Neural Correlates of Auditory Perceptual Organization Measured with Direct Cortical Recordings in Humans

by

ANDREW R. DYKSTRA

S.M., Electrical Engineering and Computer Science (2008)
Massachusetts Institute of Technology, Cambridge, MA

B.S., Electrical Engineering – Audio Option (2005)
University of Miami, Coral Gables, FL

SUBMITTED TO THE HARVARD-MIT DIVISION OF HEALTH SCIENCES AND TECHNOLOGY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE OF

DOCTOR OF PHILOSOPHY IN SPEECH AND HEARING BIOSCIENCE AND TECHNOLOGY
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August, 2011
[September 2011]

Signature of Author: _____

Speech and Hearing Bioscience Technology Program
Harvard-MIT Division of Health Sciences and Technology
August 6, 2011

Certified by: _____

Sydney S. Cash, MD, PhD
Assistant Professor, Department of Neurology
Massachusetts General Hospital and Harvard Medical School
Thesis Supervisor

Certified by: _____

Jennifer R. Melcher, PhD
Associate Professor, Department of Otology and Laryngology
Massachusetts Eye and Ear Infirmary and Harvard Medical School
Thesis Committee Chair

Accepted by: _____

Ram Sasisekharan, PhD
Underwood-Prescott Professor of Biological Engineering
Director, Health Sciences and Technology

Neural Correlates of Auditory Perceptual Organization Measured with Direct Cortical Recordings in Humans

by

ANDREW R. DYKSTRA

Submitted to the Harvard-MIT Division of Health Sciences and Technology on August 6, 2011 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Speech and Hearing Bioscience and Technology

ABSTRACT

One of the primary functions of the human auditory system is to separate the complex mixture of sound arriving at the ears into neural representations of individual sound sources. This function is thought to be crucial for survival and communication in noisy settings, and allows listeners to selectively and dynamically attend to a sound source of interest while suppressing irrelevant information. How the brain works to perceptually organize the acoustic environment remains unclear despite the multitude of recent studies utilizing microelectrode recordings in experimental animals or non-invasive human neuroimaging. In particular, the role that brain areas outside the auditory cortex might play is, comparatively, vastly understudied.

The experiments described in this thesis combined classic behavioral paradigms with electrical recordings made directly from the cortical surface of neurosurgical patients undergoing clinically-indicated invasive monitoring for localization of epileptogenic foci. By sampling from widespread brain areas with high temporal resolution while participants simultaneously engaged in streaming and jittered multi-tone masking paradigms, the present experiments sought to overcome limitations inherent in previous work, namely sampling extent, resolution in time and space, and direct knowledge of the perceptual experience of the listener.

In experiment 1, participants listened to sequences of tones alternating in frequency (i.e., ABA-) and indicated whether they perceived the tones as grouped ("1 stream") or segregated ("2 streams"). As has been reported in neurologically-normal listeners since the 1950s, patients heard the sequences as grouped when the frequency separation between the A and B tones was small and segregated when it was large. Evoked potentials from widespread brain areas showed amplitude correlations with frequency separation but surprisingly did not differ based solely on perceptual organization in the absence of changes in the stimuli. In experiment 2, participants listened to sequences of jittered multi-tone masking stimuli on which a regularly-repeating target stream of tones was sometimes superimposed and indicated when they heard the target stream. Target detectability, as indexed behaviorally, increased throughout the course of each sequence. Evoked potentials and high-gamma activity differed strongly based on the listener's subjective perception of the target tones. These results extend and constrain theories of how the brain subserves auditory perceptual organization and suggests several new avenues of research for understanding the neural mechanisms underlying this critical function.

Thesis Supervisor: Sydney S. Cash
Assistant Professor of Neurology
Massachusetts General Hospital and Harvard Medical School

Biographical note

Andrew R. Dykstra
617.895.7195
adykstra@mit.edu

EDUCATION

Massachusetts Institute of Technology — Cambridge, MA
Doctor of Philosophy — Expected: Summer, 2011
Harvard-MIT Division of Health Sciences and Technology
Speech and Hearing Bioscience and Technology
Thesis Supervisor: Sydney S. Cash
Thesis Committee Chair: Jennifer R. Melcher

Massachusetts Institute of Technology — Cambridge, MA
Master of Science — September, 2008
Electrical Engineering and Computer Science
GPA: 5.0 on a 5.0 scale
Thesis Supervisor: Jennifer R. Melcher

University of Miami, *Cum Laude* — Coral Gables, FL
Bachelor of Science — May, 2005
Electrical and Computer Engineering – Audio Option
GPA: 3.9 on a 4.0 scale

AWARDS

University of Miami

Edward Arnold Scholarship for Audio Engineering	August, 2003 – May, 2005
Eta Kappa Nu, Electrical Engineering Honor Society	May, 2003
Tau Beta Pi, Engineering Honor Society	May, 2004
Ann Bachelor Scholarship for General Engineering	January, 2004 – May, 2005

Massachusetts Institute of Technology

Speech and Hearing Bioscience and Technology Fellowship	June, 2005 – May, 2008
Neuroimaging Training Program Fellowship	June, 2008 – May, 2009
Speech and Hearing Bioscience and Technology Fellowship	June, 2009 – May, 2010
Amelia Peabody Scholarship	June, 2010 – September, 2010

PUBLICATIONS

Dykstra AR, Koh CK, Braida LD, Tramo MJ. Dissociation of Absolute Detection and Intensity Discrimination of Pure Tones following Bilateral Lesions of Auditory Cortex, *In revision at PLoS ONE.*

Dykstra AR, Nam EC, Melcher JR. A reevaluation into the effects of cardiac gating on fMRI of the auditory system. *In preparation for Human Brain Mapping.*

Dykstra AR, Halgren E, Thesen T, Carlson C, Kuzniecky R, Doyle W, Madsen J, Eskandar E, Cash SS. Widespread Brain Areas Engaged during a Classical Auditory Streaming Task Revealed by Intracranial EEG. *Frontiers in Human Neuroscience.*

Dykstra AR, Chan AM, Quinn BT, Zepeda R, Keller CJ, Cormier JE, Cash SS. Individualized Localization and Cortical Surface-based Registration of Intracranial Electrodes. *Under review at Neuroimage.*

Dykstra AR, Halgren E, Thesen T, Carlson C, Kuzniecky R, Doyle W, Madsen J, Eskandar E, Cash SS. Multiple Timescales of Acoustic Segmentation Recorded Directly from Human Cortex. *In preparation for Journal of Neuroscience.*

Dykstra AR, Halgren E, Gutschalk A, Golby A, Eskandar E, Cash SS. Intracranial Neural Correlates of Auditory Perceptual Awareness in Humans. *In preparation for Journal of Neuroscience.*

Dykstra AR, Halgren E, Melcher, J.R., Eskandar, E., Cash, S.S., Neural correlates of auditory scene analysis in humans. Auditory System Gordon Research Conference, July, 2008, New London, NH

Dykstra AR, Halgren E, Melcher JR, Eskandar E, Cash, SS. Preliminary intracranial evidence for oscillatory entrainment as a mechanism for auditory streaming. Society for Neuroscience Annual Meeting, November, 2008, Washington D.C.

Dykstra AR, Halgren E, Thesen T, Trongnetrpunya A, Kuzniecky R, Doyle W, Madsen J, Eskandar E, Cash SS. Neural Correlates of Auditory Streaming Measured with Direct Cortical Recordings in Behaving Humans. 16th Annual Meeting of the Organization for Human Brain Mapping, June, 2010, Barcelona, Spain

Dykstra AR, Chan AM, Zepeda R, Keller CJ, Quinn BT, Cash SS. Individualized localization and cortical surface-based registration of semi-chronic intracranial electrodes. 17th Annual Meeting of the Organization for Human Brain Mapping, June, 2011, Québec, QC. Canada

Dykstra AR, Halgren E, Thesen T, Carlson C, Kuzniecky R, Doyle W, Madsen J, Eskandar E, Cash SS. Multiple Timescales of Acoustic Segmentation Recorded Directly from Human Cortex. Society for Neuroscience Annual Meeting, November, 2011, Washington D.C.

TALKS

Preliminary intracranial evidence for oscillatory entrainment as a mechanism for auditory streaming. 15th International Congress on Event-Related Potentials of the Brain, April, 2009, Bloomington, IN, USA

Direct Cortical Recordings from Humans Engaged in a Classical Auditory Streaming Task. Department of Neurology, University of Heidelberg, June, 2010, Heidelberg, Germany

PROFESSIONAL EXPERIENCE and SERVICE

President, Eta Kappa Nu — August, 2004 – May, 2005
Epsilon Kappa Chapter, University of Miami

Graduate Resident Tutor, La Maison Française — August, 2006 – June, 2009
New House Dormitory, Massachusetts Institute of Technology

Teaching Assistant, HST.583 — Fall, 2008
Functional Magnetic Resonance Imaging: Data Acquisition and Analysis

Member, Admissions Committee — Spring 2008, 2009
Speech and Hearing Bioscience and Technology Program

Organizer and Lecturer, Tutorial Series — Fall, 2008
Speech and Hearing Bioscience and Technology Program

Chair, Distinguished Lecture Series Committee — Spring, 2009
Speech and Hearing Bioscience and Technology Program

Committee Member, Student-Initiated Seminar in Auditory Physiology — Spring, 2009
Eaton-Peabody Laboratory

Member, Human Brain Mapping Society

Member, Society for Neuroscience

Acknowledgments

First and foremost, I'd like to express the deepest gratitude to my supervisor, Syd Cash, who proved an extremely patient, willing, and knowledgeable mentor. I'm also especially grateful to Jennifer Melcher who, serving as my committee chairperson, acted as a second adviser and offered up loads of advice and encouragement. My committee members Peter Cariani, Eric Halgren, and Barbara Shinn-Cunningham – and particularly Eric – were extremely patient and generous with their time and offered countless helpful suggestions.

The patients who participated in the experiments described herein undergo a tremendous ordeal for their clinical well-being, and I cannot begin to describe the respect I have for them for agreeing to participate in what must seem like esoteric, irrelevant experiments with "R2-D2" sounds.

A very special thank you is owed Rindy Northrup, who opened her home to me in the warmest of ways.

This work was not possible without the support – technical and otherwise – of many colleagues, collaborators, and hospital staff, including Cashlab members Alex Chan, Nima Dehghani, Justine Cormier, Rodrigo Zepeda, Corey Keller, Jake Donaghue, Daniel Wabo, and Omar Ahmed, MGH hospital staff Emad Eskandar, Kristi Tripp, Cara Houghton, Jay Pathmanathan, Matt Bianchi, Cat Chu-Shore, Steve Stufflebeam, and Naoro Tanaka, NYU collaborators Thomas Thesen, Chad Carlson, Amy Trongnetrpuya, Olga Felsovalyi, Brian Quinn, and Matthew Davidson, and BWH hospital staff Joe Madsen, Alex Golby, Melissa Murphy, and Paul Dionne.

I'd also like to thank my SHBT 2005 classmates, with whom I've shared many an existential conversation, as well as the broader SHBT community for its close-knit environment and paternal support.

Although it may sound odd, I'd like to thank my loving cat, Baloo, for often being my solo companion while working remotely.

I'd like to thank my parents as well as my wife's parents for constantly putting up with my desire to be a "career student," for always encouraging me to persevere in doing what I enjoy, and for setting a first-rate example.

Finally, I'd like to thank my wife, Rita Denisse Cordova-Montes, whose unrelenting encouragement, love, and perseverance carried me through the depths of the dissertation process many times over. Without her, this thesis would have been impossible.

The research performed for this thesis was supported by NIDCD grant T32 DC00038, NIBIB grant T32 EB001680, an Amelia Peabody Charitable Trust scholarship, NIH grant NS18741, and NINDS grant NS062092.

Table of Contents

List of Abbreviations

A1 – Primary auditory cortex

AER – Auditory-evoked response

ASA – Auditory scene analysis

BOLD – Blood-oxygenation level dependent

BWH – Brigham and Women's Hospital

CT – Computerized tomography

ΔF – Frequency separation

ECoG – Electrocorticography

EEG – Electroencephalography

EP – Evoked potential

ERB – Equivalent rectangular bandwidth

ERSP – Event-related spectral perturbation

fMRI – Functional magnetic resonance imaging

GBA – Gamma-band activity

iEEG – Intracranial electroencephalography or intracranial electroencephalogram

ISI – Inter-stimulus interval

MEG – Magnetoencephalography

MGH – Massachusetts General Hospital

MIT – Massachusetts Institute of Technology

MNI – Montreal Neurological Institute

MRI – Magnetic resonance imaging

MTG – Middle temporal gyrus

NIH – National Institute of Health

NYU – New York University

PET – Positron emission tomography

PLST – Posterior lateral superior temporal area

RAS – Right, anterior, superior, a Freesurfer coordinate system

RMS – Root-mean squared

SOA – Stimulus-onset asynchrony

SSE – Sum-squared error

STG – Superior temporal gyrus

STP – Superior temporal plane

STS – Superior temporal sulcus

List of Figures

Chapter 1: Introduction

1.1 Auditory scene analysis

One of the primary functions of the auditory system is to decompose the complex mixture of sound arriving at the ears into a neural representation which isolates individual sound sources. This process, termed *auditory scene analysis* (ASA), is thought to be crucial for survival and communication in noisy environments such as a crowded restaurant or busy street corner. The abilities to understand speech or perceive music – uniquely human abilities – are both thought to rely on the ability of the auditory system to first separate incoming information into components where, when the system functions properly, each component in the representation reflects a single source in the environment. Failures to segregate incoming acoustic information result in illusory grouping of sounds from multiple sources as well as in sound sources going undetected entirely. Even in the absence of traditionally-defined presbyacusis, older listeners often report difficulty in situations which tax the segregating abilities of the auditory system. Figure 1.1 shows a cartoon which highlights the conditions under which the auditory system's ability to segregate sound is necessary (*B. Shinn-Cunningham, personal communication*). In this example, there are several conversations proceeding simultaneously, yet most young and early middle-aged listeners have no problem understanding the message of the person with whom they are engaged - the cocktail-party effect (Cherry 1953). Furthermore, this ability is dynamic in that a listener can focus their attention on different sources at different moments in time.

At the most basic level of analysis, the perceptual organization of the auditory scene begins with the quanta – i.e. most basic attributes – of auditory objects. These attributes include pitch, loudness, duration timbre, and location in space, and are largely perceptually independent

from each other. That is, a change in one perceptual attribute (e.g. pitch) does not tend to covary with a change in another perceptual attribute (e.g. loudness). Similar to the way in which it's defined in vision, an auditory object can be thought of as a unitary perceptual consisting of an ensemble of basic attributes which are bound together to form a perceptual whole. This is the auditory version of the binding problem defined over 100 years ago by the Gestalt psychologists (Köhler 1947; Koffka 1935) and elaborated upon by several recent authors (for reviews see *Neuron* **24** 1999). At any moment during our auditory experience, there can multiple objects present in the auditory scene which may or may not correspond to physical sound sources present in the acoustic environment.

Much is now known about the object-level perceptual grouping cues that are inherent in sound as well as the schema-driven processes which allow listeners to segregate one sound source from another (Bregman, 1994; Darwin, 1997; Carlyon, 2004). In general sound energy tends to group perceptually when the energy (i) is circumscribed in time-frequency space, (ii) is inherently harmonic, (iii) possesses common onsets across frequency, or (iv) is perceived to arise from a similar spatial location. Auditory events (abrupt changes in sound energy from one moment to the next) can be further grouped together across time into "streams" if the basic perceptual attributes constituting successive events are sufficiently similar to one another. This grouping or segregation of successive auditory events is known as auditory streaming, typical examples of which include melody (successive grouping of different pitches) or rhythm (temporal patterns of related auditory events) (Handel 1993; Cariani 2011; Meyer 1956; Bregman 1990).

Figure 1.1. The cocktail-party problem.

How auditory perceptual organization is implemented in the neural architecture of the auditory system is not nearly as well understood as the perceptual phenomena themselves [for recent reviews, see (Micheyl et al., 2007b; Snyder and Alain, 2007b; Bidet-Caulet and Bertrand, 2009; Bee and Micheyl, 2008; Shamma and Micheyl, 2010; Shamma et al., 2010)], despite several recent studies utilizing both microelectrode recordings in experimental animals (Fay, 1998, 2000; Fishman et al., 2001, 2004; Kanwal et al., 2003; Bee and Klump, 2004, 2005; Micheyl et al., 2005; Schul and Sheridan, 2006; Pressnitzer et al., 2008; Elhilali et al., 2009a; Itatani and Klump, 2009, 2010; Bee et al., 2010) and human neuroimaging techniques (Sussman et al., 1999; Gutschalk et al., 2005, 2007; Snyder et al., 2006; Snyder and Alain, 2007a; Wilson et al., 2007; Cusack, 2005; Deike et al., 2004, 2010; Kondo and Kashino, 2009; Schadwinkel and Gutschalk, 2010a, 2010b, 2011). In particular, there is an extreme paucity of data regarding the role of brain areas outside of the classically-defined auditory cortex in ASA [but see (Pressnitzer et al., 2008; Cusack, 2005; Kondo and Kashino, 2009)].

1.1.1 Auditory streaming

One commonly-used paradigm to study the basic neural mechanisms involved in auditory perceptual organization involves presenting a listener with triplets of the form ABA-ABA- (Figure 1.2), where A and B are short-duration sounds (usually pure tones), and the dash is a silent gap (van Noorden, 1975). When either the frequency separation is small or the presentation rate slow, the tones are perceived as grouped and as arising from a single environmental sound source.

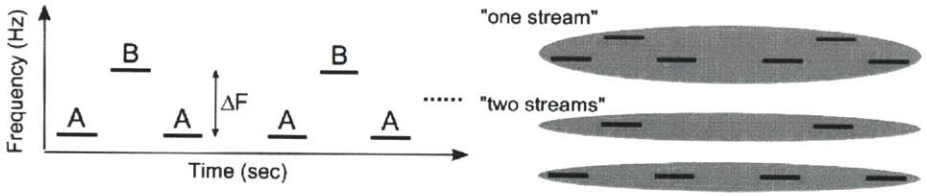

Figure 1.2. Depiction of auditory streaming paradigm.

