal cortical processing. First, in the normal population, the robustness of cortical speech-encoding in noise is correlated with brainstem timing. Wible et al. (2005) showed that a strong correlation exists between brainstem timing and the effects of background noise on the cortical response, placing children with LD and delayed brainstem timing on the opposite end of this continuum with respect to those with normal timing and normal learning children Second, abnormal brainstem timing is associated with reduced cortical discrimination of fine acoustic differences (MMNs). Thus, as a group, individuals with delayed brainstem timing do not show a significant MMN response to an oddball stimulus, even though their basic cortical representation of the same sound (the P1/N1 complex) is normal. At the individual level, MMN was small or absent in more than 40% of individuals with LD and abnormal brainstem timing as opposed to only 10-15% among typically developing children and children with LD and normal brainstem timing (Banai et al., 2005b).

Third, Abrams et al. (in press) have shown a relationship between the degree of delay in brainstem timing and the degree of laterality in cortical auditory processing. Thus, individuals with delayed brainstem timing showed a smaller degree of left/right cortical asymmetry in response to the speech sound /da/.

Taken together, this series of studies suggests that abnormal processing at the auditory brainstem and cortex are intimately linked. While it is tempting to interpret the findings that a single deficit at the level of the brainstem is related to a wide array of abnormalities in cortical function to support a bottom-up causal relationship between the midbrain and the cortex, this is not necessarily the case. On the one hand, developmental studies indicate that the brainstem responses probably mature at an earlier age than cortical potentials (see Hood (1998) and Johnson et al. (2006) for maturation of the click- and speech- ABRs respectively, and Sharma et al. (1997), Cunningham et al. (2000) and Ponton et al. (2002) for maturation of cortical AEPs). Thus, a deficit in brainstem timing would result in degraded input to the still-developing cortex. On the other hand, similar genetic or environmental factors leading to abnormal brainstem timing could also cause abnormal cortical function. Indeed, Strata et al. (2005) reported that in rats, anoxia results in deficits in both the auditory brainstem and cortex and that the cortical abnormalities were more pronounced/severe. A third possibility was suggested by Galaburda (1999) who claimed that cortical ectopias, emerging at a relatively early developmental stage actually affect lower brain regions (i.e. the thalamus) to which they are connected and thus are responsible for temporal processing deficits observed in ectopic mice (and humans with dyslexia). Recent studies indicate that language experience affects encoding at the level of the brainstem. Krishnan et al. (2005) have shown that brainstem encoding of Mandarin speech-sound differs between native speakers of Mandarin and English speakers. These findings suggest that encoding at the level of the brainstem could be malleable to top-down effects (e.g. experience and context). A potential explanation for top-down influences on sensory processing is provided by the Reverse Hierarchy Theory (RHT, Ahissar and Hochstein, 2004). The RHT suggests that conscious perception is typically based on the highest possible representation of the stimulus

along the perceptual hierarchy. With repeated exposures, higher levels are thus likely not only to use input from lower levels, but also influence the ways the lower levels encode incoming stimuli in a context dependent manner. How top-down influences interact with developmental factors in accounting for the speech-ABR deficits in cildren with LD is at present unknown. Alternatively, the differences between Mandarin and English speakers could be accounted for by the formers greater exposure to the specific statistics of Mandarin pitch patterns. In support of this view Xu, Krishnan and Gandour (2006) have recently shown that the more robust pitch encoding in Mandarin speakers was specific to naturally occurring pith contours but not to slightly unnatural pitch contours that could still be heard as good quality Mandarin words.

The functional significance of abnormal speech-ABR

How abnormal speech encoding in the brainstem affects behavior is still poorly understood. In the two following sections we discuss speech perception, literacy-related and cognitive abilities in individuals with LD and abnormal speech-ABR. The data were obtained by pooling together data from our previously published studies (Abrams et al., in press; Banai et al., 2005a; King et al., 2002; Wible et al., 2004, 2005) and reclassifying participants with LD into normal and abnormal speech-ABR groups based on the norms presented in Table 1.

Speech-ABR and speech perception

We hypothesized that abnormal speech-ABR should manifest itself in difficulties in speech perception. To test this hypothesis, a speech discrimination task was administered to study participants in our lab. Discrimination thresholds were determined using an adaptive protocol and a four-interval 2-alternative forced choice task. Stimuli were taken from the /da-ga/

continuum. On every trial, participants heard two pairs of 100 ms syllables (e.g. /da-da/ and /da-ga/) and were required to select the pair in which the two sounds differed from each other. Initially, the endpoints of the continuum were used and following correct responses the F3 frequency of the /ga/ token was made more similar to that of the /da/ token which served as an anchor. Just noticeable difference (JND) was determined for each subject at the 69% percent correct level. See Bradlow et al., (1999) and King et al. (2002) for further description, but note that the stimuli were slight modifications of the stimuli described in these two papers. JNDs were measured in quiet and in background noise.

At the group level, the two groups of children with LD (those with normal and abnormal speech-ABRs) had significantly higher JNDs compared to normal learning children, but did not differ significantly from each other in either quiet or in the presence of background noise as shown in Table 2 (means are for 43 normal learning children, 35 children with LD and normal speech-ABR and 33 children with LD and abnormal speech-ABR). This finding is surprising, and suggests that abnormal speech-ABR is not necessary or sufficient for abnormal speech perception. However, if difficulties in phonological processing are related to abnormal phonological representations which may be the result of difficulty in the perception of fine acoustic differences it makes sense that speech discrimination will be impaired in the majority of persons with LD, irrespective of their brainstem status.

Speech-ABR, literacy-related and other cognitive abilities

All study participants in our lab are routinely tested on a psycoeducational test battery that provides information on their current level of performance on literacy-related tasks, phonological awareness and other cognitive abilities. Group means and standard deviations for 75 – 90 normal learning children, 34 - 44 children with learning disability and normal speech-ABR and 30 - 49 children with learning disability and abnormal speech-ABR are shown in Table 2. Literacy was measured using the reading and spelling subtests of the Wide Range Achievement Test (WRAT, Wilkinson, 1993) and the Word Attack subtest of the Woodcock Johnson Revised (WJ-R, Woodcock and Johnson, 1989, 1990). Phonological processing was measured using three subtests taken from the Comprehensive Test of Phonological Processing (CTOPP, Wagner et al., 1999) – Elision, Phoneme Reversal and Segmenting Nonwords, as well as the Memory for Words subtest from the WJR. Non-verbal cognitive ability was estimated using the Test of Non-verbal Intelligence (TONI-3, Brown et al., 1997) and the Brief Cognitive Scale (WJ). In addition, the Listening Comprehension and the Cross Out (a measure of visual speed of processing) subtests of the WJR were also administered. Children with LD scored lower than normal learning children on all of these measures, but children with LD and normal or abnormal speech-ABR did not differ from each other.

The analyses of the speech perception and psychoeducational data leads us to conclude, at-present, that on the one hand, the cognitive profiles of children with LDs with either normal or abnormal speech-ABRs are similar. On the other hand, available data suggest that more than 80% of LD individuals with abnormal brainstem timing are poor readers (Banai et al., 2005b). This figure is much higher than the proportion of poor readers in our larger sample (50-60% poor readers) of children with an LD diagnosis (not specifically selected for poor reading) and reflects

the generally estimated proportion of poor readers from the total LD population in the US (Snow et al., 1998). The implication is that speech-ABRs can serve to help organize the highly heterogeneous population of LDs into more homogenous subgroups, at least with respect to the physiological correlates of their LD.

Furthermore, following auditory training programs for LDs, both auditory cortical processing and speech discrimination tend to significantly improve in LDs with abnormal brainstem processing, compared to LDs whose brainstem processing is normal, even though the degree of speech perception deficits is similar in these two groups before training (Hayes et al., 2003; King et al., 2002). These outcomes suggest that, at least for speech discrimination, the etiology of the deficit may differ between children with LD with normal and abnormal speech-ABR, hence the different effects of training. Enrolling in a demanding training program is resource intensive. If further studies support these findings and perhaps extend them to other training programs and outcome measures, the speech-ABR may help to determine when to refer a child to training, and reduce the frustration of parents and educators from the uncertainty of outcomes.