When the frequency separation is large or the presentation rate fast, the A and B tones are perceived as segregated, each forming their own auditory "stream." Figure 1.3 shows the common percepts of a typical listener across the parameter space defined by frequency separation and presentation rate (van Noorden, 1975). Phenomenally, given the typical tone durations and inter-stimulus intervals used during this paradigm, when the A and B tones group (i.e., 1 stream), a distinctive galloping rhythm is heard which is segmented into successive 3-tone triplets occurring approximately every half second. When the A and B tones segregate, the percept of a galloping rhythm is lost; what can be heard in its place are two isochronous rhythms, one of the A tones at a rate half that of the B tones. Furthermore, the most common report among listeners hearing two streams is that of a perceptual foreground and background, either of which is occupied by either A tones or B tones at a given moment.

Sussman et al. (1999) were the first to use such sounds while simultaneously measuring neural activity (Sussman et al., 1999). They recorded event-related potentials while subjects

ignored alternating tone sequences which varied in their rate of presentation. When the presentation rate was slow, conditions known to produce the percept of individual tones heard in succession, infrequent deviant tones did not evoke a mismatch negativity; when the rate was fast, conditions known to produce the percept of separate auditory streams (one of the high tones and one of the low tones), same infrequent deviant tones did evoke a mismatch negativity response, indicating that change detection occurs within, but not across, auditory streams.

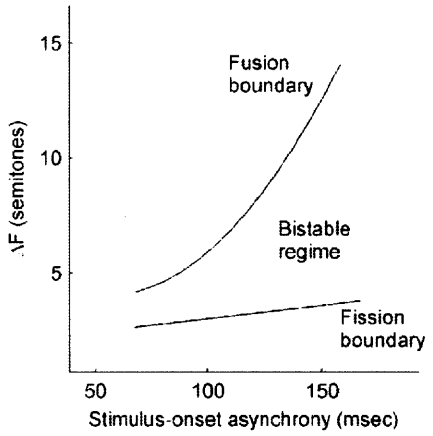

Figure 1.3. "Fission" and "fusion" boundaries during the ABA auditory streaming paradigm, depicting the boundaries at which one stream (fusion) or two streams (fission) can no longer be perceived.

Fishman et al. (2001) were the first to explicitly correlate neural responses with changes in a stimulus parameter – the presentation rate of alternating-frequency pure tones – known to produce changes in how the tones are perceptually organized by human listeners (Fishman et al., 2001). Multi-unit activity and current source densities were recorded from several sites in primary auditory cortex (A1) of awake monkeys. When the presentation rate was slow, responses to both A and B tones were found in sites with best frequencies close to the A-tone frequency. In contrast, when the presentation rate was fast, responses to the B tones were differentially suppressed, leaving only responses to the A tones. This suppression was made

stronger by increasing the frequency separation between the A and B tones, again paralleling human psychophysical results with similar parameter manipulations. These results have since been replicated several times by both animal neurophysiological studies (Fishman et al., 2004; Micheyl et al., 2005; Bee et al., 2010) and human neuroimaing (Gutschalk et al., 2005; Snyder et al., 2006; Wilson et al., 2007; Snyder and Alain, 2007a).

1.1.2 Informational masking

Another commonly-used paradigm to study auditory perceptual organization, at least from a psychoacoustical perspective, involves presenting a listener with a sequence of tones placed randomly in time and frequency with, on some sequences, a regularly-repeating stream of target tones (Figure 1.4). On some trials in which the regularly-repeating target stream of tones is present, they will "pop out" from the otherwise random mixture as an independent stream in the form of a simple isochronous rhythm. On other trials, they will go entirely unperceived.

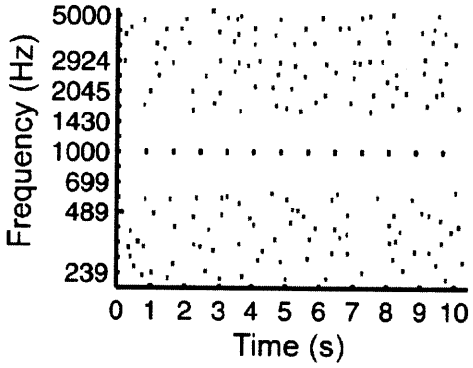

Figure 1.4. Depiction of the jittered multi-tone masking stimulus.

This task produces large changes in subjective perception despite physically-identical stimuli in that the target tones sometimes are detected but sometimes go undetected. In the non-jittered version of this task (i.e., when the masker tones are always synchronous with the target), randomly varying the frequency content of the masker tones while holding the target tone

constant can produce as much as a 40 dB shift in detection threshold relative to the fixed masker case (Neff and Green, 1987; Kidd et al., 2003, 2008). Furthermore, it has been shown using stimuli with a protected frequency region (as in the stimuli shown in Figure 1.4) that the behavioral masking effect of the multi-tone background is thought not to be entirely due to energetic masking effects at the level of the auditory periphery or, for that matter, at any tonotopically-organized site along the ascending auditory pathway (Neff et al., 1993; Kidd et al., 1994, 2008; Micheyl et al., 2007a). Instead, researchers have conceptualized an information-processing bottleneck in the brain (Overath et al., 2007; Gutschalk et al., 2008; Zylberberg et al., 2010) that gates a sensory stimulus' access to perceptual awareness. Behaviorally, this type of masking has been termed "informational masking." Consistent with the idea of an information-processing bottleneck in non-primary auditory cortex, Gutschalk et al. (2008) measured long-latency auditory evoked fields which differed substantially based on whether or not the target tones (as shown in Figure 1.4) were detected (Gutschalk et al., 2008). Detected targets evoked a large N1m-like response which localized to the superior temporal plane either on the lateral aspect of Heschl's gyrus or just posterior to it, in non-primary auditory cortex. Undetected targets evoked activity similar to the case in which the target tones were not present at all. The only other neural study to use jittered multi-tone masking stimuli focused on the effects of selectively attending to the regularly-repeating target tones and found that doing so enhances it's representation (as indexed by power in the MEG response at the frequency of occurrence of the target tones) (Elhilali et al., 2009b).

1.1.3 Multistable perception

Multistable perception, or the tendency of a given physical stimulus to elicit two or more distinct but stable percepts, is a well-known phenomenon that can occur in a variety of sensory

modalities, including vision, somatosensation and, most interestingly for the purpose of this thesis, audition. The phenomenon is highlighted by the classic Necker cube shown in Figure 1.5, which can be experienced in two ways by an observer, with the foreground of the figure being located (i) in the upper-right portion or (ii) lower-left portion (Necker, 1832). It is now well known that alternations in how the cube is perceived can occur both spontaneously and with effort, and that this phenomenon can be extended to several other types of stimuli in the visual, somatosensory, and auditory systems. The temporal dynamics of such switching behavior are also similar across the various sensory modalities (Pressnitzer and Hupé, 2006).

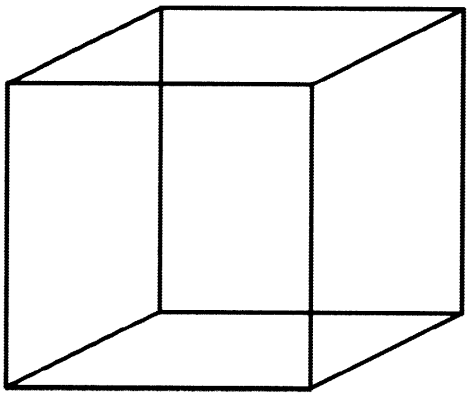

Figure 1.5. Necker cube demonstrating the concept of multistable perception.

From a neural point of view, such multistable stimuli are of extreme interest due to their utility in isolating brain responses related directly to one's perceptual experience, the so-called Neural Contents of Consciousness (Koch, 2009). By comparing responses to the same physical stimulus for different perceptual experiences of the same stimuli, it is possible to dissociate brain activity related to parameters of the stimulus vs. brain activity related directly to perception. This approach is widely used in visual scene analysis (Leopold and Logothetis, 1999; Rees, 2009; Sterzer et al. 2009), but there are surprisingly few examples of it in other sensory modalities, including the auditory domain (Sterzer et al. 2009). Current theories of the neural

basis for whether a sensory stimulus will (or will not) reach perceptual awareness posit the necessity for the instantiation of large-scale feedback loops between frontoparietal areas and sensory cortex (in the case of vision, visual cortex, etc.) in order for a subject to become aware of it (Del Cul et al. 2007, 2009; Gaillard et al. 2009; Libedinsky and Livingstone 2011).

The clearest demonstration of neural covariates of perception comes from studies utilizing a phenomenon known as binocular rivalry (Tong et al. 2006) in which distinct images are presented to each eye. Instead of perceiving a mixture of the two images, observers report perceiving each image in alternation. As could be expected, when the stimuli consist of images which drive different neural populations (e.g., a face and a house, which would elicit activity in the "face" and "house areas" of the inferior temporal cortex when presented in isolation), activity in these different populations correlates with which image has access to consciousness at successive moments in time (Leopold and Logothetis 1999; Blake and Logothetis 2002; Rees 2007). Activity in early visual areas (i.e. primary visual cortex and the lateral geniculate nucleus) has also been shown to covary with subjective visual perception during binocular rivalry as well as other bistable visual paradigms (Haynes and Rees 2006).

Only four studies to date have shown neural correlates of subjective auditory perception (Hillyard et al. 1971; Cusack et al. 2005; Gutschalk et al. 2005, 2008). The results from these studies suggest that correlates of auditory bistability may be found in secondary auditory cortex (Gutschalk et al. 2005, 2008) as well as supra-modal brain areas (Hillyard et al. 1971; Cusack et al. 2005).

The simplest paradigm capable of generating bistable perception is detection. In the auditory domain, a near-threshold sound with the same trial-to-trial intensity will only sometimes be perceived by the listener (Hillyard et al., 1971). The two aforementioned ASA paradigms –

auditory streaming and the multi-tone masking paradigm – also produce bistable perception (Anstis and Saida, 1985; Carlyon et al., 2001; Neff and Green, 1987; Kidd et al., 2003). In the case of streaming, the same physical sequence can be heard as either grouped (i.e., "one stream") or segregated (i.e., "two streams"). In the case of the multi-tone masking paradigm, target tones can go either detected or undetected.

A qualitative model for bistable perception is depicted in Figure 1.6. The horizontal axis depicts a linear variation of a given stimulus parameter (e.g. frequency separation between A and B tones in the ABA streaming paradigm). The vertical axis depicts some uni-dimensional measure of the neural response to such stimuli. A parametric variation of a given stimulus or stimulus feature could produce neural activity patterns which vary linearly or categorically as shown by the blue and red curves, respectively. Noise in the response of a population showing a linear relationship with the stimulus, when fed to a population showing a more categorical relationship, could engender sufficient trial-to-trial variability for bistable perception. While such activity patterns have been widely reported in vision [for reviews see (Logothetis, 1998; Leopold and Logothetis, 1999; Sterzer et al., 2009)], only limited evidence for such a mechanism exists in the auditory system (Hillyard et al., 1971; Cusack, 2005; Gutschalk et al., 2005, 2008; Riecke et al., 2007, 2009; Kondo and Kashino, 2009; Lemus et al., 2009a, 2009b; Schadwinkel and Gutschalk, 2011).

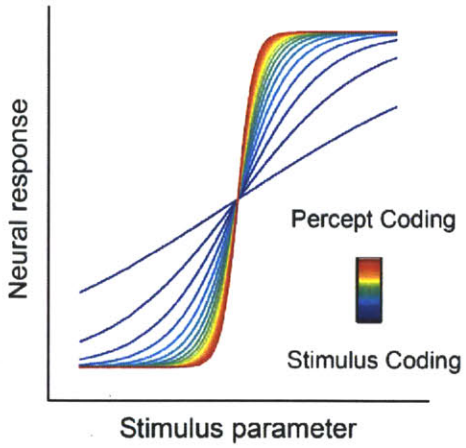

Figure 1.6. Qualitative model of bistable perception.

The y-axis can be thought of as a uni-dimensional projection of an N-dimensional space rather than the amplitude of a given measure of neural activity. Presumably, such a state-space exists in which the neural activity associated with two different percepts given a fixed stimulus can be robustly dissociated.

1.2 Intracranial EEG

Since the 1950s, epileptologists have used intraoperative surface electrocorticography in epileptic neurosurgical candidates to map both epileptogenic and eloquent cortex (Penfield and Jasper, 1954). Recently, basic cognitive neuroscientists have realized the opportunity that such clinically indicated invasive recordings offer and have begun taking advantage of the unique high spatiotemporal resolution afforded in such settings (Engel et al., 2005). Currently, such recordings are even sometimes semi-chronic, allowing researchers to sample from widespread brain areas simultaneously while patients are awake and behaving. The technique circumvents limitations inherent in commonly used noninvasive neuroimaging methods (e.g., electro- or magneto encephalography and fMRI. As more and more epilepsy centers around the country and around the world have begun their own neurosurgery units, intracranial EEG has become an increasingly popular method to map the neural correlates of normal cognitive function.

Although each epilepsy center is slightly idiosyncratic in its surgical procedures for invasive monitoring, many centers use some form of semi-chronic electrocorticography, i.e. sub-dural electrodes resting directly on the pial surface. An intraoperative photograph of a typical example from such a case is shown in Figure 1.7. Panel A of this figure shows the exposed pial surface of one patient during the implant procedure, where the craniotomy has been performed and the dura has been peeled away from the parenchyma. Panel B shows the same patient's brain with an 8x8 array of platinum disk electrodes overlaid. As can be seen from Figure 1.7, widespread cortical brain areas are sampled simultaneously from such an array of electrodes, allowing for coverage of distinct functional anatomical networks without suffering from common interpretational issues associated with noninvasive methods such as the ill-posed inverse problem in EEG and MEG and the unspecified link between neural activity and hemodynamics in fMRI.

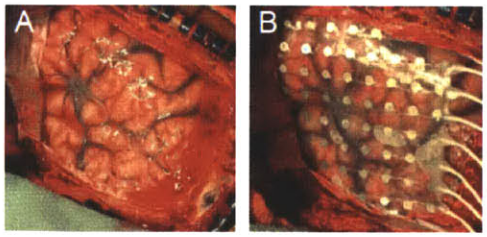

Figure 1.7. Intraoperative photographs of a sub-dural grid implant.

A second commonly used method for invasive monitoring of epileptogenic cortex involves inserting linear arrays of penetrating depth electrodes to target mesial structures of the temporal, frontal, and parietal lobes. A depiction of such an array is shown in Figure 1.8. Panel A shows a 1x8 electrode array (in red dots) which has been overlayed onto a resliced image in the cortical volume (slice orientation shown in B).

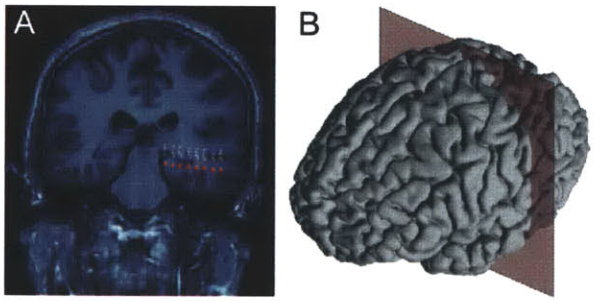

Figure 1.8. Localization of a depth-electrode array.

These electrode arrays often sample from medial temporal and medial frontal lobe structures, such as hippocampal and peri-hippocampal areas as well as cingulate cortex, structures that are often - if not always - inaccessible to either EEG or MEG. These depth electrode arrays are also often combined with arrays resting on the pial surface.

Many neuroscientific advances have come from the various methods used for iEEG.

While a comprehensive review of all such studies is beyond the scope of this thesis, a subset of the intracranial studies which focused on the auditory system, or relevant from other sensory systems, are reviewed briefly below.

Several groups have attempted to use iEEG to define functional auditory areas in the superior temporal plane and lateral superior temporal cortex, including anatomical boundaries defined by the transverse gyrus of Heschl, the planum temporale, and the superior temporal gyrus (Liégeois-Chauvel et al., 1991, 1994; Howard et al., 1996, 2000; Brugge et al., 2003, 2008; Oya et al., 2007). Based on overall responsiveness, relative latency, tonotopic organization, and functional connectivity measures, these studies indicate an areal sub-definition of the posterior superior temporal plane that includes the posteromedial and anterolateral portions of the transverse gyrus as well as a circumscribed area of the posterior STG, with only the posteromedial portion of Heschl's gyrus defined as the core area. One seminal study (Bitterman et al., 2008) has demonstrated that single neurons in the human auditory cortex (likely the core area) exhibit frequency tuning that is much sharper than previously shown in experimental animals, either at the level of the cortex of at the auditory periphery.

Some of the earliest work to use iEEG for neurocognitive purposes came from Halgren and colleagues (Halgren et al., 1980). Halgren et al. (1980) used an auditory change-detection paradigm and demonstrated large closed-field potentials in the hippocampus, peaking between 300 and 600 msec post stimulus onset, to infrequent auditory events when those events were attended. Given it's latency range, the authors interpreted this large peak to be associated with the scalp-recorded P3b component, though the nature of the potential fields found in the hippocampus indicate that they would be very difficult to observe from outside the head. Further work with similar auditory oddball-like tasks demonstrated a series of endogenous components

to rare events in trains of acoustic stimuli, beginning with an auditory-specific mismatch negativity between 100 and 200 msec, localized to the superior temporal plane followed by other auditory-specific components (included the P2, N2a and N2b) and diffuse supramodal P3a-like components located in anterior and posterior cingulate gyrus, supramarginal gyrus, temporal pole, middle temporal gyrus, parahippocampal gyri, fusiform gyrus, and widespread frontal areas (Halgren et al., 1995a, 1995b; Baudena et al., 1995). Finally, large monophasic waveforms reminiscent of the scalp-recorded P3b were found in more circumscribed areas including the hippocampus and parahippocampal areas (though these likely do not propagate to the scalp), ventrolateral prefrontal cortex, and posterior superior parietal cortex (Halgren et al., 1995a, 1995b; Baudena et al., 1995; Halgren et al., 1998). More recent work has expanded on the auditory change detection paradigm with intraoperative electrocorticography (Bekinschtein et al., 2009) to examine high-gamma activity (Edwards et al., 2005) – thought to be a marker of functionally-active cortex, correlating with both summed synaptic activity and multiunit firing (Steinschneider et al., 2008) – in response to deviant tones.