Summary

Evidence accumulating during nearly a decade of research suggests that a substantial sub-population of LDs exhibit abnormal encoding of speech at the level of the brainstem. In particular, abnormal onset of the response and reduction of its magnitude over the FFR period distinguish normal from abnormal responses suggesting less precisely timed neural response to complex sounds in a subgroup of children with LD. The abnormal speech-ABR, in turn, shows a relationship to cortical processing and literacy deficits. The importance of these relationships and

the relative ease with which speech-ABR may be measured has recently led to its translation as a clinical tool – the BioMAP™ (Biological Marker of Auditory Processing, Bio-Logic, Mundelein, IL), designed to provide knowledge about physiological encoding of sound during the course of LD diagnosis. Further research and clinical use of the speech-ABR should lead to a refinement of our understanding of the neural bases of auditory processing and improve clinical diagnosis and treatment. Further research comparing children with LD with normal vs. abnormal speech-ABR on other perceptual, language and cognitive measures, familiality of LD and medical history is required to establish whether abnormal speech-ABR is associated with any specific phenotype among individuals with learning problems. Developmental cross-sectional or longitudinal studies are required to determine the relationships between abnormal brainstem function and the emergence of learning problems.

Acknowledgments

Supported by NIH/NIDCD RO1-01510, F32DC008052 and the Hugh Knowles Center for Clinical and Basic Science in Hearing and its Disorders. We would like to thank Nicole Russo, Judy Song, Trent Nicol, Gabriella Musacchia, Krista Johnson and Erika Skoe for their help at various stages of this project.

Statement regarding conflict of interest: Nina Kraus is a co-inventor of the BioMAP™ but has no financial stake in the commercialization process. Northwestern University holds the patent and licensed it to Biologic for commercialization. Karen Banai and Daniel Abrams are not involved with BioMAP™ and declare no conflict of interest.

References

Abrams DA, Nicol T, Zecker S, Kraus N (in press) Auditory brainstem timing predicts cerebral asymmetry for speech. J Neurosci.

Ahissar M, Hochstein S (2004) The reverse hierarchy theory of visual perceptual learning. Trends Cogn Sci 8: 457-464.

Banai K, Nicol T, Zecker S, Kraus N (2005a) Cortical and cognitive consequences of brainstem timing deficits. In: Association for Research in Otolaryngology Midwinter Meeting. New Orleans: ARO.

Banai K, Nicol T, Zecker SG, Kraus N (2005b) Brainstem timing: implications for cortical processing and literacy. J Neurosci 25: 9850-9857.

Bishop DV (2006) Developmental cognitive genetics: how psychology can inform genetics and vice versa. Q J Exp Psychol (Colchester) 59: 1153-1168.

Boston JR, Møller AR (1985) Brainstem auditory-evoked potentials. Crit Rev Biomed Eng 13: 97-123.

Bradlow AR, Kraus N, Nicol TG, McGee TJ, Cunningham J et al. (1999) Effects of lengthened formant transition duration on discrimination and neural representation of synthetic CV syllables by normal and learning-disabled children. J Acoust Soc Am 106: 2086-2096.

Brown L, Sherbenou RJ, Johnsen SK (1997) Test of nonverbal intelligence (3rd edition). Austin, Tx: Pro-ed.

Clarey JC, Paolini AG, Grayden DB, Burkitt AN, Clark GM (2004) Ventral cochlear nucleus coding of voice onset time in naturally spoken syllables. Hear Res 190: 37-59.

Cunningham J, Nicol T, Zecker S, Kraus N (2000) Speech-evoked neurophysiologic responses in children with learning problems: development and behavioral correlates of perception. Ear Hear 21: 554-568.

Cunningham J, Nicol T, Zecker SG, Bradlow A, Kraus N (2001) Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. Clin Neurophysiol 112: 758-767.

Delgutte B (1980) Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers. J Acoust Soc Am 68: 843-857.

Delgutte B, Kiang NY (1984a) Speech coding in the auditory nerve: I. Vowel-like sounds. J Acoust Soc Am 75: 866-878.

Delgutte B, Kiang NY (1984b) Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics. J Acoust Soc Am 75: 897-907.