High-gamma activity has been a topic of much recent interest, particular in intracranial studies given the fact that it is much more readily observable when compared with noninvasive methods such as EEG and MEG. Crone et al. (2001) was the first to examine high-gamma activity in response to acoustic stimuli (including tones and phonemes) and demonstrated greater STG activity during discrimination of phonemes than words (Crone et al., 2001). Since then, several studies have examined high gamma activity in response to several types of acoustic stimuli including tones (Edwards et al., 2005), speech sounds (Steinschneider et al., 1999, 2011; Nourski et al., 2009; Edwards et al., 2009, 2010; Chang et al., 2010a, 2010b), click trains (Brugge et al., 2008, 2009), pitch (Griffiths et al., 2010), and long-duration complex auditory

scenes (Bidet-Caulet et al., 2007a). Other studies have shown that high gamma activity is a strong predictor of the fMRI BOLD response in the auditory cortex (Mukamel et al., 2005). A comprehensive review on the topic of event-related gamma-band activity across a wide range of cognitive tasks has recently been published (Jerbi et al., 2009).

The only intracranial studies which were explicitly focused on mechanisms of auditory scene analysis were those from Bidet-Caulet and colleagues (Bidet-Caulet et al., 2007a, 2007b). In both studies, listeners were presented with complex auditory scenes consisting of three amplitude-modulated pure tones at different carrier and modulation frequencies. Depending on the stimulus onset asynchrony (SOA) of the different tones, they could be perceived as grouped (in the case of zero SOA) or segregated (in the case of non-zero SOA). In one study, listeners' attention was directed away from the stimuli with an orthogonal target detection task in order to examine automatic mechanisms of scene analysis (Bidet-Caulet et al., 2007a). Comparing identical acoustic stimuli with differing preceding contexts, the authors reported that transient evoked and sustained broadband (gamma and high-gamma range) activity was enhanced when the context lead to the test stimulus being perceived as a single stream vs. two segregated streams. However, another plausible interpretation is that a larger change in the acoustics of the stimulus occurred for the 1-stream stimulus vs. the 2-stream stimulus, despite the fact that they were subsequently identical or how they were then perceptually organized. In the other study which used very similar stimuli, listeners' explicitly attended to one stream or the other; selective attention enhanced both steady-state and transient evoked responses in medial Heschl's gyrus and lateral STG, respectively (Bidet-Caulet et al., 2007b).

1.3 Organization of dissertation research

The principal aim of this thesis was to explore how the brain performs various ASA functions. To this end, we combined electrical recordings made directly from the cortical surface of neurosurgical patients with what are now considered to be classic behavioral paradigms in the ASA literature, including (i) a simple auditory streaming paradigm consisting of alternating tone triplets and (ii) an informational masking paradigm consisting of a randomly-varying background masker stimulus and a regularly-repeating target stream. These two behavioral paradigms are extremely well studied psychophysically, each having a literature dating back at least 40 years. This affords us the opportunity to compare psychophysical results from our patient population (note: the fact that they are neurosurgical patients is entirely tangential to the aims of this thesis) with that of a large body of literature, providing inherent controls for the fact that the subjects used in our study to examine the neural correlates of ASA are not neurologically normal. Furthermore, by comparing neural responses to acoustic stimuli that are physically identical but perceived differently, we were able to dissociate brain activity related to the processing of stimulus information vs brain activity directly related to the subjective experience of a listener.

Chapter 2 outlines the general methods used, including details about the idiosyncratic nature of attempting to perform well-controlled cognitive experiments in a clinical setting. Chapter 3 describes an ad hoc method devised for accurately localizing intracranial electrodes with respect to each individual's neuroanatomy. Chapter 4 reports the results from the main experiment in the thesis, demonstrating that the brain areas engaged during the classic auditory streaming paradigm are much more widespread than has been previously shown. Surprisingly, there were few consistent correlates of perceptual organization, *per se*, in the absence of

differences in the physical stimulus. Chapter 5 builds on chapter 4 to examine brain responses over the entire duration of each tone sequence rather than focusing exclusively on triplet-locked responses. Several timescales of acoustic segmentation were exhibited in the responses across widespread brain areas. Generally, areas near the posterior superior temporal gyrus – the putative location of secondary or tertiary unimodal auditory cortex – showed responses that persisted throughout the duration of the acoustic stimulation, while areas in frontal and parietal cortex – putative supramodal areas – showed responses only at the onset or offset of sound. Chapter 6 describes the results from experiments using the multi-tone informational masking paradigm. Evoked potentials and high-gamma activity showed strong correlations with whether or not the target stream was detected (i.e., with the subjective perceptual experience of the listener). Chapter 7 discusses the results with reference to current theories of the neural mechanisms of auditory perceptual organization.

1.4 References

Anstis, S., and Saida, S. (1985). Adaptation to auditory streaming of frequency modulated tones. *Journal of Experimental Psychology. Human Perception and Performance* 11, 257-271.

Baudena, P., Halgren, E., Heit, G., and Clarke, J. M. (1995). Intracerebral potentials to rare target and distractor auditory and visual stimuli. III. Frontal cortex. *Electroencephalogr Clin Neurophysiol* 94, 251-264.

Bee, M. A., and Klump, G. M. (2005). Auditory stream segregation in the songbird forebrain: effects of time intervals on responses to interleaved tone sequences. *Brain Behav. Evol* 66, 197-214.

Bee, M. A., and Klump, G. M. (2004). Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. *J. Neurophysiol* 92, 1088-1104.

Bee, M. A., and Micheyl, C. (2008). The cocktail party problem: what is it? How can it be solved? And why should animal behaviorists study it? *J Comp Psychol* 122, 235-251.

Bee, M. A., Micheyl, C., Oxenham, A. J., and Klump, G. M. (2010). Neural adaptation to tone sequences in the songbird forebrain: patterns, determinants, and relation to the build-up of auditory streaming. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol* 196, 543-557.

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., and Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proc. Natl. Acad. Sci. U.S.A* 106, 1672-1677.

Bidet-Caulet, A., and Bertrand, O. (2009). Neurophysiological mechanisms involved in auditory perceptual organization. *Front Neurosci* 3, 182-191.

Bidet-Caulet, A., Fischer, C., Bauchet, F., Aguera, P.-E., and Bertrand, O. (2007a). Neural substrate of concurrent sound perception: direct electrophysiological recordings from human auditory cortex. *Front Hum Neurosci* 1, 5.

Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P.-E., Giard, M.-H., and Bertrand, O. (2007b). Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. *J. Neurosci* 27, 9252-9261.

Bitterman, Y., Mukamel, R., Malach, R., Fried, I., and Nelken, I. (2008). Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. *Nature* 451, 197-201.

Blake R, Logothetis NK (2002) Visual competition. Nat Rev Neurosci 3:13-21.

Bregman, A. S. (1994). *Auditory scene analysis: the perceptual organization of sound.* MIT Press.

Brugge, J. F., Nourski, K. V., Oya, H., Reale, R. A., Kawasaki, H., Steinschneider, M., and Howard, M. A., 3rd (2009). Coding of repetitive transients by auditory cortex on Heschl's gyrus. *J. Neurophysiol* 102, 2358-2374.

Brugge, J. F., Volkov, I. O., Garell, P. C., Reale, R. A., and Howard, M. A. (2003). Functional connections between auditory cortex on Heschl's gyrus and on the lateral superior temporal gyrus in humans. *J. Neurophysiol* 90, 3750-3763.

Brugge, J. F., Volkov, I. O., Oya, H., Kawasaki, H., Reale, R. A., Fenoy, A., Steinschneider, M., and Howard, M. A. (2008). Functional localization of auditory cortical fields of human: click-train stimulation. *Hear. Res* 238, 12-24.

Cariani PA (2011) Towards a theory of information processing in the auditory cortex In The Auditory Cortex Springer.

Carlyon, R. P., Cusack, R., Foxton, J. M., and Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *J Exp Psychol Hum Percept Perform* 27, 115-127.

Carlyon, R. P. (2004). How the brain separates sounds. *Trends Cogn. Sci. (Regul. Ed.)* 8, 465-471.

Chang, E. F., Edwards, E., Nagarajan, S. S., Fogelson, N., Dalal, S. S., Canolty, R. T., Kirsch, H. E., Barbaro, N. M., and Knight, R. T. (2010a). Cortical Spatio-temporal Dynamics Underlying Phonological Target Detection in Humans. *J Cogn Neurosci*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20465359 [Accessed September 16, 2010].

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., and Knight, R. T. (2010b). Categorical speech representation in human superior temporal gyrus. *Nat Neurosci*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20890293 [Accessed October 26, 2010].

Cherry EC (1953) Some Experiments on the Recognition of Speech, with One and with Two Ears. J Acoust Soc Am 25:975.

Crone, N. E., Boatman, D., Gordon, B., and Hao, L. (2001). Induced electrocorticographic gamma activity during auditory perception. Brazier Award-winning article, 2001. *Clin Neurophysiol* 112, 565-582.

Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *J Cogn Neurosci* 17, 641-651.

Darwin, C. J. (1997). Auditory grouping. *Trends Cogn. Sci. (Regul. Ed.)* 1, 327-333.

Deike, S., Gaschler-Markefski, B., Brechmann, A., and Scheich, H. (2004). Auditory stream segregation relying on timbre involves left auditory cortex. *Neuroreport* 15, 1511-1514.

Deike, S., Scheich, H., and Brechmann, A. (2010). Active stream segregation specifically involves the left human auditory cortex. *Hear. Res* 265, 30-37.

Del Cul A, Dehaene S, Reyes P, Bravo E, Slachevsky A (2009) Causal role of prefrontal cortex in the threshold for access to consciousness. Brain 132:2531-2540.

Del Cul A, Baillet S, Dehaene S (2007) Brain dynamics underlying the nonlinear threshold for access to consciousness. PLoS Biol 5:e260.

Edwards, E., Nagarajan, S. S., Dalal, S. S., Canolty, R. T., Kirsch, H. E., Barbaro, N. M., and Knight, R. T. (2010). Spatiotemporal imaging of cortical activation during verb generation and picture naming. *Neuroimage* 50, 291-301.

Edwards, E., Soltani, M., Deouell, L. Y., Berger, M. S., and Knight, R. T. (2005). High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. *J. Neurophysiol* 94, 4269-4280.

Edwards, E., Soltani, M., Kim, W., Dalal, S. S., Nagarajan, S. S., Berger, M. S., and Knight, R. T. (2009). Comparison of time-frequency responses and the event-related potential to auditory speech stimuli in human cortex. *J. Neurophysiol* 102, 377-386.

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J., and Shamma, S. A. (2009a). Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron* 61, 317-329.

Elhilali, M., Xiang, J., Shamma, S. A., and Simon, J. Z. (2009b). Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. *PLoS Biol* 7, e1000129.

Engel, A. K., Moll, C. K. E., Fried, I., and Ojemann, G. A. (2005). Invasive recordings from the human brain: clinical insights and beyond. *Nat. Rev. Neurosci* 6, 35-47.

Fay, R. R. (1998). Auditory stream segregation in goldfish (Carassius auratus). *Hear. Res* 120, 69-76.

Fay, R. R. (2000). Spectral contrasts underlying auditory stream segregation in goldfish (Carassius auratus). *J. Assoc. Res. Otolaryngol* 1, 120-128.

Fishman, Y. I., Reser, D. H., Arezzo, J. C., and Steinschneider, M. (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear. Res* 151, 167-187.

Fishman, Y. I., Arezzo, J. C., and Steinschneider, M. (2004). Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *J. Acoust. Soc. Am* 116, 1656-1670.

Gaillard R, Dehaene S, Adam C, Clémenceau S, Hasboun D, Baulac M, Cohen L, Naccache L

(2009) Converging intracranial markers of conscious access. PLoS Biol 7:e61.

Griffiths, T. D., Kumar, S., Sedley, W., Nourski, K. V., Kawasaki, H., Oya, H., Patterson, R. D., Brugge, J. F., and Howard, M. A. (2010). Direct recordings of pitch responses from human auditory cortex. *Curr. Biol* 20, 1128-1132.

Gutschalk, A., Micheyl, C., and Oxenham, A. J. (2008). Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biol* 6, e138.

Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., and Oxenham, A. J. (2005). Neuromagnetic correlates of streaming in human auditory cortex. *J. Neurosci* 25, 5382-5388.

Gutschalk, A., Oxenham, A. J., Micheyl, C., Wilson, E. C., and Melcher, J. R. (2007). Human cortical activity during streaming without spectral cues suggests a general neural substrate for auditory stream segregation. *J. Neurosci* 27, 13074-13081.

Halgren, E., Baudena, P., Clarke, J. M., Heit, G., Liégeois, C., Chauvel, P., and Musolino, A. (1995a). Intracerebral potentials to rare target and distractor auditory and visual stimuli. I. Superior temporal plane and parietal lobe. *Electroencephalogr Clin Neurophysiol* 94, 191-220.

Halgren, E., Baudena, P., Clarke, J. M., Heit, G., Marinkovic, K., Devaux, B., Vignal, J. P., and Biraben, A. (1995b). Intracerebral potentials to rare target and distractor auditory and visual stimuli. II. Medial, lateral and posterior temporal lobe. *Electroencephalogr Clin Neurophysiol* 94, 229-250.

Halgren, E., Marinkovic, K., and Chauvel, P. (1998). Generators of the late cognitive potentials in auditory and visual oddball tasks. *Electroencephalogr Clin Neurophysiol* 106, 156-164.

Halgren, E., Squires, N. K., Wilson, C. L., Rohrbaugh, J. W., Babb, T. L., and Crandall, P. H. (1980). Endogenous potentials generated in the human hippocampal formation and amygdala by infrequent events. *Science* 210, 803-805.

Haynes J-D, Rees G (2006) Decoding mental states from brain activity in humans. Nat Rev Neurosci 7:523-534.

Handel S (1993) Listening: An Introduction to the Perception of Auditory Events. MIT Press.

Hillyard, S. A., Squires, K. C., Bauer, J. W., and Lindsay, P. H. (1971). Evoked potential correlates of auditory signal detection. *Science* 172, 1357-1360.

Howard, M. A., 3rd, Volkov, I. O., Abbas, P. J., Damasio, H., Ollendieck, M. C., and Granner, M. A. (1996). A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. *Brain Res* 724, 260-264.

Howard, M. A., Volkov, I. O., Mirsky, R., Garell, P. C., Noh, M. D., Granner, M., Damasio, H., Steinschneider, M., Reale, R. A., Hind, J. E., et al. (2000). Auditory cortex on the human posterior superior temporal gyrus. *J. Comp. Neurol* 416, 79-92.

Itatani, N., and Klump, G. M. (2009). Auditory streaming of amplitude-modulated sounds in the songbird forebrain. *J. Neurophysiol* 101, 3212-3225.

Itatani, N., and Klump, G. M. (2010). Neural correlates of auditory streaming of harmonic complex sounds with different phase relations in the songbird forebrain. *J. Neurophysiol.* Available at: http://www.ncbi.nlm.nih.gov/pubmed/21068270 [Accessed December 23, 2010].

Jerbi, K., Ossandón, T., Hamamé, C. M., Senova, S., Dalal, S. S., Jung, J., Minotti, L., Bertrand, O., Berthoz, A., Kahane, P., et al. (2009). Task-related gamma-band dynamics from an intracerebral perspective: review and implications for surface EEG and MEG. *Hum Brain Mapp* 30, 1758-1771.

Kanwal, J. S., Medvedev, A. V., and Micheyl, C. (2003). Neurodynamics for auditory stream segregation: tracking sounds in the mustached bat's natural environment. *Network* 14, 413-435.

Kidd, G., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (1994). Reducing informational masking by sound segregation. *J. Acoust. Soc. Am* 95, 3475-3480.

Kidd, G., Mason, C. R., and Richards, V. M. (2003). Multiple bursts, multiple looks, and stream coherence in the release from informational masking. *J. Acoust. Soc. Am* 114, 2835-2845.

Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (2008). "Informational Masking," in Auditory Perception of Sound Sources (New York, NY: Springer), 143-190.

Koch C (2009) The Neurobiology of Consciousness In The Cognitive Neurosciences Cambridge, MA, USA: MIT Press, p. 1137-1150.

Koffka K (1935) Principles of Gestalt Psychology. Routledge.

Köhler W (1947) Gestalt psychology: An introduction to new concepts in modern psychology. Liveright Publishing Corporation.

Kondo, H. M., and Kashino, M. (2009). Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *J. Neurosci* 29, 12695-12701.

Lemus, L., Hernández, A., and Romo, R. (2009a). Neural codes for perceptual discrimination of acoustic flutter in the primate auditory cortex. *Proc. Natl. Acad. Sci. U.S.A* 106, 9471-9476.

Lemus, L., Hernández, A., and Romo, R. (2009b). Neural encoding of auditory discrimination in ventral premotor cortex. *Proc. Natl. Acad. Sci. U.S.A* 106, 14640-14645.

Leopold, and Logothetis (1999). Multistable phenomena: changing views in perception. *Trends Cogn. Sci. (Regul. Ed.)* 3, 254-264.

Libedinsky C, Livingstone M (2011) Role of prefrontal cortex in conscious visual perception. J Neurosci 31:64-69.

Liégeois-Chauvel, C., Musolino, A., and Chauvel, P. (1991). Localization of the primary auditory area in man. *Brain* 114 (Pt 1A), 139-151.

Liégeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., and Chauvel, P. (1994). Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. *Electroencephalogr Clin Neurophysiol* 92, 204-214.

Logothetis, N. K. (1998). Single units and conscious vision. *Philos. Trans. R. Soc. Lond., B, Biol. Sci* 353, 1801-1818.

Meyer LB (1956) Emotion and Meaning in Music. New York, NY: Academic Press.

Micheyl, C., Shamma, S., and Oxenham, A. J. (2007a). "Hearing out repeating elements in randomly varying multitone sequences: a case of streaming," in Hearing - from basic research to application (Berlin: Springer), 267-274.

Micheyl, C., Carlyon, R. P., Gutschalk, A., Melcher, J. R., Oxenham, A. J., Rauschecker, J. P., Tian, B., and Courtenay Wilson, E. (2007b). The role of auditory cortex in the formation of auditory streams. *Hear. Res* 229, 116-131.

Micheyl, C., Tian, B., Carlyon, R. P., and Rauschecker, J. P. (2005). Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron* 48, 139-148.

Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., and Malach, R. (2005). Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. *Science* 309, 951-954.

Necker, L. A. (1832). Observations on some remarkable optical phaenomena seen in Switzerland; and on an optical phaenomenon which occurs on viewing a figure of a crystal or geometrical solid. *London and Edinburgh Philosophical Magazine and Journal of Science* 1, 329-337.

Neff, D. L., and Green, D. M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. *Percept Psychophys* 41, 409-415.

Neff, D. L., Dethlefs, T. M., and Jesteadt, W. (1993). Informational masking for multicomponent maskers with spectral gaps. *J. Acoust. Soc. Am* 94, 3112-3126.

van Noorden, L. (1975). Temporal coherence in the perception of tone sequences.

Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., Howard, M. A.,

3rd, and Brugge, J. F. (2009). Temporal envelope of time-compressed speech represented in the human auditory cortex. *J. Neurosci* 29, 15564-15574.

Overath, T., Cusack, R., Kumar, S., von Kriegstein, K., Warren, J. D., Grube, M., Carlyon, R. P., and Griffiths, T. D. (2007). An information theoretic characterisation of auditory encoding. *PLoS Biol* 5, e288.

Oya, H., Poon, P. W. F., Brugge, J. F., Reale, R. A., Kawasaki, H., Volkov, I. O., and Howard, M. A., 3rd (2007). Functional connections between auditory cortical fields in humans revealed by Granger causality analysis of intra-cranial evoked potentials to sounds: comparison of two methods. *BioSystems* 89, 198-207.

Penfield, W., and Jasper, H. H. (1954). *Epilepsy and the functional anatomy of the human brain.* Little, Brown.

Pressnitzer, D., and Hupé, J.-M. (2006). Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Curr. Biol* 16, 1351-1357.

Pressnitzer, D., Sayles, M., Micheyl, C., and Winter, I. M. (2008). Perceptual organization of sound begins in the auditory periphery. *Curr. Biol* 18, 1124-1128.

Rees G (2007) Neural correlates of the contents of visual awareness in humans. Philos Trans R Soc Lond , B, Biol Sci 362:877-886.

Rees G (2009) Visual Awareness In The Cognitive Neurosciences Cambridge, MA, USA: MIT Press, p. 1151-1164.

Riecke, L., Esposito, F., Bonte, M., and Formisano, E. (2009). Hearing illusory sounds in noise: the timing of sensory-perceptual transformations in auditory cortex. *Neuron* 64, 550-561.

Riecke, L., van Opstal, A. J., Goebel, R., and Formisano, E. (2007). Hearing illusory sounds in noise: sensory-perceptual transformations in primary auditory cortex. *J. Neurosci* 27, 12684-12689.

Schadwinkel, S., and Gutschalk, A. (2010a). Activity Associated with Stream Segregation in Human Auditory Cortex is Similar for Spatial and Pitch Cues. *Cereb Cortex*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20237241 [Accessed September 14, 2010].

Schadwinkel, S., and Gutschalk, A. (2010b). Functional dissociation of transient and sustained fMRI BOLD components in human auditory cortex revealed with a streaming paradigm based on interaural time differences. *Eur. J. Neurosci* 32, 1970-1978.

Schadwinkel, S., and Gutschalk, A. (2011). Transient bold activity locked to perceptual reversals of auditory streaming in human auditory cortex and inferior colliculus. *J. Neurophysiol* 105, 1977-1983.

Schul, J., and Sheridan, R. A. (2006). Auditory stream segregation in an insect. *Neuroscience*

138, 1-4.

Shamma, S. A., and Micheyl, C. (2010). Behind the scenes of auditory perception. *Curr. Opin. Neurobiol* 20, 361-366.

Shamma, S. A., Elhilali, M., and Micheyl, C. (2010). Temporal coherence and attention in auditory scene analysis. *Trends Neurosci.* Available at: http://www.ncbi.nlm.nih.gov/pubmed/21196054 [Accessed January 27, 2011].

Snyder, J. S., Alain, C., and Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *J Cogn Neurosci* 18, 1-13.

Snyder, J. S., and Alain, C. (2007a). Sequential auditory scene analysis is preserved in normal aging adults. *Cereb. Cortex* 17, 501-512.

Snyder, J. S., and Alain, C. (2007b). Toward a neurophysiological theory of auditory stream segregation. *Psychol Bull* 133, 780-799.

Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., and Howard, M. A. (1999). Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *J. Neurophysiol* 82, 2346-2357.

Steinschneider, M., Fishman, Y. I., and Arezzo, J. C. (2008). Spectrotemporal analysis of evoked and induced electroencephalographic responses in primary auditory cortex (A1) of the awake monkey. *Cereb. Cortex* 18, 610-625.

Steinschneider, M., Nourski, K. V., Kawasaki, H., Oya, H., Brugge, J. F., and Howard, M. A., 3rd (2011). Intracranial Study of Speech-Elicited Activity on the Human Posterolateral Superior Temporal Gyrus. *Cereb Cortex.* Available at: http://www.ncbi.nlm.nih.gov/pubmed/21368087 [Accessed July 14, 2011].

Sterzer, P., Kleinschmidt, A., and Rees, G. (2009). The neural bases of multistable perception. *Trends Cogn. Sci. (Regul. Ed.)* 13, 310-318.

Sussman, E., Ritter, W., and Vaughan, H. G. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology* 36, 22-34.

Wilson, E. C., Melcher, J. R., Micheyl, C., Gutschalk, A., and Oxenham, A. J. (2007). Cortical FMRI activation to sequences of tones alternating in frequency: relationship to perceived rate and streaming. *J. Neurophysiol* 97, 2230-2238.

Zylberberg, A., Fernández Slezak, D., Roelfsema, P. R., Dehaene, S., and Sigman, M. (2010). The brain's router: a cortical network model of serial processing in the primate brain. *PLoS Comput. Biol* 6, e1000765.

Chapter 2: General Methods

2.1 Participants

Participants in the experiments described in this thesis were patients with drug-resistant partial epilepsy undergoing clinically-indicated invasive monitoring for localization of the seizure focus prior to its surgical removal. In total, we recruited 27 such patients (Table 1) from three different epilepsy centers – Brigham and Women's Hospital (N=3), Massachusetts General Hospital (N=18), and New York University Medical Center (N=6) – to participate in two different experiments described below. Each patient was implanted with either (i) sub-dural disc electrode arrays resting directly on the cortical surface, (ii) linear penetrating depth arrays in order to directly sample medial structures such as the hippocampus or cingulate cortex, or both (*Adtech Medical, Racine, WI, USA*). In every case, the decision to implant intracranial electrodes as well as where to implant them was made solely on clinical grounds; the experimental protocols described herein were not considered. In total, collated across all patients and all behavioral tasks, 2,474 individual sites were sampled, providing coverage of nearly every area of the brain.

Patient	Sex	Age	Hemisphere of Implant	Type of electrodes implanted	Number of contacts	Seizure Focus	Task	Hospital
S1	M	21	Left	Sub-dural grid/strip	72	Left mesial temporal	Auditory streaming	BWH
S2	M	27	Bilateral	Sub-dural grid/strip	64	Unclear	Auditory streaming Informational masking	BWH
S3	F	31	Right	Linear depth	34	Right mesial temporal	Auditory streaming	MGH
S4	F	55	Bilateral	Linear depth	56	Left temporal neocortex	Auditory streaming	MGH
S5	M	45	Bilateral	Linear depth	64	Multifocal	Auditory streaming	MGH
S6	M	29	Bilateral	Linear depth	64	Right mesial temporal	Auditory streaming	MGH
S7	F	58	Bilateral	Linear depth	78	Multifocal	Auditory streaming	MGH
S8	F	45	Left	Linear depth	40	Left mesial temporal	Auditory streaming	MGH
S9	M	28	Bilateral	Linear depth	68	Multifocal	Auditory streaming	MGH
S10	M	22	Left	Sub-dural grid/strip Linear depth	120	Left cingulate cortex	Auditory streaming	MGH
S11	F	52	Left	Sub-dural grid/strip Linear depth	100	Left anterior temporal	Auditory streaming	MGH
S12	F	65	Bilateral	Linear depth	68	Multifocal	Auditory streaming	MGH
S13	M	19	Right	Sub-dural grid/strip Linear depth	116	Right parietal neocortex	Auditory streaming	MGH
S14	M	29	Left	Sub-dural grid/strip	120	Left anterior temporal	Auditory streaming	MGH
S15	M	23	Bilateral	Linear depth	80	Left frontal neocortex	Auditory streaming	MGH
S16	M	25	Left	Sub-dural grid/strip Linear depth	96	Multifocal	Auditory streaming	MGH
S17	F	42	Bilateral	Linear depth	80	Unclear	Auditory streaming	MGH
S18	F	?	Right	Sub-dural grid/strip	162	?	Auditory streaming	NYU
S19	F	42	Bilateral	Sub-dural grid/strip	182	Right mesial temporal	Auditory streaming	NYU
S20	F	38	Bilateral	Sub-dural grid/strip	170	Left mesial temporal	Auditory streaming	NYU
S21	M	37	Left	Sub-dural grid/strip	88	Left mesial temporal	Auditory streaming	NYU
S22	F	?	Left	Sub-dural grid/strip	145	?	Auditory streaming	NYU
S23	M	25	Left	Sub-dural grid/strip	133	Left mesial temporal	Auditory streaming	NYU
S24	M	30	Right	Sub-dural grid/strip	34	Unclear	Informational masking	BWH
S25	F	19	Left	Sub-dural grid/strip	60	Unclear	Informational masking	MGH
S26	M	34	Bilateral	Linear depth	80	Right mesial temporal	Informational masking	MGH
S27	M	31	Left	Sub-dural grid/strip	100	Left mesial temporal	Informational masking	MGH

Table 2.1. Patient information.

2.2 Stimuli and behavioral tasks

Two behavioral paradigms were used. The first was a classic auditory-streaming paradigm that has been used in many psychoacoustic studies dating to the early 1970s. The second was a modified version of a classic informational-masking paradigm. Both paradigms involve listening to sequences of pure tones and making simple, subjective judgments about them. What both paradigms have in common is that they allow for the comparison of neural activity between conditions in which physical stimuli remain constant but subjective perception differs greatly. Such a comparison allows for the examination of the neural mechanisms which are directly involved in perception, eliminating confounds of changing physical stimuli which often covary with perception. Each paradigm is described in more detail below as well as in the relevant respective chapters (4 and 6) of this thesis.

2.2.1 Auditory streaming

The first experimental paradigm used was a classical auditory-streaming paradigm devised in the early 1970s (van Noorden, 1975) and elaborated upon by Bregman (Bregman, 1994). The stimuli used in this paradigm consist of sequences of alternating pure tones of the form ABA-ABA-..., where A and B denote short-duration pure tones of different frequencies and the dash denotes a silent period (see Figures 1.2 and 4.1). Most often, the listener's task is to listen to each stimulus sequence and indicate by button press whether they perceive the tones as grouped or segregated; that is, whether they hear a single "stream" comprised of both A and B tones in the form a "galloping" rhythm or two parallel isochronous streams, one comprised of the A tones at a fast (relative) rate and another comprised of the B tones at a slow (relative) rate. Manipulations of precise stimulus parameters such as the frequency separation (ΔF) between A and B tones or overall presentation rate can engender the percept of one (small ΔF or slow

presentation rates) or two (large ΔF or fast presentation rates) streams (see Figure 1.3). For the purposes of this thesis, we held the presentation rate constant and varied ΔF between 0 and 100 % or, in musical terms, between 0 and 12 semitones. Sequences with a 0-semitone ΔF should, by definition, always be heard as a single stream while sequences with a 12-semitone ΔF are almost always heard as two streams. Our participants were asked to indicate, at the end of each sequence, whether they were hearing one or two streams at the end of each sequence.

2.2.2 Informational masking

The second paradigm used was a modified version of another classical auditory task, the multi-tone masking paradigm. In the classic version of this task, listeners are presented with a sequence of inharmonic complex tones separated by short intervals and asked to report whether there was a single component which repeated throughout the sequence (Neff and Green, 1987). Recently, a modified version of this paradigm was devised for use with AER experiments in which the target tones repeat at a regular interval but the other components of the sequence are jittered in time and frequency (Gutschalk et al., 2008) (see Figures 1.4 and 6.1). The task of the listener is to indicate as soon as they begin to hear the regularly-repeating target stream. What makes this task difficult is the fact that the frequency of the repeating target tone is unknown at the start of each sequence. The jittered nature of the masker tones allows for the isolation of time-locked activity elicited by the targets. Our participants listened to approximately 7-second long sequences comprised of a randomly varying multi-tone masker stimulus and, on some trials, a regularly-repeating target tone (at one of six predetermined frequencies) and indicated by button press as soon as they began to hear the target stream "pop out" from the background. The masker-only trials served as catch trials in order to measure the participants' behavioral ability to detect the targets accurately.

2.3 Data acquisition

2.3.1 Imaging

At each center (MGH, BWH, and NYU), preoperative high-resolution T1-weighted MRI was acquired for each patient prior to their implant procedure. All scans were acquired by the radiology department at each of the three centers. The precise scan sequence differed depending on hospital site, but were either sagittal MPRAGE (on Siemens scanners at MGH), SPGR (on GE scanners at BWH and NYU), or similar. For patients at MGH and BWH (NYU), postoperative CT (MRI) scans were acquired in order to localize electrodes and verify the absence of hemorrhage and/or mass effect. In both postoperative CT and postoperative MRI, electrodes produce a visible artifact which was subsequently used to localize electrodes to the individual's reconstructed pial surface (computed from the preoperative MRI scans using Freesurfer).

2.3.2 Behavioral interface

All stimuli were presented via Etymotic ER-2 insert earphones at a comfortable listening level from the on-board sound card of a portable laptop computer running Presentation software (*Neurobehavioral Systems Inc., Albany, CA*). Experiments were run at the patient's bedside in their hospital room. The laptop computer was placed on a table directly in front of the patient in order for the patient to view experimental instructions and a fixation cross presented on the screen. Patients entered their responses by button press via a multicolored button box (*Cedrus Corporation, San Pedro, CA*) interfaced with Presentation via USB. Efforts were made to ensure that, during each experiment, the patient's hospital room was as quiet as possible, including turning off non-essential monitoring equipment, closing the hallway door, and turning off the

patient's television. However, the nature of the hospital environment precludes ideal conditions for psychoacoustic experiments.

2.3.3 Intracranial electroencephalography

The iEEG for patients at MGH and BWH was recorded by each hospital's clinical acquisition system (*XLTEK, Natus Medical Corporation, Oakville, Ontario, Canada*) which is capable of recording 128 channels simultaneously. Data were sampled at either 250 or 500 Hz and digitized with 32-bit resolution. All data were referenced to an either an inverted disc electrode placed against the inner skull table or a C2 reference placed on the back of the neck.

The iEEG for patients at NYU was recorded by a customized acquisition system sampling at 30 kHz and 64-bit resolution (Bijan Pesaran et al., personal communication). All data were subsequently downsampled to 500 Hz for analysis. A screw bolted to the skull was used as the reference electrode.

2.4 Data analysis

2.4.1 Imaging and electrode localization

Pre- and post-operative imaging data were analyzed using a combination of open-source packages (Freesurfer, http://surfer.nmr.mgh.harvard.edu/fswiki) and custom in-house software written in the MATLAB programming environment (*The Mathworks Inc., Natick, MA, USA*). High-resolution T1-weighted preoperative MRI volumes were used to render a three-dimensional cortical surface for each patient in the Freesurfer environment. The same MRI volumes were simultaneously coregistered with either postoperative CT (MGH and BWH) or postoperative MRI (NYU). Individual RAS coordinates in each patients native anatomical space were obtained for each electrode by visual inspection of the postoperative CT or MRI volume. These coordinates were then overlaid onto the patient's individual pial surface (in the case of grid electrodes) or onto a single slice of the preoperative MRI volume (in the case of depth electrodes). For group analyses, electrodes across patients were projected into the same anatomical space using either spherical (for grid electrodes) or volumetric (for depth electrodes) registration methods These methods are described in more detail in Chapter 3 of this thesis.

2.4.2 Auditory streaming task

2.4.2.1 Behavior data

Behavioral responses were recorded at the end of each stimulus sequence by asking the listener to report whether they were hearing one or two streams at the end of the sequence. Early on in the study, attempts were made to get participants to continuously update their percept throughout the course of each sequence, but this proved difficult for the patients to understand, so behavioral responses were subsequently only entered at the end of each 6.5-10 second long sequence. For each ΔF condition, the proportion of trials that were heard by the listener as two

streams was calculated and plotted as a function of ΔF. The large previous behavioral literature using this paradigm provides an inherent behavioral check indicating whether the patient understood the task correctly, an important point given the fact that our participants were neurological patients. If the patients understand the task, the proportion of trials heard as two streams should monotonically increase as ΔF increases.

2.4.2.2 Intracranial EEG

After appropriate preprocessing (see Chapter 4 of this thesis), the iEEG was epoched relative to (i) the onset of each stimulus sequence and (ii) the onset of each ABA- triplet in the stimulus sequence. Trials were binned according to either ΔF or percept (one vs. two streams). Statistical effects of ΔF and percept in both the EP and high-gamma power amplitude were computed using non-parametric cluster-based Monte-Carlo estimates of permutation distributions (Chapter 4). Waveform differences across conditions and across patients were quantified using a dissimilarity index in order to account for the highly-variable waveforms we observed from one patient and one electrode to the next. A further analysis was carried out on the sequence-length epochs in order to quantify the time course on which a particular electrode was response (Chapter 5). A waveshape index was computed whose value was close to -1 if a given site responded in a sustained or steady-state manner throughout the stimulus and close to 1 if the site responded only at the onset or offset of the stimuli.

2.4.3 Informational masking task

2.4.3.1 Behavioral data

As in the auditory streaming task, behavioral responses were entered with a button box. Participants listened to sequences of jittered multi-tone sequences and pressed a button to indicate the moment at which they began to hear a regularly-repeating target stream of tones.

Button presses for target stimuli that were indeed present were considered "hits" while button presses in the absence of target stimuli were considered "false alarms." Since the task was to detect *repeating* targets, it was assumed that, at minimum, the two target tones preceding a button press were also detected. Behavioral curves were constructed to show the percentage of hits and false alarms as a function of time since the sequence began. The inherent behavioral check in this experiment is the ratio of hits to false alarms. Ideally, for the purposes of comparing brain activity to detected vs. undetected target tones, the hit rate should be near 50% while the false alarm rate should be close to 0%. This would indicate that the patient understood the task, but was not able to perform near ceiling.