Eggermont JJ, Ponton CW (2002) The neurophysiology of auditory perception: from single units to evoked potentials. Audiol Neurootol 7: 71-99.

Frisina RD (2001) Subcortical neural coding mechanisms for auditory temporal processing. Hear Res 158: 1-27.

Galaburda AM (1999) Developmental dyslexia: a multilevel syndrome. Dyslexia 5: 183-191.

Galbraith GC, Amaya EM, de Rivera JM, Donan NM, Duong MT et al. (2004) Brain stem evoked response to forward and reversed speech in humans. Neuroreport 15: 2057-2060.

Galbraith GC, Arbagey PW, Branski R, Comerci N, Rector PM (1995) Intelligible speech encoded in the human brain stem frequency-following response. Neuroreport 6: 2363-2367.

Grontved A, Walter B, Gronborg A (1988a) Auditory brain stem responses in dyslexic and normal children. A prospective clinical investigation. Scand Audiol 17: 53-54.

Grontved A, Walter B, Gronborg A (1988b) Normal ABR's in dyslexic children. Acta Otolaryngol Suppl 449: 171-173.

Hall JW (1992) Handbook of auditory evoked responses. Needham Heights, Mass: Allyn & Bacon.

Hayes EA, Warrier CM, Nicol TG, Zecker SG, Kraus N (2003) Neural plasticity following auditory training in children with learning problems. Clin Neurophysiol 114: 673-684.

Heim S, Keil A (2004) Large-scale neural correlates of developmental dyslexia. Eur Child Adolesc Psychiatry 13: 125-140.

Hood LJ (1998) Clinical applications of the auditory brainstem response. San Diego, CA: Singular Publishing Group, Inc.

Irvine DR (1992) Physiology of the auditory brainstem. In: The mammalian auditory pathway:Neurophysiology (Popper A.N., Fay R.R., eds.), pp 153-231. Ney York, NY: Springer-Verlag.

Jacobsen J (1985) The auditory brainstem response. San Diego, CA: College-Hill Press.

Jerger S, Martin RC, Jerger J (1987) Specific auditory perceptual dysfunction in a learning disabled child. Ear Hear 8: 78-86.

Johkura K, Matsumoto S, Hasegawa O, Kuroiwa Y (1998) Defective auditory recognition after small hemorrhage in the inferior colliculi. J Neurol Sci 161: 91-96.

Johnson KL, Nicol TG, Kraus N (2005) Brain Stem Response to Speech: A Biological Marker of Auditory Processing. Ear Hear 26: 424-434.

Johnson KL, Nicol TG, Kraus N (2006) Maturation of the Speech-Evoked Auditory Brainstem Response. In: American Auditory Society Meeting, March 5-7, Scottsdale, AZ.

Keith RW, Jacobson JT (1985) Physiological responses in multiple sclerosis and other demyelinating diseases. In: The auditory brainstem response (Jacobson J.T., ed.). San Diego, CA: College-Hill Press.

King C, Warrier CM, Hayes E, Kraus N (2002) Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. Neurosci Lett 319: 111-115.

Kraus N, McGee T (1992) Electrophysiology of the human auditory system. In: The mammalian auditory system (A Popper, R Fay, eds.), pp 335-404. New-York: Springer Verlag.

Kraus N, Nicol T (2005) Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. Trends Neurosci 28: 176-181.

Krishnan A (2002) Human frequency-following responses: representation of steady-state synthetic vowels. Hear Res 166: 192-201.

Krishnan A, Xu Y, Gandour J, Cariani P (2005) Encoding of pitch in the human brainstem is sensitive to language experience. Brain Res Cogn Brain Res 25: 161-168.

Krishnan A, Xu Y, Gandour JT, Cariani PA (2004) Human frequency-following response: representation of pitch contours in Chinese tones. Hear Res 189: 1-12.

Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. J Neurophysiol 60: 1799-1822.

Lauter JL, Wood SB (1993) Auditory-brainstem synchronicity in dyslexia measured using the REPs/ABR protocol. Ann N Y Acad Sci 682: 377-379.

Lyytinen H, Guttorm TK, Huttunen T, Hamalainen J, Leppanen PHT et al. (2005) Psychophysiology of developmental dyslexia: A review of findings including studies of children at risk for dyslexia. Journal of Neurolinguistics 18: 167-195.

Marler JA, Champlin CA (2005) Sensory processing of backward-masking signals in children with language-learning impairment as assessed with the auditory brainstem response. J Speech Lang Hear Res 48: 189-203.

Martin WH, Mishler ET (2002) Intraoperative monitoring of auditory evoked potentials and facial nerve electromyography. In: Handbook of clinical audiology (Katz J., ed.), pp 323-348. Baltimore, MD: Williams & Wilkins Company.

Mason SM, Mellor DH (1984) Brain-stem, middle latency and late cortical evoked potentials in children with speech and language disorders. Electroencephalogr Clin Neurophysiol 59: 297-309.

McAnally KI, Stein JF (1997) Scalp potentials evoked by amplitude-modulated tones in dyslexia. J Speech Lang Hear Res 40: 939-945.

Møller AR (1999) Neural mechanisms of BAEP. Electroencephalogr Clin Neurophysiol Suppl 49: 27-35.

Musiek FE, Charette L, Morse D, Baran JA (2004) Central deafness associated with a midbrain lesion. J Am Acad Audiol 15: 133-151; quiz 172-133.

Musiek FE, Gollegly KM (1985) ABR in eight nerve and low brainstem lesions. In: The auditory brainstem response (Jacobson J.T., ed.). San Diego, CA: College-Hill Press.

Oertel D (1997) Encoding of timing in the brain stem auditory nuclei of vertebrates. Neuron 19: 959-962.

Plyler PN, Ananthanarayan AK (2001) Human frequency-following responses: representation of second formant transitions in normal-hearing and hearing-impaired listeners. J Am Acad Audiol 12: 523-533.

Ponton C, Eggermont JJ, Khosla D, Kwong B, Don M (2002) Maturation of human central auditory system activity: separating auditory evoked potentials by dipole source modeling. Clin Neurophysiol 113: 407-420.

Purdy SC, Kelly AS, Davies MG (2002) Auditory brainstem response, middle latency response, and late cortical evoked potentials in children with learning disabilities. J Am Acad Audiol 13: 367-382.

Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc Natl Acad Sci U S A 97: 11800-11806.

Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS et al. (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat Neurosci 2: 1131-1136.

Russo N, Nicol T, Musacchia G, Kraus N (2004) Brainstem responses to speech syllables. Clin Neurophysiol 115: 2021-2030.

Schreiner CE, Langner G (1988) Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. J Neurophysiol 60: 1823-1840.

Sharma A, Kraus N, McGee TJ, Nicol TG (1997) Developmental changes in P1 and N1 central auditory responses elicited by consonant-vowel syllables. Electroencephalogr Clin Neurophysiol 104: 540-545.

Sinex DG, Chen GD (2000) Neural responses to the onset of voicing are unrelated to other measures of temporal resolution. J Acoust Soc Am 107: 486-495.

Snow CE, Burns MS, Griffin P, eds. 1998. Preventing reading difficulties in young children. Washington DC: National Academic Press.

Song JH, Banai K, Russo NM, Kraus N (2006) On the Relationship between Speech- and Nonspeech-Evoked Auditory Brainstem Responses. Audiol Neurootol 11: 233-241.

Strata F, Deipolyi AR, Bonham BH, Chang EF, Liu RC et al. (2005) Perinatal anoxia degrades auditory system function in rats. Proc Natl Acad Sci U S A 102: 19156-19161.

Wagner RK, Torgesen JK, Rachotte CA (1999) Comprehensive test of phonological processing (CTOPP). Austin: Pro-Ed, Inc.

Wible B, Nicol T, Kraus N (2004) Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. Biol Psychol 67: 299-317.

Wible B, Nicol T, Kraus N (2005) Correlation between brainstem and cortical auditory processes in normal and language-impaired children. Brain 128: 417-423.

Wilkinson GS (1993) Wide range achievement test (WRAT3). Wilmington, DE: Wide Range, Inc.

Woodcock RW, Johnson MB (1989, 1990) Woodcock Johnson psycho-educational battery: Tests of cognitive ability (WJ-R). Allen, TX: DLM Teaching Resources.