2.4.3.2 Intracranial EEG

After appropriate preprocessing, the iEEG was epoched relative to the onset of target tones and binned according to whether the target tones were detected or undetected. In two control conditions, the iEEG was epoched relative to target tones in isolation and relative to virtual target tones – had they been present – in the masker-only stimulus. The targets-only condition provided an EP and high-gamma template against which to compare the response to the detected target tones and the masker-only condition provided a template against which to compare the response to the undetected targets. Theoretically, the response to the virtual targets in the masker-only condition should be flat since the time-locking was respect to randomly-presented tones in the absence of a regularly-repeating target. The main comparison of interest was detected vs. undetected targets. Statistical effects of detection in both EP amplitude and high-gamma activity were computed using non-parametric cluster-based Monte-Carlo estimates of permutation distributions (Chapter 6).

2.5 References

Bregman, A. S. (1994). *Auditory scene analysis: the perceptual organization of sound.* MIT Press.

Gutschalk, A., Micheyl, C., and Oxenham, A. J. (2008). Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biol* 6, e138.

Neff, D. L., and Green, D. M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. *Percept Psychophys* 41, 409-415.

van Noorden, L. (1975). Temporal coherence in the perception of tone sequences.

Chapter 3: Individualized Localization and Cortical Surface-Based Registration of Intracranial Electrodes

*This chapter was submitted for publication to *Neuroimage* on July 5, 2011

Andrew R. Dykstra[1,3], Alexander M. Chan[2,3], Brian T. Quinn[4], Rodrigo Zepeda[3], Corey K. Keller[3], Justine E. Cormier[3], Joseph R. Madsen[5], Emad N. Eskandar[6], Sydney S. Cash[3]

1 Program in Speech and Hearing Bioscience and Technology, Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, **2** Program in Medical Engineering and Medical Physics, Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, **3** Cortical Physiology Laboratory, Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, **4** Comprehensive Epilepsy Center, New York University School of Medicine, New York, NY, **5** Department of Neurosurgery, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, **6** Department of Neurosurgery, Massachusetts General Hospital and Harvard Medical School

ABSTRACT

In addition to its widespread clinical use, the intracranial electroencephalogram (iEEG) is increasingly being employed as a tool to map the neural correlates of normal cognitive function as well as for developing neuroprosthetics. Despite recent advances, and unlike other established brain mapping modalities (e.g. functional MRI, magneto- and electroencephalography), registering the iEEG with respect to neuroanatomy in individuals – and coregistering functional results across subjects – remains a significant challenge. Here we describe a method which coregisters high-resolution preoperative MRI with postoperative computerized tomography (CT) for the purpose of individualized functional mapping of both normal and pathological (e.g., interictal discharges and seizures) brain activity. Our method accurately (within 3mm, on average) localizes electrodes with respect to an individual's neuroanatomy. Furthermore, we outline a principled procedure for either volumetric or surface-based group analyses. We demonstrate our method in five patients with medically-intractable epilepsy undergoing invasive monitoring of the seizure focus prior to its surgical removal. The straight-forward application of this procedure to all types of intracranial electrodes, robustness to deformations in both skull and brain, and the ability to compare electrode locations across groups of patients makes this procedure an important tool for basic scientists as well as clinicians.

3.1 Introduction

In a large subset of patients with complex partial epilepsy, pharmacological intervention is ineffective (Engel et al. 2005). If non-invasive measures (e.g. EEG, PET, fMRI) fail to sufficiently localize the epileptogenic zone or if that zone abuts or overlaps eloquent cortex, arrays of electrodes placed either directly on the cortical surface or into deep structures (e.g. hippocampus, amygdala) may be indicated. Since its inception (Penfield & Jasper 1954; Engel et al. 2005), iEEG has been the gold-standard method for localizing seizure foci and delineating eloquent cortex in patients with medically-intractable epilepsy. Owing to its high spatiotemporal resolution and simultaneous coverage of wide areas of cortex, iEEG is increasingly being used as a tool to study the neural correlates of normal cognitive function (e.g., Crone et al. 2001; Yoshor et al. 2007; Brugge et al. 2008; Jerbi et al. 2009; Sahin et al., 2009) and examine spontaneous brain activity (e.g., Canolty et al. 2006; He et al. 2008; Cash et al. 2009). It has proven particularly informative in studying certain aspects of brain activity (e.g. gamma-band activity) which are less observable with non-invasive methods (i.e., EEG or MEG). More recently, intracranial electrodes are also being used as a recording platform from which to design brain-computer interfaces for various neuroprosthetic purposes, including communication aids for patients who suffer from amyotrophic lateral sclerosis or stroke (e.g., Wilson et al. 2006; Leuthardt et al. 2006; Felton et al. 2007; Schalk et al. 2008; Schalk 2010; Shenoy et al. 2008) .

For clinical as well as scientific purposes – including seizure-onset localization, eloquent-cortex mapping, and cross-subject comparison, as well as relating results to other neuroanatomical structures (e.g. white matter tracts) and the larger neuroimaging literature – knowledge of electrode location with respect to the patient's individual neuroanatomy is critical.

This is especially true given the spatial specificity of iEEG (due to its proximity to neural generators). Despite its importance, registering iEEG with a patient's individual cortical folding pattern remains a major challenge. Several solutions to this problem have been proposed, utilizing photography (Wellmer et al. 2002; Mahvash et al. 2007; Dalal et al. 2008), 2D radiography (Miller et al. 2007), postoperative MRI (Bootsveld et al. 1994; Kovalev et al. 2005; Wang et al., personal communication), or postoperative CT (Grzeszczuk et al. 1992; Winkler et al. 2000; Noordmans et al. 2001; Nelles et al. 2004; Hunter et al. 2005; Tao et al. 2009; Hermes et al. 2010), each with inherent limitations.

Here, we describe an electrode-localization procedure which combines the coregistration of high-resolution preoperative MRI with postoperative CT and the 3D rendering of each patient's cortical surface (Dale et al. 1999; Fischl et al. 1999a) or volumetric reslicing to align the slice with the long axis of depth-electrode arrays. The parenchymal shift introduced by the implantation of subdural electrodes (Hill et al. 1998; Hill et al. 2000; Hastreiter et al. 2004; Miyagi et al. 2007; Dalal et al. 2008; Hermes et al. 2010) is accounted for by using an optimization algorithm that minimizes an energy function defined by inter-electrode distances and global deformation of electrode configuration. This method minimizes assumptions about the nature of the parenchymal shift introduced by the implant and allows for accurate localization of electrodes near highly convex cortical regions. We extend previous work by providing a procedure for cortical surface-based inter-subject registration of each individual's electrode ensemble, allowing for surface-based group analyses of studies utilizing subdural electrodes. Given the fact that subdural electrodes are positioned on the cortical surface, this method of group analysis should prove more accurate when compared to the standard volumetric-based analyses (Collins et al. 1994; Miller et al. 2007; Ritzl et al. 2007; Talairach & Tournoux 1988).

We validate our registration method by comparison with intraoperative photographs, using prominent anatomical landmarks in order to determine the distance between estimated and actual electrode locations in reference to the 3D cortical reconstruction.

3.2 Methods

3.2.1 Patients

Patient	Hemisphere	Sex	Age	Electrode Arrays	Seizure Focus
P1	Left	Male	21	Four 2x8 One 1x8	Left Anterior Temporal
P2	Left	Male	22	One 8x8 One 1x8 Four 1x4	Left Cingulate Cortex
P3	Left	Male	29	One 8x8 One 1x8 Four 1x4	Left Anterior Temporal and Left Parietal
P4	Left	Female	23	One 8x8 Five 1x4	Anterior Middle Frontal
P5	Left	Male	20	One 8x8 Four 1x4	Posterior Superior Temporal and Middle Frontal

Table 3.1. Patient information.

Five patients with medically-intractable epilepsy underwent clinically-indicated invasive monitoring of the seizure focus prior to its surgical removal (Table 1). Patients gave their informed consent, and all procedures were approved by the Institutional Review Boards at Partners Healthcare (Massachusetts General Hospital and Brigham and Women's Hospital) and the Massachusetts Institute of Technology. Prior to electrode implantation, each patient underwent high-resolution T1-weighted MRI. During the implantation procedure, patients were placed under general anesthesia, a craniotomy was performed, and the overlying dura was reflected to expose the lateral aspect of the cortical surface (Fig. 3.1A and Fig. 3.1B). In each patient, arrays of platinum electrodes embedded in silastic sheets (2.3mm exposed diameter, 10mm center-to-center spacing, *Adtech Medical, Racine, WI*) were placed over temporal, frontal, and parietal cortex. In some cases inferior temporal, interhemispheric, and occipital cortex was covered as well. The reflected dura was then sewn over the electrode array and the skull was

replaced. After a brief recovery period, each patient underwent postoperative CT in order to assess electrode placement and to verify the absence of hemorrhage and mass effect.

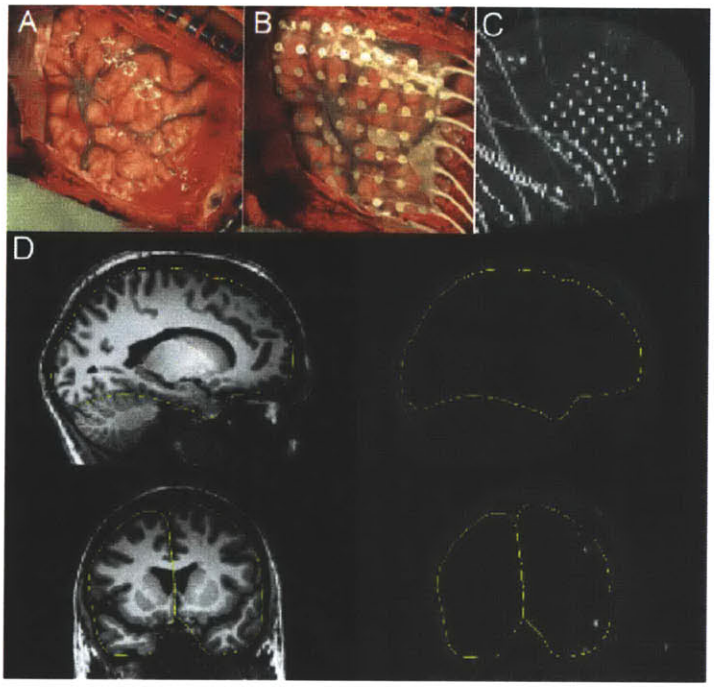

Figure 3.1. Intraoperative photographs, MRI-CT coregistration, and maximal-intensity projection.

(**A**) Reflected dura, exposed pial surface and overlaid electrode array (**B**) from a typical craniotomy. (**C**) The sagittal maximal-intensity projection of the postoperative CT scan, showing most of the electrode sites in a single view. (**D**) Illustration of the accuracy of the coregistration between the preoperative MRI and the postoperative CT. The left panels show sagittal (top) and coronal (bottom) views of a single subject's (patient 5) MRI; the right panels show the same orientations for the postoperative CT. Electrode sites can be seen as bright spots in the coronal CT section. The yellow trace outlines the pial surface in both the MRI and CT.

3.2.2 Coregistration of preoperative MRI with postoperative CT

High-resolution postoperative CT was automatically registered to the same patient's preoperative MRI using the mutual-information-based transform algorithm provided by SPM (http://www.fil.ion.ucl.ac.uk/spm/; Wells et al. 1996) in the Freesurfer environment (http://surfer.nmr.mgh.harvard.edu/fswiki/FreeSurferWiki). Using the CT scan as the movable

volume, the registration was visually checked and, if necessary, manually adjusted. Fig. 3.1D shows an example of this registration for a single patient.

3.2.3 Manual selection of electrode coordinates

Electrode coordinates were obtained in the patient's native anatomical space by visual inspection in the Freesurfer environment using *tkmedit*. Nearly all electrodes from an individual patient can be visualized in a 2D image by computing the maximum intensity projection of the CT volume in the plane approximately perpendicular to the long axis of the electrode arrays (usually sagittal for subdural arrays as shown in Fig. 3.1C and Fig. 3.3A, and usually coronal or horizontal for stereotactically-placed depth arrays). Doing so significantly reduces the time it takes to manually localize each electrode in three dimensions and transcribe its coordinate (either native Freesurfer RAS or MNI space). After locating each electrode in the 2D image, the final electrode coordinates were obtained by traversing slices in the 3rd dimension (through the plane in Fig. 3.1C) and selecting the approximate center of the hyper intensity created by each electrode. Each electrode's coordinate (in native RAS space) was then manually transcribed. This procedure yielded an initial estimate of a subdural electrode's coordinate on the cortical surface or a depth electrode's position within the volume.

3.2.4 Volumetric reslicing for visualization of depth-electrode arrays

In order to visualize all electrodes from a signal depth array, new slices from the 3D MRI volume were computed which aligned the new slices with long axis of each array. In general, stereotactic depth arrays are placed orthogonal or near-orthogonal to the sagittal plane; thus, reslicing was performed to yield peri-coronal or peri-horizontal images. The peri-coronal or peri-horizontal plane for each new slice was defined by the most superfical and deepest contacts. Results using such volumetric reslicing are shown in Fig. 3.2.

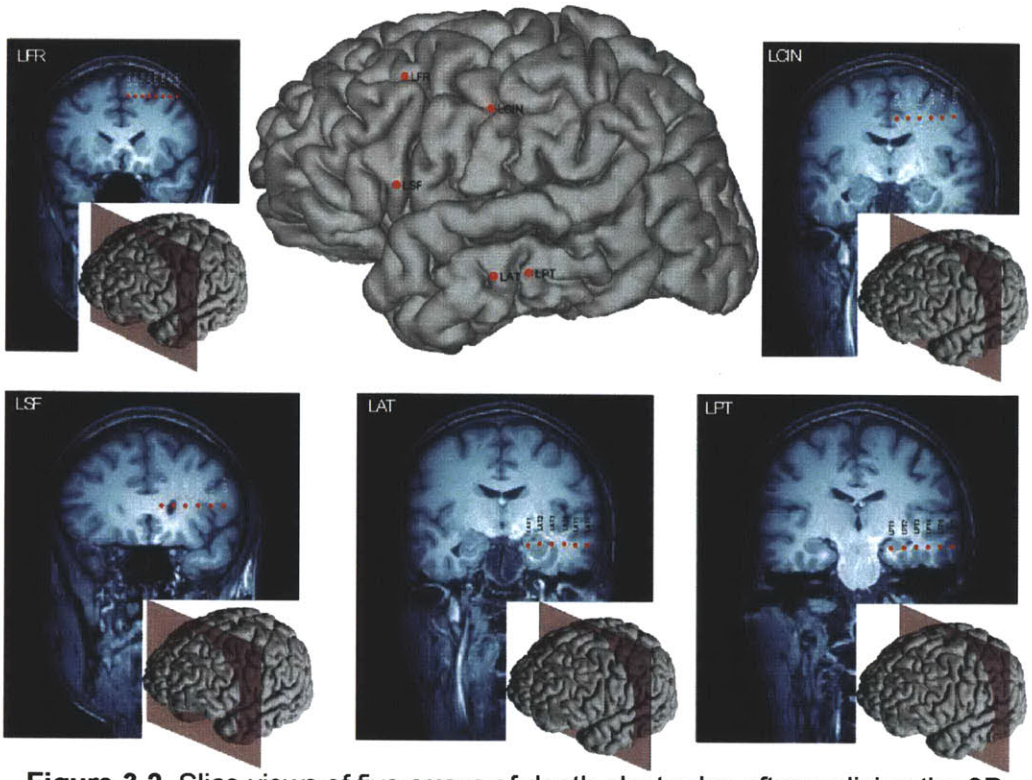

Figure 3.2. Slice views of five arrays of depth electrodes after reslicing the 3D MRI volume to the long axis of each array. The middle panel shows the entry point for each array overlaid on the reconstructed cortical surface.

3.2.5 Construction of pial and dural surfaces

3D rendering of each patient's cortical surface was performed in the Freesurfer environment using high-resolution T1-weighted MRI. Talairach registrations (using the Freesurfer talairach volume as a movable template) and white and gray matter surfaces were inspected and manually corrected if necessary. After a satisfactory pial surface was obtained, a smoothed pial surface (effectively a dural surface, Fig. 3.3B) was computed for both hemispheres using the morphological closing algorithm built into Freesurfer (Schaer et al. 2008).

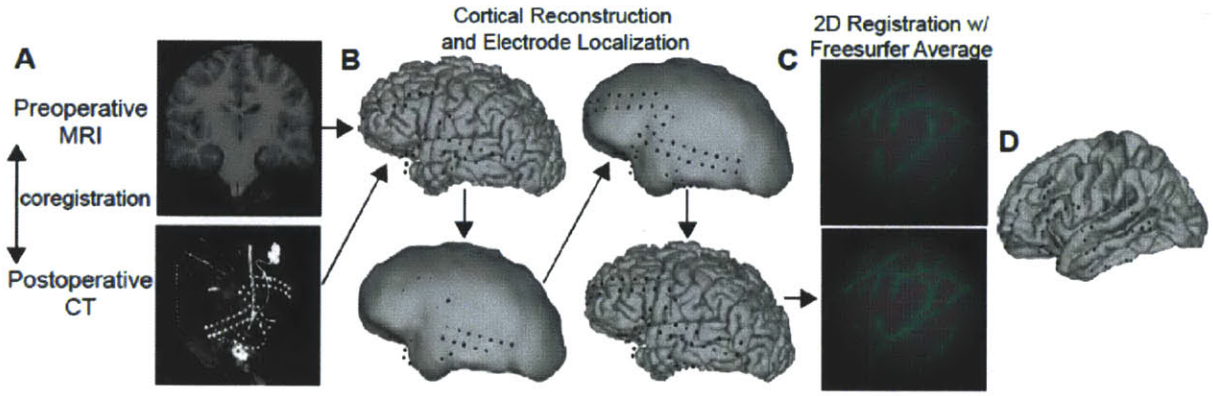

Figure 3.3. Outline of electrode localization and inter-subject mapping procedure.

(**A**) The preoperative MRI is coregistered with the postoperative CT volume. The lower panel shows the maximal intensity projection of the CT volume in the sagittal dimension, which shows all the electrodes in a sagittal plane. (**B**) Due to the parenchymal shift from the implant procedure, some electrodes initially appear as though buried in the gray matter. To correct for this, each electrode coordinate is projected first to a smoothed pial surface (effectively a dural surface) and subsequently back to the pial surface. (**C**) 2D registration with the Freesurfer average brain. An inflated spherical surface is computed from the individual's pial surface and aligned with that of the Freesurfer average. (**D**) Projection of each electrode coordinate from the individual's pial surface to that of the Freesurfer average.