Xu Y, Krishnan A, Gandour JT (2006) Specificity of experience-dependent pitch representation in the brainstem. Neuroreport 17: 1601-1605.

Table 1. Normative speech-ABR values based on 88 typically developing 8-12 years old children. A. Transient measures. B. Sustained measures (12-47 ms). Mean ± s.d. values are shown.

Speech-ABR measures		
A. Transient	Peak Latencies (ms)	Peak Amplitudes (μV)
V	6.68 ± 0.25	0.31 ± 0.15
A	7.59 ± 0.31	-0.67 ± 0.17
C	17.86 ± 0.48	-0.32 ± 0.13
D	22.29 ± 0.43	-0.33 ± 0.17
E	30.99 ± 0.44	-0.39 ± 0.13
F	39.54 ± 0.44	-0.44 ± 0.19
O	47.95 ± 0.52	-0.19 ± 0.11
B. Sustained	Magnitude	
RMS	0.20 ± 0.03	
FFT F0 (μV)	0.081 ± 0.032	
FFT F1 (μV)	0.034 ± 0.009	

Table 2. Speech, literacy related and cognitive abilities in normal learning (NL) and learning disabled groups (LD) with normal and abnormal speech-ABR. Mean ± s.d. values are shown. Values in bold type indicate that the highlighted group was significantly different ($p \leq 0.037$) from the other groups on a Scheffe post-hoc comparison

	NL	LD normal speech-ABR	LD abnormal speech-ABR	F (p)
Speech Perception (JNDs, Hz)				
Quiet	**105 ± 49**	149 ± 75	155 ± 69	6.99 (0.001)
Noise	232 ± 130	317 ± 110[&]	285 ± 141	4.46 (0.014)
Phonological Abilities (CTOPP scores)				
Elision	**11.7 ± 2.1**	7.9 ± 2.9	7.9 ± 2.8	41.02 (<0.001)
Phoneme Reversal	**11.0 ± 2.5**	7.7 ± 1.8	7.6 ± 1.9	39.84 (<0.001)
Segmenting Nonwords	**10.8 ± 1.9**	9.4 ± 2.3	8.6 ± 2.4	13.17 (<0.001)
Literacy (standard scores)				
Reading	**115 ± 11**	89 ± 10	85 ± 13	140.4 (<0.001)
Spelling	**115 ± 14**	88 ± 8	85 ± 11	140.6 (<0.001)
Word Attack	**117 ± 14**	90 ± 9	88 ± 11	114.0 (<0.001)
Other Cognitive Abilities (standard scores)				
Memory For Words	**108 ± 16**	94 ± 11	95 ± 12	21.2 (<0.001)
Listening Comprehension	**120 ± 16**	106 ± 19	105 ± 19	22.2 (<0.001)
Cross Out	**114 ± 13**	100 ± 16	98 ± 19	15.4 (<0.001)
Brief Cognitive Scale	**123 ± 12**	100 ± 13	100 ± 15	72.8 (<0.001)
TONI-3[$]	**117 ± 16**	107 ± 16	101 ± 11	10.0 (<0.001)

[&]This group was significantly different from NLs but not from the other group of LDs.

[$]This test was completed by 61 Normal learning children, 23 children with LD and normal speech-ABR and 21 children with LD and abnormal speech-ABR.

Figure Legends

Figure 1. Top. Amplitude vs. time waveform of the syllable /da/. **Bottom**. Example of a typical speech-ABR waveform recorded to a 40 ms 80 dB /da/ (stimuli were presented at a rate of 11/s, response is the average of 6000 presentations) showing the onset and the FFR portion of the response. The stimulus has been shifted by ~7 ms (representing the delay in neural conduction at the brainstem) to demonstrate the similarities between the stimulus and the response over the FFR period. The thin horizontal lines intersecting the stimulus and response represent 0 μV.

Figure 2. Mean FFT magnitude (average spectra from 23 to 44 ms) during the periodic portion of the response for 90 normal learning children. Spectral peaks are observed at regions corresponding to F0 and F1 in the /da/ stimulus (stimulus F0: 103-125 Hz; F1: 220-720 Hz), however, F0 is more strongly represented. The thick line denotes mean magnitude; thin dashed lines are ±1 s.d. of the mean.