3.2.6 "Snapping" electrode coordinates to the cortical surface

Due to brain deformation known to be caused by the implant procedure, initial electrode coordinates obtained from CT scans (coregistered with preoperative MRI) appear buried within the parenchyma instead of on the cortical surface when overlaid onto the cortical reconstruction (left panels of Fig. 3.3B). Our method accounts for this deformation automatically in two steps: (1) pulling the initial coordinate estimates to the dural surface via an energy minimization algorithm and (2) projecting the coordinates from the dural surface to the closest vertex on the pial surface (lower right panel of Fig. 3.3B).

The electrode coordinates are initially pulled to the dural surface using a constrained energy minimization algorithm. The algorithm must fulfill the constraint that all electrodes lie on the dural surface, while minimizing the displacement between the original and current electrode locations, as well as the deformation in the spatial configuration of the electrodes. The energy function is therefore composed of a displacement term as well as a deformation term and is defined as,

$$E(e)=E_{displacement}+E_{deformation}=\sum_{i=1}^{N}\|e_i-e_i^0\|^2+\sum_{i=1}^{N}\sum_{j=i+1}^{N}\alpha_{ij}\left(d_{ij}-d_{ij}^0\right)^2$$

with the constraint,

$$\forall i, \|e_i-s_i\|=0$$

where N is the number of electrodes, e_i is the location of electrode i, e_i^0 are the original coordinates for electrode i, d_{ij} is the distance between electrodes i and j, d_{ij}^0 is the original distance between that same pair of electrodes, α_{ij} is a parameter which takes values 1 or 0 determining whether the pair of electrodes i and j contribute to the energy function, and s_i is the location on the dural surface closest to electrode i.

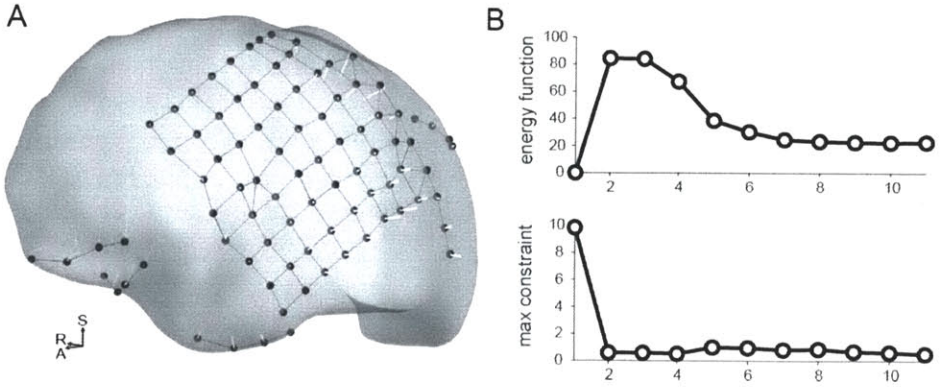

Figure 3.4. Illustration of the energy-minimization procedure used to project the electrode coordinates onto the cortical surface.

(**A**) Dural surface showing each electrode's path from it's initial volumetric location estimate to it's final location on the dural surface. (**B**) The top panel shows the value of the energy function across successive iterations of the algorithm. Note how the function approaches a stable value as the number of iterations increases. The bottom panel shows the value of the constraint function.

Intuitively, the energy algorithm places a series of "springs" between each electrode and its original position (displacement springs), as well as springs between the electrodes themselves (deformation springs). The energy in each spring increases if the electrodes are pulled apart or pushed together further than their equilibrium distance (0 for displacement springs, α_{ij} for deformation springs).

Ideally, the pairs of electrodes which define adjacent contacts on a grid or strip of electrodes would be included in the deformation term of the energy function ($\alpha_{ij}=1$), while distant pairs of electrodes would not exert an influence on the energy minimization ($\alpha_{ij}=0$). This would result in a mesh-like structure of springs which prevents individual electrodes from greatly deviating from the initial spatial configuration, but allows for "bending" or "flexing" of the silastic grids and strips (Fig. 3.4A). To automatically determine the pairs of electrodes included in the energy function, the distances between all pairs of electrodes are first computed and sorted. 2.5**N* of the shortest distances are averaged to obtain a "reference" distance. This reference distance approximates the distance between two adjacent grid electrodes. Any pair of electrodes, *i* and *j*, with a distance smaller than 150% of this reference distance is included in the energy function (i.e. α_{ij} is set to 1 for the pair of electrodes *i* and *j*). Empirically, this threshold yielded the best tradeoff between the desire to preserve local inter-electrode distances and

minimize long distance electrode pairings that could prevent flexibility of the grids or strips.

The optimization was performed in MATLAB using the *fmincon* algorithm in the optimization toolbox. The optimization was set to terminate when the total constraint function was less than $0.01*N$, the change in the energy function was less than 0.03, or the change in any $\boldsymbol{e}_i$ was less than 0.1. The total number of iterations was limited to 50.

After the electrodes are pulled to the dural surface, they are projected to the closest vertex (in Euclidean distance) on the pial surface. This second projection is necessary for functional mapping within an individual as well as accurate projection to the Freesurfer average brain (see section 2.8).

3.2.7 Validation

Validation of our localization procedure was carried out in two of five patients through visual inspection of intraoperative photographs and comparison with estimated electrode locations in reference to the reconstructed cortical surface. For electrodes which were visible in intraoperative photographs, local anatomical landmarks including prominent sulci and gyri were used to estimate, manually, an electrode's actual position on the reconstructed surface. Subsequently, for each electrode, the Euclidean distance between the location estimated by the MR-CT procedure and that determined by visual inspection of the photographs was computed. We used this distance as a measure of error.

3.2.8 Surface-based coregistration with Freesurfer average brain

For each patient, we computed a 2D spherical rendering of the pial surface and coregistered it with the average spherical surface provided in Freesurfer. Electrode coordinates from an individual were then transformed to the individual's registered spherical surface using a one-to-one look-up table, projected onto the spherical Freesurfer average surface by nearest-

neighbor transformation. Finally, these coordinates were overlaid onto the template pial surface by one-to-one look-up table (Fig. 3.3D). Given that this registration method is known to yield better alignment of structural as well as functional brain areas across subjects (Fischl et al. 1999b) and the fact that subdural electrodes most often rest on or near gyral crowns, projecting electrode coordinates from an individual's pial surface onto a template brain by spherical means could prove useful for group studies.

3.3 Results

3.3.1 Individual localization

Fig. 3.3B shows the results of our projection method for an individual subject implanted with four 2x8 and one 1x8 strip arrays. As can be seen in the upper and lower left-hand panels of Fig. 3.3B, several electrodes are either invisible or outside the brain as defined by the smoothed pial surface. After the snapping procedure outlined in section 2.5 was performed (Fig. 3.4), the electrodes now appear on the smoothed pial surface (upper right-hand panel of Fig. 3.3B), and were finally snapped to the original pial surface by nearest-neighbor mapping (lower right-hand panel of Fig. 3.3B).

3.3.2 Cross-subject registration

Unlike more established brain-mapping modalities such as fMRI, the ability to cross-register functional intracranial recordings across various individuals has proven difficult. The accurate localization of electrodes within individuals could allow for cognitive generalizations if those locations could be compared across subjects.

Fig. 3.3C and Fig. 3.3D show the surface-based coregistration of a single subject with the Freesurfer average surface (Fischl et al., 1999b). The 2D spherical surface of the individual subject is shown in the lower panel of Fig. 3.3C (Fischl et al., 1999a). Gyri and sulci are indicated by green and red, respectively. This surface has been coregistered with the Freesurfer average surface, shown in the upper panel of Fig. 3.3C using the procedure described in Fischl et al., 1999b. Each electrode coordinate was projected from the individual's pial surface (lower-right panel of Fig. 3.3B) to the same individual's spherical surface by a one-to-one mapping of the vertex. Electrode coordinates were then projected onto the Freesurfer average brain by nearest-neighbor mapping defined by Euclidean distance in the 2D spherical space. Electrodes

from an individual subject after having been projected onto the Freesurfer average are shown in Fig. 3.3D. Fig. 3.5 shows the results of the spherical registration procedure collapsed across all five subjects in the study after each subject's cortical surface was aligned with that of the Freesurfer average. While electrodes from different patients rarely overlap, this mapping allows direct comparison across subjects.

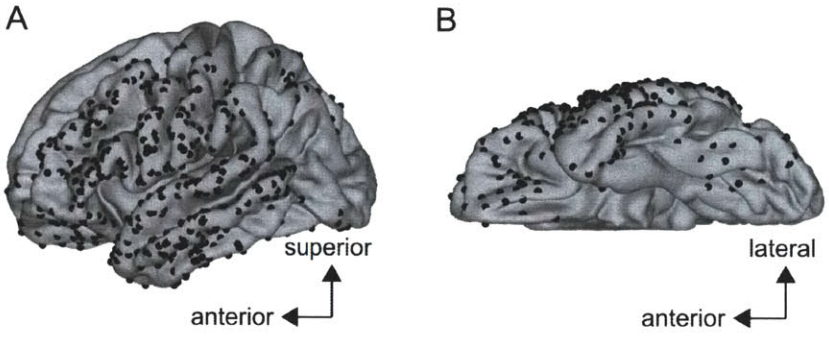

Figure 3.5. Electrode coordinates collapsed across all four patients mapped onto the Freesurfer average brain.

Each black dot represents an electrode from one of the four patients. (**A**) Lateral view of the left hemisphere. (**B**) Inferior view of the left hemisphere.

3.3.3 Validation

Estimates of the error in the electrode localization procedure were obtained in 2 of the 5 subjects in the study by manual inspection of electrodes visible in intraoperative photographs (Fig. 3.1A and Fig. 3.1B). An electrode's true location on the cortical surface was estimated with reference to major anatomical landmarks (e.g. gyri/sulci and/or vascular features). The estimated difference between the true electrode locations and those obtained by our localization procedure was 2.52mm and 3.00mm in patient 3 and patient 5, respectively. These values were found to qualitatively typical of the error distance across other patients. Interquartile ranges as well as minimum and maximum errors are given in table 2.

Patient	Median	Lower Quartile	Upper Quartile	Min	Max
P3	2.52	1.27	3.66	0.05	5.56
P5	3.00	2.04	4.01	0.52	8.16

Table 3.2. Estimated error of the localization procedure. All values in mm.

3.3.4 Example of method's utility

To illustrate application of the localization method, Fig. 3.6 shows the spatiotemporal voltage pattern of a single interictal discharge in patient 5 from 100 ms before to 1000 ms after the peak of the discharge. As can be seen from this figure, the discharge initiates at around 70ms along the border between the inferior and middle frontal gyri and subsequently spreads dorsally and ventrally to include the posterior middle frontal gyrus, anterior inferior frontal gyrus, and possibly anteroventral temporal areas. The initial spike was followed by a broader slow wave that initiated at approximately 130ms and continued throughout the remainder of the times shown in the figure. Our method has also recently been used to compare low-frequency (< 0.1 Hz) resting-state fMRI networks with corticocortical evoked potentials elicited by single-pulse stimulation with intracranial EEG electrodes (Keller et al., *in press*).

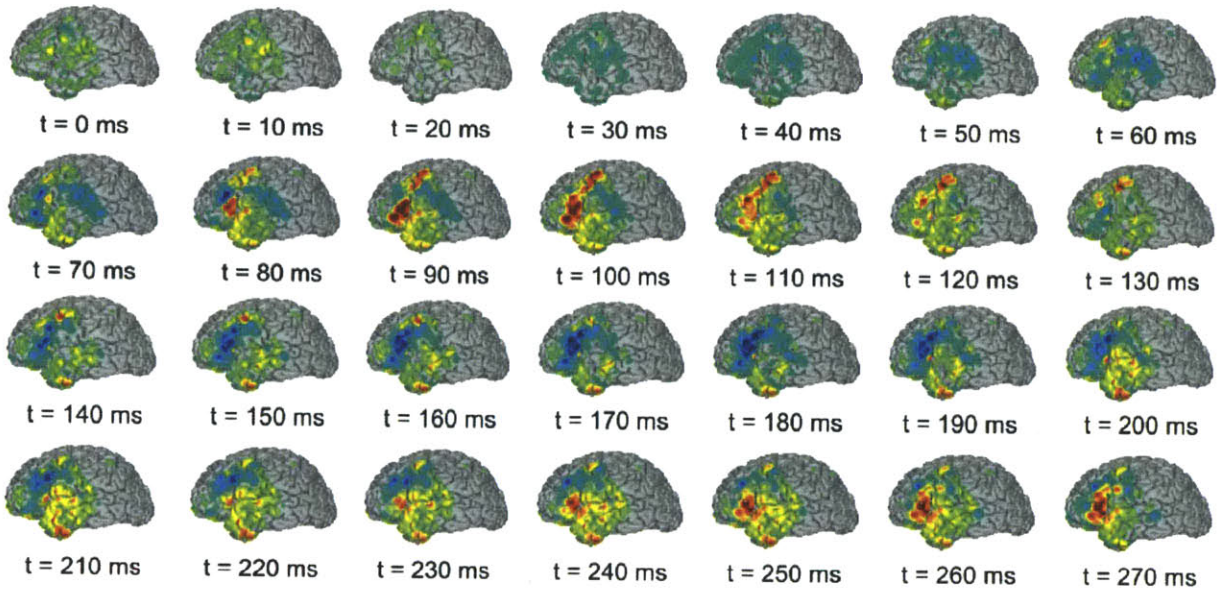

Figure 3.6. Spatiotemporal voltage map of an interictal discharge from patient 5. Color intensity indicates amplitude of positive (red) or negative (blue) voltage. The discharge initiates near the border of the middle and inferior frontal gyri at approximately 70ms into the epoched time window. A post-discharge undershoot can be seen beginning around 130ms and continuing until the start of the slow wave 230ms.

3.4 Discussion

The localization of intracranial electrodes is a critical issue in electrocorticography for clinical as well as scientific purposes. We have demonstrated a method for accurately aligning a postoperative CT scan with preoperative MRI for the purpose of individualized localization of semi-chronic intracranial electrodes. The method coregisters preoperative MRI with postoperative CT using a combination of an automatic mutual information-based procedure (Wells et al., 1996) and visual inspection. After this alignment, electrode coordinates were obtained manually using Freesurfer and projected onto the pial surface using a combination of an energy-minimization function and nearest-neighbor mapping. Our method affords patient-specific localization accuracy with estimated error on the order of half a centimeter or less,

comparable with previously-demonstrated methods using either a similar CT-MRI method (Hermes et al., 2010) or a method utilizing X-rays and intraoperative photographs (Dalal et al., 2008).

The main advantages of our method include (i) the same initial coregistration procedure operates on either subdural electrodes or depth electrodes accessing the mesial temporal-lobe structures (Fig. 3.2), (ii) it minimizes assumptions about the parenchymal shift (Hill et al. 1998; Hill et al. 2000; Hastreiter et al. 2004; Miyagi et al. 2007; Dalal et al. 2008; Hermes et al. 2010) introduced by the implant by using an optimization procedure to project the electrodes to the smoothed pial surface, (iii) it provides a procedure for cortical-surface-based inter-subject mapping (Fig. 3.5), (iv) – as mentioned by others (Hermes et al., 2010) – it can accurately localize electrodes positioned anywhere in the volume, rather than solely those in the vicinity of the craniotomy or away from highly convex areas such as the ventral temporal lobes, and (v) the method relies on very little commercial software (with the exception being MATLAB). With respect to (ii), our method provides a particular advantage over local-norm-based projection (Hermes et al., 2010) in that it applies equally well to different electrode configurations (e.g., 8x8 vs. 1x8) because it does not rely on the computation of a normal vector for which there may be no unique solution, e.g. in the case of 1xN array.

Our method is useful for basic scientific as well as clinical purposes. Knowledge of electrode location is critical for the interpretation of cognitive experiments and allows clinicians to extraoperatively visualize ictal and interictal activity at anatomically correct locations, possibly mitigating risk to the patient by reducing the need for intraoperative mapping. The surface-based inter-subject mapping procedure is especially useful for group studies. It has been shown previously that 2D spherical alignment of individual brains provides far greater

localization accuracy of both structural and functional neuroanatomical features when compared with volumetric-based alignment methods (Fischl et al., 1999b). This is especially likely to be the case for subdural electrode studies given that each electrode rests on or near the crown of a gyrus. A surface-based procedure is also likely to yield better overlap between electrocorticography findings and those from non-invasive methods such as functional MRI, scalp EEG, and magnetoencephalography.

Both the MRI-CT coregistration and the pial-surface projection are likely to introduce some localization error. The pial-surface projection error occurs due to some electrodes positioning between adjacent gyri over the intervening sulcus. However, in order to perform inter-subject surface-based registration, the error introduced by the pial-surface projection is unavoidable; thus, most of the correctable error is likely due to subtle inaccuracies in the MRI-CT registration. In our method, this registration was semi-automated, using a mutual-information-based algorithm shown previously to provide good alignment of within-subject multimodal images (Wells et al., 1996). This procedure is rendered more difficult, however, by the deformation of the skull caused by the craniotomy, and manual adjustments were often required. In our experience, scans yielding high-resolution isotropic or near-isotropic voxels provided the best initial coregistration estimate with the preoperative MR volume. In principle, it should be possible to use only the hemisphere opposite the craniotomy in computing the MRI-CT registration. This is likely to yield better automatic registration estimates but would require manual editing of the CT volume in order to mask the hemisphere with the craniotomy-induced deformations.

One limitation of our method is the time it takes to manually localize each electrode in the coregistered CT volume. Automation of the electrode localization would significantly reduce

the time spent during this step. A possible automation algorithm would use the dural surface as an inclusive volumetric mask for the CT volume after the MRI and CT have been coregistered. This would prevent the extraneous metal (e.g. wires and connectors) often included within the CT's field of view from generating false alarms during automatic electrode detection. A similar procedure has previously been described for coregistering preoperative and postoperative MR images (Kovalev et al., 2005). However, such an automation procedure would only be useful as a first pass electrode localization, and manual validation would likely still be required.

3.5 References

Bootsveld, K. et al., 1994. Localization of intracranial EEG electrodes using three dimensional surface reconstructions of the brain. *European Radiology*, 4(1). Available at: http://www.springerlink.com/content/wq4t753481u36412/ [Accessed September 27, 2010].