Figure 1

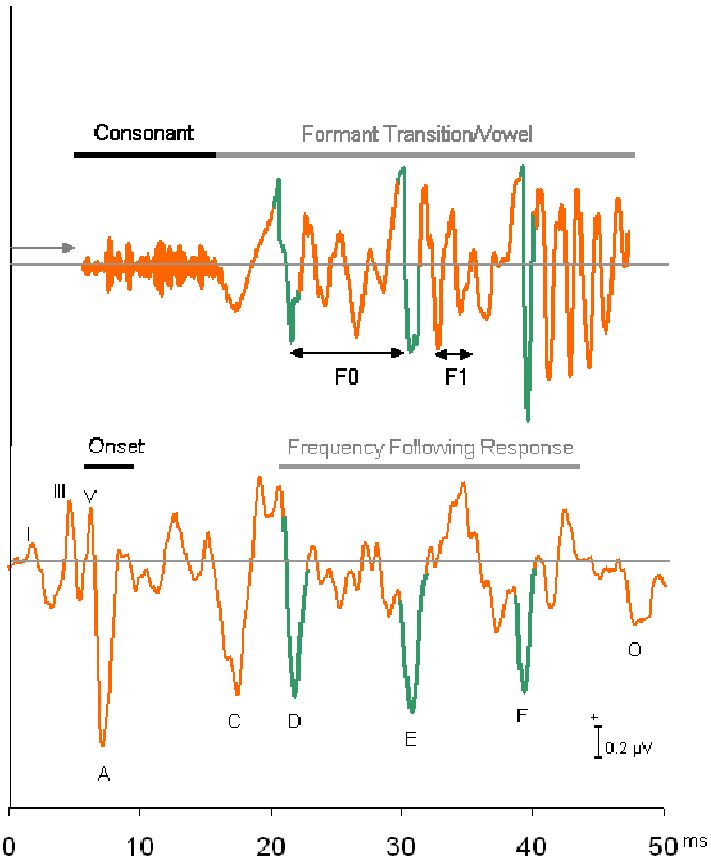
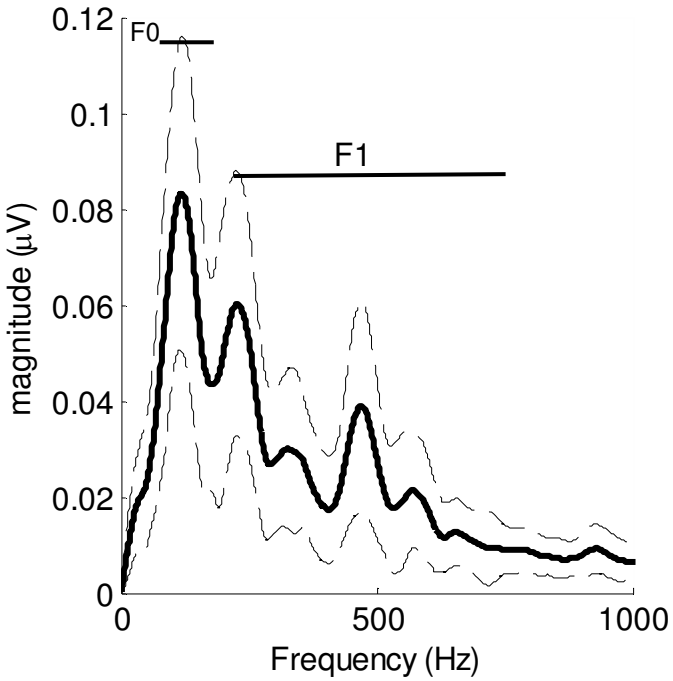

Consonant
Formant Transition/Vowel
F0
F1
Onset
Frequency Following Response
I
III
V
A
C
D
E
F
O
0.2 µV
0
10
20
30
40
50 ms

Figure 2


0.12
0.1
0.08
0.06
0.04
0.02
0
magnitude (μV)
F0
F1
0
500
1000
Frequency (Hz)