Brugge, J.F. et al., 2008. Functional localization of auditory cortical fields of human: click-train stimulation. *Hearing Research*, 238(1-2), pp.12-24.

Canolty, R.T. et al., 2006. High gamma power is phase-locked to theta oscillations in human neocortex. *Science (New York, N.Y.)*, 313(5793), pp.1626-1628.

Cash, S.S. et al., 2009. The human K-complex represents an isolated cortical down-state. *Science (New York, N.Y.)*, 324(5930), pp.1084-1087.

Collins, D.L. et al., 1994. Automatic 3D inter-subject registration of MR volumetric data in standardized Talairach space. *Journal of Computer Assisted Tomography*, 18(2), pp.192-205.

Crone, N.E. et al., 2001. Induced electrocorticographic gamma activity during auditory perception. Brazier Award-winning article, 2001. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 112(4), pp.565-582.

Dalal, S.S. et al., 2008. Localization of neurosurgically implanted electrodes via photograph-MRI-radiograph coregistration. *Journal of Neuroscience Methods*, 174(1), pp.106-115.

Dale, A.M., Fischl, B. & Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*, 9(2), pp.179-194.

Engel, A.K. et al., 2005. Invasive recordings from the human brain: clinical insights and beyond. *Nature Reviews. Neuroscience*, 6(1), pp.35-47.

Felton, E.A. et al., 2007. Electrocorticographically controlled brain-computer interfaces using motor and sensory imagery in patients with temporary subdural electrode implants. Report of four cases. *Journal of Neurosurgery*, 106(3), pp.495-500.

Fischl, B., Sereno, M.I. & Dale, A.M., 1999a. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, 9(2), pp.195-207.

Fischl, B., Sereno, M.I., Tootell, R.B. et al., 1999b. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, 8(4), pp.272-284.

Grzeszczuk, R. et al., 1992. Retrospective fusion of radiographic and MR data for localization of subdural electrodes. *Journal of Computer Assisted Tomography*, 16(5), pp.764-773.

Hastreiter, P. et al., 2004. Strategies for brain shift evaluation. *Medical Image Analysis*, 8(4),

pp.447-464.

He, B.J. et al., 2008. Electrophysiological correlates of the brain's intrinsic large-scale functional architecture. *Proceedings of the National Academy of Sciences of the United States of America*, 105(41), pp.16039-16044.

Hermes, D. et al., 2010. Automated electrocorticographic electrode localization on individually rendered brain surfaces. *Journal of Neuroscience Methods*, 185(2), pp.293-298.

Hill, D.L. et al., 1998. Measurement of intraoperative brain surface deformation under a craniotomy. *Neurosurgery*, 43(3), pp.514-526; discussion 527-528.

Hill, D.L. et al., 2000. Sources of error in comparing functional magnetic resonance imaging and invasive electrophysiological recordings. *Journal of Neurosurgery*, 93(2), pp.214-223.

Hunter, J.D. et al., 2005. Locating chronically implanted subdural electrodes using surface reconstruction. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 116(8), pp.1984-1987.

Jerbi, K. et al., 2009. Task-related gamma-band dynamics from an intracerebral perspective: review and implications for surface EEG and MEG. *Human Brain Mapping*, 30(6), pp.1758-1771.

Kovalev, D. et al., 2005. Rapid and fully automated visualization of subdural electrodes in the presurgical evaluation of epilepsy patients. *AJNR. American Journal of Neuroradiology*, 26(5), pp.1078-1083.

Leuthardt, E.C. et al., 2006. Electrocorticography-based brain computer interface--the Seattle experience. *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, 14(2), pp.194-198.

Mahvash, M. et al., 2007. Coregistration of digital photography of the human cortex and cranial magnetic resonance imaging for visualization of subdural electrodes in epilepsy surgery. *Neurosurgery*, 61(5 Suppl 2), pp.340-344; discussion 344-345.

Miller, K.J. et al., 2007. Cortical electrode localization from X-rays and simple mapping for electrocorticographic research: The "Location on Cortex" (LOC) package for MATLAB. *Journal of Neuroscience Methods*, 162(1-2), pp.303-308.

Miyagi, Y., Shima, F. & Sasaki, T., 2007. Brain shift: an error factor during implantation of deep brain stimulation electrodes. *Journal of Neurosurgery*, 107(5), pp.989-997.

Nelles, M. et al., 2004. Fusion of MRI and CT with subdural grid electrodes. *Zentralblatt Für Neurochirurgie*, 65(4), pp.174-179.

Noordmans, H.J. et al., 2001. Localization of implanted EEG electrodes in a virtual-reality environment. *Computer Aided Surgery: Official Journal of the International Society for*

Computer Aided Surgery, 6(5), pp.241-258.

Penfield, W. & Jasper, H.H., 1954. *Epilepsy and the functional anatomy of the human brain*, Little, Brown.

Ritzl, E.K. et al., 2007. Transforming electrocortical mapping data into standardized common space. *Clinical EEG and Neuroscience: Official Journal of the EEG and Clinical Neuroscience Society (ENCS)*, 38(3), pp.132-136.

N. Sahin *et al.*, Sequential processing of lexical, grammatical, and phonological information within Broca's area, *Science* 326 (2009), pp. 445–449.

Schaer, M. et al., 2008. A surface-based approach to quantify local cortical gyrification. *IEEE Transactions on Medical Imaging*, 27(2), pp.161-170.

Schaer, M. et al., 2008. A surface-based approach to quantify local cortical gyrification. *IEEE Transactions on Medical Imaging*, 27(2), pp.161-170.

Schalk, G. et al., 2008. Two-dimensional movement control using electrocorticographic signals in humans. *Journal of Neural Engineering*, 5(1), pp.75-84.

Schalk, G., 2010. Can Electrocorticography (ECoG) Support Robust and Powerful Brain-Computer Interfaces? *Frontiers in Neuroengineering*, 3, p.9.

Shenoy, P. et al., 2008. Generalized features for electrocorticographic BCIs. *IEEE Transactions on Bio-Medical Engineering*, 55(1), pp.273-280.

Talairach, J. & Tournoux, P., 1988. *Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system : an approach to cerebral imaging*, Thieme.

Tao, J.X. et al., 2009. The accuracy and reliability of 3D CT/MRI co-registration in planning epilepsy surgery. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 120(4), pp.748-753.

Wellmer, J. et al., 2002. Digital photography and 3D MRI-based multimodal imaging for individualized planning of resective neocortical epilepsy surgery. *Epilepsia*, 43(12), pp.1543-1550.

Wells, W.M. et al., 1996. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1), pp.35-51.

Wilson, J.A. et al., 2006. ECoG factors underlying multimodal control of a brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, 14(2), pp.246-250.

Winkler, P.A. et al., 2000. Usefulness of 3-D reconstructed images of the human cerebral cortex for localization of subdural electrodes in epilepsy surgery. *Epilepsy Research*, 41(2),

pp.169-178.

Yoshor, D. et al., 2007. Receptive fields in human visual cortex mapped with surface electrodes. *Cerebral Cortex (New York, N.Y.: 1991)*, 17(10), pp.2293-2302.

Chapter 4: Widespread Brain Areas Engaged during a Classical Auditory Streaming Task Revealed by Intracranial EEG

*This chapter was accepted for publication in *Frontiers in Human Neuroscience* on July 19, 2011

Andrew R. Dykstra[1,2], Eric Halgren[3], Thomas Thesen[4], Chad E. Carlson[4], Werner Doyle[4], Joseph R. Madsen[5], Emad N. Eskandar[6], Sydney S. Cash[2]

1 Program in Speech and Hearing Bioscience and Technology, Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, **2** Cortical Physiology Laboratory, Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, **3** Departments of Radiology and Neurosciences, University of California San Diego, San Diego, CA, **4** Comprehensive Epilepsy Center, New York University School of Medicine, New York, NY, **5** Department of Neurosurgery, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, **6** Department of Neurosurgery, Massachusetts General Hospital and Harvard Medical School

ABSTRACT

The auditory system must constantly decompose the complex mixture of sound arriving at the ear into perceptually-independent streams constituting accurate representations of individual sources in the acoustic environment. How the brain accomplishes this task is not well understood. The present study combined a classic behavioral paradigm with direct cortical recordings from neurosurgical patients with epilepsy in order to further describe the neural correlates of auditory streaming. Participants listened to sequences of pure tones alternating in frequency and indicated whether they heard one or two "streams." The intracranial EEG was simultaneously recorded from sub-dural electrodes placed over temporal, frontal, and parietal cortex. Like healthy subjects, patients heard one stream when the frequency separation between tones was small and two when it was large. Robust evoked-potential correlates of frequency separation were observed over widespread brain areas. Waveform morphology was highly variable across individual electrode sites both within and across gross brain regions. Surprisingly, few evoked-potential correlates of perceptual organization were observed after controlling for physical stimulus differences. The results indicate that the cortical network engaged during the streaming task are more complex and widespread than has been demonstrated by previous work, and that, by-and-large, correlates of bi-stability during streaming are probably located on a spatial scale not assessed – or in a brain area not examined – by the present study.

4.1 Introduction

The auditory system is constantly faced with the challenge of decomposing the complex mixture of sound arriving at the eardrums into an accurate representation of the acoustic environment. This decomposition, termed auditory scene analysis [ASA, (Bregman, 1994)], is critical for survival and communication and its failure is a common symptom reported by elderly individuals and those with sensorineural hearing loss. Despite its importance in daily life, the neural mechanisms of auditory scene analysis remain unclear (Shamma and Micheyl, 2010; Shamma et al., 2010; Snyder and Alain, 2007b; Micheyl et al., 2007; Bidet-Caulet and Bertrand, 2009; Carlyon, 2004; Winkler et al., 2009; Elhilali and Shamma, 2008; Nelken and Bar-Yosef, 2008). One aspect of ASA – auditory streaming (the segregation of time-varying acoustic energy into distinct perceptual objects) – can be studied in a controlled setting using sequences of pure-tone triplets of the form ABA-ABA- (Bregman, 1994; van Noorden, 1975; Miller and Heise, 1950), where A and B denote tones of different frequencies separated by a silent gap (Figure 1A). Many psychophysical studies dating back to the 1950s have shown that when the frequency gap (ΔF) between the A and B tones is small, listeners hear the sequence as a single stream comprised of both A and B tones and that when ΔF is large, they hear the sequence as two isochronous streams, one of A tones and one of B tones (van Noorden, 1975; Miller and Heise, 1950) (see http://web.mit.edu/~adykstra/Public/streaming_demo.wav for a demo). Interestingly, percepts evoked by sequences with intermediate ΔF are bistable (i.e., can be heard as either one stream or two) and can switch between two stable states, either spontaneously or with effort (van Noorden, 1975; Anstis and Saida, 1985; Carlyon et al., 2001).

Recent interest in the neural underpinnings of auditory streaming has produced several studies using ABA tone sequences while recording from the auditory cortex in a variety of

species including insects (Schul and Sheridan, 2006), fish (Fay, 1998, 2000), bats (Kanwal et al., 2003), songbirds (Bee and Klump, 2004, 2005; Bee et al., 2010; Itatani and Klump, 2009, 2010), ferrets (Elhilali et al., 2009), non-human primates (Fishman et al., 2001, 2004; Micheyl et al., 2005), and humans (Sussman et al., 1999; Gutschalk et al., 2005, 2007; Snyder et al., 2006; Snyder and Alain, 2007a; Wilson et al., 2007; Cusack, 2005; Deike et al., 2004, 2010; Kondo and Kashino, 2009; Schadwinkel and Gutschalk, 2010a, 2010b). A prevailing model from these

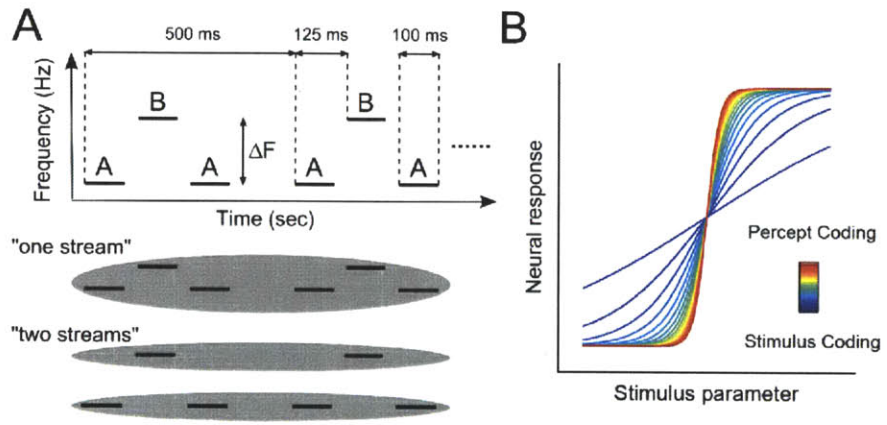

Figure 4.1. Behavioral paradigm and conceptual model.

(**A**) Schematic illustration of the alternating-tone stimuli used in the experiment and how those stimuli are perceptually organized by the listener. The frequency of the B-tone was held constant at 1000 Hz and the frequency separation between the A- and B- tone varied between 0 and 12 semitones, resulting in A-tone frequencies between 500 and 1000 Hz. (**B**) Conceptual model of varying neural responses to parametric manipulation of the acoustic parameter (frequency separation). A linear variation of the neural response is to be expected if that response is coding the stimulus parameter, whereas a sigmoidal (i.e. categorical) response is to be expected if the response is coding the percept directly.

studies posits that a two-stream percept will be evoked whenever the A and B tones excite non-overlapping populations of neurons [but see (Elhilali et al., 2009)]. However, inherent limitations in previous work related to spatiotemporal resolution, sparsity of coverage, and lack

of direct behavioral measures in experimental animals preclude straight-forward interpretation. A general extension of this model is schematized in Figure 1B. Specifically, a parametric variation of a given stimulus or stimulus feature could produce neural activity patterns which vary linearly or categorically as shown by the blue and red curves, respectively. Noise in the response of a population showing a linear relationship with the stimulus, when fed to a population showing a more categorical relationship, could engender sufficient trial-to-trial variability for bistable perception. While such activity patterns have been widely reported in vision [for reviews see (Leopold and Logothetis, 1999; Logothetis, 1998; Sterzer et al., 2009)], only limited evidence for such a mechanism exists in the auditory system (Cusack, 2005; Gutschalk et al., 2005, 2008; Kondo and Kashino, 2009).

Here, we report the results from experiments in which direct cortical recordings were made from widespread brain areas of neurosurgical patients with epilepsy (Engel et al., 2005) while they participated in a classical auditory streaming paradigm. Our aims were to better characterize the neurophysiological correlates of auditory streaming, extend them into brain areas outside the auditory cortex and frequency regions less observable with non-invasive measure (Crone et al., 2001), and test the idea of neuronal variability as a mechanism for perceptual bi-stability in the auditory modality (Almonte et al., 2005; Deco and Romo, 2008; Deco et al., 2008; Moreno-Bote et al., 2007; Gigante et al., 2009; Shpiro et al., 2009) by comparing evoked responses to physically-identical stimuli when they were perceived as one vs. two streams. Our participants listened to ABA tone sequences and indicated at the end of each sequence whether they were hearing one or two streams at the end of the sequence. For each electrode sampled in a given patient, we compared responses across ΔF conditions as well as perceptual report in an attempt to identify correlates of both during a classical auditory streaming

task. We hypothesized that when a participant perceived one (two) stream(s), the evoked response would be similar to those conditions which consistently engender a one-stream (two-stream) percept. Responses from widespread brain areas showed robust correlates with ΔF but, surprisingly, rarely differed based on percept *per se*.

4.2 Methods

4.2.1 Ethics Statement

All procedures were approved by the Institutional Review Boards at Partners Healthcare (MGH and BWH), the New York University (NYU) Langone Medical Center, and the Massachusetts Institute of Technology (MIT) in accordance with NIH guidelines. Written informed consent was obtained from all patients prior to their participation.

4.2.2 Listeners

Twelve patients with intractable epilepsy underwent invasive monitoring in order to localize the epileptogenic zone prior to its surgical removal.

Patient	Hemisphere	Sex	Age	Stimulation	ΔF conditions	Seizure Focus
S1	Left	M	21	Binaural	0,5,6,7,12	Left Anterior Temporal
S2	Left	M	22	Binaural	0,5,6,7,12	Left Cingulate Cortex
S3	Left	F	52	Binaural	0,5,6,7,12	Left Anterior Temporal
S4	Right	F	42	Binaural	0,5,6,7,12	Right Anteromesial Temporal
S5	Left	M	37	Binaural	0,5,6,7,12	Left Anteromesial Temporal
S6	Right	M	19	Binaural	0,2,4,6,8,10,12	Right Parietal
S7	Left	F	38	Monaural (R)	0,2,4,6,8,10,12	Left Mesial Temporal
S8	Left	M	29	Binaural	0,2,4,6,8,10,12	Left Anterior Temporal / Left Parietal
S9	Left	M	25	Monaural (R)	0,2,4,6,8,10,12	Left Anteromesial Temporal

Table 4.1. Patient information.

Each patient was implanted with an array of sub-dural platinum-iridium electrodes embedded in silastic sheets (2.3mm exposed diameter, 10mm center-to-center spacing; *Ad-tech Medical, Racine, WI*) placed directly on the cortical surface. Prior to implantation, each patient underwent high-resolution T1-weighted MRI. Subsequent to implantation, patients implanted at Massachusetts General Hospital (MGH) and Brigham and Women's Hospital (BWH) underwent postoperative computerized tomography (CT); patients implanted at NYU underwent postoperative MRI. Electrode coordinates obtained from postoperative scans were co-registered with preoperative MRI and overlaid onto the patient's reconstructed cortical surface using

Freesurfer (Dale et al., 1999; Fischl et al., 1999a) and custom MATLAB (*The Mathworks, Framingham, MA*) scripts (A. Dykstra et al., submitted; H. Wang et al., Comprehensive Epilepsy Center, NYU School of Medicine, personal communication]. Electrode coordinates were then projected onto the Freesurfer average brain using a spherical registration between the individual's cortical surface and that of the Freesurfer average (Fischl et al., 1999b). The data from three patients were excluded from analysis due to excessive noise caused by technical malfunction; the data reported here are from the remaining nine patients (Table 1).

4.2.3 Stimuli and procedure

Stimuli were long sequences of pure-tone triplets of the form ABA-ABA-..., where A and B represent individual tones and the dash represents a silent gap (Figure 3A). Each tone was 100ms in duration with 10ms raised-cosine on- and off-ramps. The inter-stimulus interval (ISI) between the first A-tone and B-tone, as well as between the B-tone and second A-tone, was 25ms; the ISI between the second A-tone and subsequent triplet was 150ms. Stimulus onset asynchrony (SOA) between successive A-tones was 250ms; SOA between successive B-tones was 500ms; triplet onset asynchrony was also 500ms. Total duration of each sequence varied between 6.5 and 10 seconds (13 and 20 triplets, respectively) depending on the listener (for P1-P5, duration varied between 6.5 and 7.5 seconds; for P6-P9, duration was 10 seconds). The B-tone frequency was fixed at 1 kHz. The A-tone frequency varied between 0 and 12 semitones below the B-tone. Listeners P1, P2, P3, P4, and P5 participated in conditions in which the frequency separation was 0, 5, 6, 7, or 12 semitones, where 1 semitone is an approximately 6% frequency difference. Listeners P6, P7, P8, and P9 participated in conditions in which the frequency separation was 0, 2, 4, 6, 8, 10, or 12 semitones. Each patient listened to between 200 and 378 triplets for a given frequency separation. All sounds were generated digitally in

MATLAB, stored as .wav files, and converted to analog waveforms by the on-board sound card of a laptop equipped with Presentation software (*Neurobehavioral Systems, Albany, CA*). Stimuli were presented at a comfortable listening level via Etymotic ER-2 insert earphones (*Etymotic Research, Inc., Elk Grove Village, IL*), diotically (when possible) or monaurally contralateral to the hemisphere of implantation. Patients were instructed to listen to the sounds and to indicate at the end of each sequence whether, at the end of the sequence, they were hearing a single "stream" comprised of all tones or two "streams," one comprised of A tones and the other of B tones. Responses were made by button press with a response box (*Cedrus Corporation, San Pedro, CA*) interfaced with Presentation via USB. Response windows were unconstrained, and the subsequent stimulus began one second after a response to the previous stimulus was entered.

4.2.4 Data acquisition

Intracranial EEG (iEEG) data at MGH and BWH were acquired with standard clinical EEG monitoring equipment (*XLTEK, Natus Medical Inc., San Carlos, CA*) at a sampling rate of 250 Hz (P1) or 500 Hz (P2,P3,P6,P8). At NYU, iEEG data were acquired with a customized system at a sampling rate of 30 kHz (P4,P5,P7,P9). All data were subsequently re-sampled to 500 Hz for analysis. All data were referenced to either an inverted intracranial electrode (i.e. facing the inner skull table) remote from the electrodes of interest (P1,P2,P3,P6,P8) or a screw bolted to the skull (P4,P5,P7,P9). For each patient, clinically-indicated, high-resolution T1-weighted structural MRI scans were acquired prior to surgery. High-resolution CT (P1,P2,P3,P6,P8) or structural MRI (P4,P5,P7,P9) scans were acquired subsequent to surgery for the purpose of electrode localization.

4.2.5 Data pre-processing

Intracranial EEG data were bandpass filtered offline between 1 and 190 Hz and notch filtered at

60 Hz and its harmonics using zero-phase shift FIR filters. Independent component analysis using the *runica* algorithm (Bell and Sejnowski, 1995) in EEGLAB (Delorme and Makeig, 2004) was performed on the "raw" data. Components dominated by large artifacts were identified and removed by inspection. The component data were then back-projected in order to remove the artifacts from the original data.

The iEEG was epoched relative to the onset of sound sequences (yielding long epochs encompassing the entire sequence) as well as to the onset of individual ABA triplets (yielding short epochs of 0.5 seconds) and binned with respect to either ΔF or perceptual report within a given ΔF. For triplet-locked epochs, the first triplet in each sequence was discarded. Epochs were baseline-corrected with respect to either the 500 ms preceding sequence onset (for sequence-locked epochs) or the 50 ms preceding triplet onset (for triplet-locked epochs). Epochs containing large artifacts were rejected automatically using joint probability and kurtosis algorithms in EEGLAB (Delorme et al., 2007). Specifically, trials with joint probabilities or kurtosis values more than 4 and 5 standard deviations from the normalized mean of these measures, respectively, were rejected as artifact. Additional epochs found to contain large epileptiform activity were rejected by visual inspection.

4.2.6 Statistical analysis

A modified version of the cluster-based, non-parametric statistical procedure outlined by Maris and Oostenveld (Maris and Oostenveld, 2007) was used to test for effects of ΔF and bistability on triplet-locked EP amplitude. Spearman (nonparametric) rank correlation (in the case of a multiple-level factor, e.g., ΔF) and unpaired t-test (in the case of two-level factors, e.g., percept) were used as the sample-level (i.e. individual time point within a single channel) statistics in order to evaluate possible effects of ΔF (five levels for P1-P5 and seven levels for

P6-P9) and bistability (always two levels), respectively. Contiguous, statistically-significant samples (defined as $p<0.05$) within a single electrode were used to define the cluster-level statistic, which was computed by summing the sample-level statistics within a cluster. Statistical significance at the cluster level was determined by computing a Monte Carlo estimate of the permutation distribution of cluster statistics using 1000 re-samples of the original data (Ernst, 2004). For multiple-level factors (ΔF), the estimate of the permutation distribution was performed by 1000 re-samples of the condition labels associated with each level in the factor. Within a single electrode, a cluster was taken to be significant if it fell outside the 95% confidence interval of the permutation distribution for that electrode. The determination of significant clusters was performed independently for each electrode. This method controls the overall false alarm rate within an electrode across time points; no correction for multiple comparisons was performed across electrodes.

Due to the known buildup effects of auditory streaming (i.e., 2-stream percepts become more likely as time since sequence onset increases and the fact that listeners only reported what they heard at the end of each stimulus sequence, two independent analyses were carried out. The first used only the data from the second half of each sequence while the second used all data after removing the onset response (0-0.5 sec after stimulus onset). The method of analysis did not effect the results, and only the results from the second analysis are shown.

4.2.7 Dissimilarity index

In order to further evaluate possible effects of perceptual bistability on the evoked waveforms, we computed a dissimilarity index between waveforms from individual trials and a template waveform within individual channels in which significant EP-ΔF correlations were found. Qualitatively, this index is defined as the difference between the sum-squared error (SSE)

computed for the condition of interest (i.e, a specific ΔF or percept) and the minimum SSE computed across all conditions, normalized by the difference between the maximum SSE and minimum SSE computed across all conditions. The index was computed by normalizing the average sum-squared error between the trial and the template, as follows:

$$SSE_{ij} = \sum_{t=1}^{T} \left(X_{ij}^{t} - X_{0}^{t} \right)^2 ,$$

where X_0 is the template waveform and X_{ij} is the individual-trial waveform for trial i in condition j, t is the individual time point, and T is the overall number of significant time points in condition j. The average SSE for condition j was computed as:

$$SSE_{j} = \frac{\sum_{i=1}^{N} SSE_{ij}}{N} ,$$

where N is the number of trials.

The index was then defined as:

$$DI_{j} = \frac{SSE_{j} - \min\limits_{j} SSE_{j}}{\max\limits_{j} SSE_{j} - \min\limits_{j} SSE_{j}}$$

Except for trials from the 0-semitone condition, the template was defined as the average EP for the 0-semitone condition. The template to which individual trials from the 0-semitone condition were compared was the average EP from the 0-semitone condition including all waveforms but the one from the trial i ("leave one out"). This index provides a measure of how dissimilar two waveforms are from each other. Although this index is biased to show a significant correlation with ΔF, it provides a means to (i) collapse waveforms across individual electrode sites and patients into a single quantitative metric and (ii) quantitatively compare responses to one- vs. two-stream percepts in a way that circumvents variable latencies and durations of percept- or ΔF-based effects across sites.

4.2.8 High-gamma power

Waveforms of high-gamma power were constructed using the wavelet transforms built into EEGLAB (specifically, the *newtimef* function). Sequence-length (between 6.5 seconds and 10 seconds) epochs were used to compute the event-related spectral perturbation (ERSP) which was baseline corrected to the 500ms preceding stimulus onset. The number of wavelet cycles used varied logarithmically with respect to frequency from three cycles at the lowest frequency tested (5 Hz) to 10 at the highest (190 Hz), yielding approximate temporal resolution of < 500 ms at 8 Hz and < 125 ms in the gamma band. High-gamma power waveforms were constructed by summing the power in frequencies from 80-190 Hz for each time point in the full time-frequency representation. These waveforms were then baseline-corrected by subtracting the mean power in each trial computed across the 500 ms preceding stimulus onset. Triplet-locked gamma-power epochs were constructed by time-locking with respect to each triplet onset and subsequently binned across the various ΔF and percept conditions in the same way as the evoked potentials. The same statistical procedures described above were applied to the high-gamma waveforms.

4.3 Results

12 patients with intractable epilepsy listened to sequences of alternating pure tones (Figure 1A) and indicated at the end of each sequence whether, at the end of the sequence, they were hearing the tones as grouped ("1 stream") or segregated ("2 streams") while we simultaneously recorded the intracranial EEG (Figure 2). Three patients were excluded from analysis for technical reasons (see Materials and Methods). Summed across the remaining nine patients (Table 1), we recorded from nearly 700 electrodes in the left hemisphere and 250 electrodes in the right hemisphere, mostly on lateral cortex of the temporal, frontal and parietal lobes (Figure 2E).

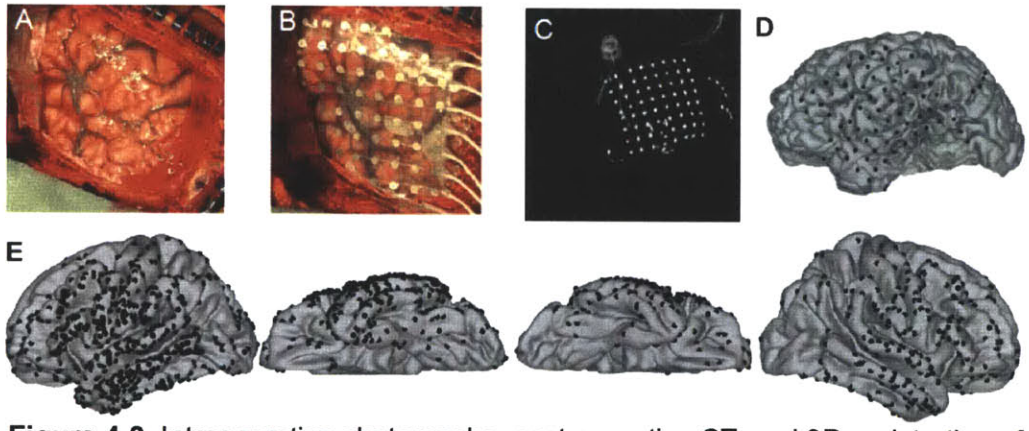

Figure 4.2. Intraoperative photographs, post-operative CT, and 3D registration of electrode coordinates on the cortical surface.

(**A**) and (**B**) Intraoperative photographs showing the reflected dura, exposed pial surface, and overlaid electrode array (**B**) of an example subject who participated in the study. (**C**) Maximal-intensity projection of sagittally-oriented CT scan showing all of the intracranial electrodes collapsed onto a single plane. (**D**) Electrodes overlaid onto a 3D rendering of the patient's cortical surface.(**E**) Summary of all individual electrode sites. Electrode coordinates from all 9 participants in the study were co-registered and overlaid onto the Freesurfer average surface. In total, we sampled from nearly 1000 sites, mostly over lateral cortex.

4.3.1 Behavior

Figure 3 shows the probability of hearing two streams as a function of ΔF averaged

across all nine patients included in the analysis. Patients reported hearing a single stream when the ΔF was small and two streams when ΔF was large. At intermediate ΔF, the percept was bistable, i.e. patients sometimes reported hearing one stream and sometimes reported hearing two streams. A Kruskal-Wallis test confirmed a main effect of ΔF ($\chi^2(1,8) = 34.1$; $p < 0.0001$).

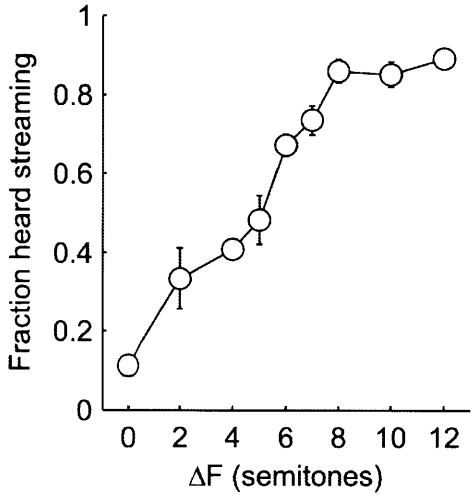

Figure 4.3. Behavioral results.

Subjects heard one stream when the frequency separation was small and two when it was large. Intermediate frequency separations perceptually bistable. Error bars represent the standard error of the mean across subjects.

4.3.2 Evoked Potentials: ΔF

In order to assess putative correlates of streaming, we tested for correlations between triplet-locked EP amplitude and ΔF which, when parametrically varied, produced changes in how the sequences were perceptually organized. In light of the known effects of perceptual buildup in streaming tasks, two analyses were carried out: one using only the triplet-locked responses from the second half of each sequence and another using all the

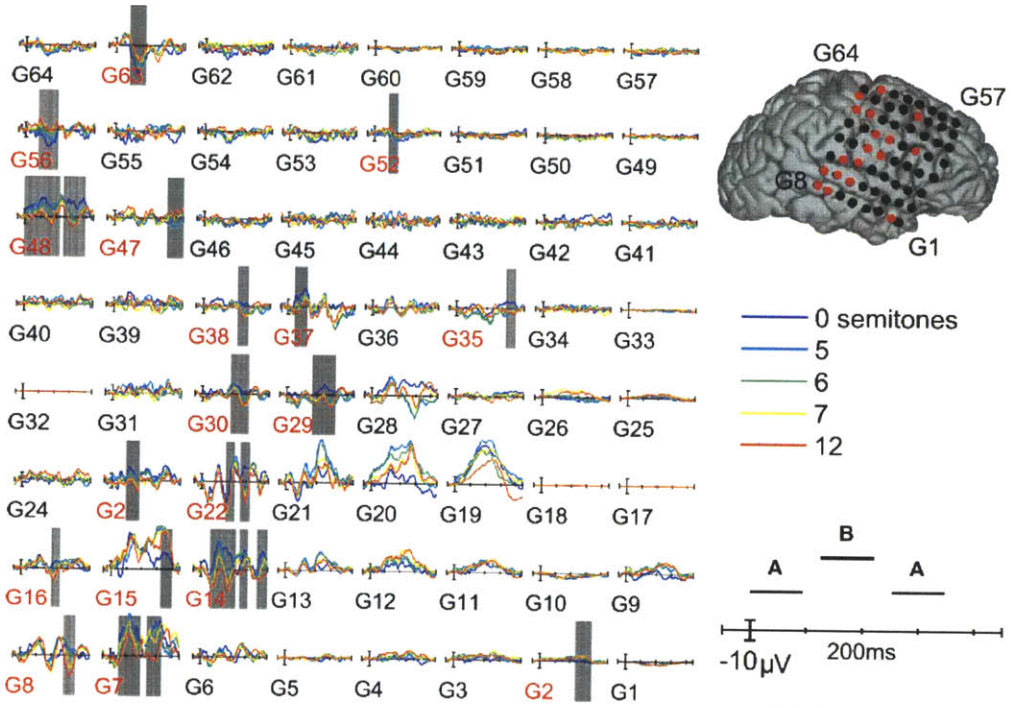

Figure 4.4. Example responses from an individual subject.

Triplet-locked evoked potentials from an 8x8 electrode array over the right hemisphere whose configuration on the individual's cortical surface is shown at right. Different frequency-separation conditions are represented by different color waveforms as shown at lower right. Significant correlations between the acoustic parameter (frequency separation) and EP amplitude are indicated by gray shading behind the waveforms in each plot. The responses showing significant correlations with frequency separation are also indicated by red text or dots.

responses to all triplets save for the first (see Materials and Methods). The results did not differ based on which analysis was used, thus only the second analysis is reported here. Significant correlations between ΔF and EP amplitude were determined by cluster-based nonparametric permutation statistics (Materials and Methods). Figure 4 shows the average triplet-locked evoked responses across an 8x8 grid of electrodes for the different ΔF conditions from a single patient (P4). The positions of each electrode are overlaid onto the patient's cortical surface rendering. Examples from other patients can be found in the chapter 4 appendix. As can be seen, waveform morphology was complex and highly variable between different electrode sites,

yet evoked responses in varying time windows and spatial positions robustly correlated with ΔF. The majority of sites which showed strong correlations with ΔF were over or adjacent to the posterior superior temporal gyrus (pSTG). However, several other sites also showed responses which correlated with ΔF. The sites which showed significant ΔF correlations across all nine patients included in the analysis are summarized in Figure 5, where electrode sites from each individual have been overlaid onto a template brain by spherical surface registration of each patient's pial surface with that of the Freesurfer average (See Materials and Methods). Across patients, a widespread set of brain areas showing significant correlations with ΔF included pSTG (as was expected), middle temporal gyrus, pre- and post-central gyri (mainly ventrally), inferior and middle frontal gyri, and the supra-marginal gyrus.

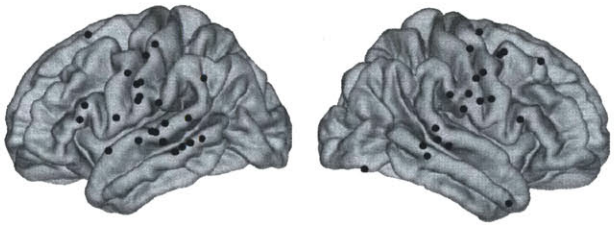

Figure 4.5. Summary of electrode sites that showed significant EP amplitude correlations with frequency separation.

4.3.3 Evoked Potentials: Bistable Perception

After having established significant correlations with a physical stimulus parameter (ΔF) known to produce changes in perceptual organization, we explicitly tested whether the same electrode sites showed significant triplet-locked EP differences based solely on how the sequences were perceptually organized (i.e., we compared EPs between sequences perceived as one stream vs. two streams within a given ΔF condition). For a given ΔF, responses were binned and averaged according to whether the listener reported hearing one or two streams. As for the

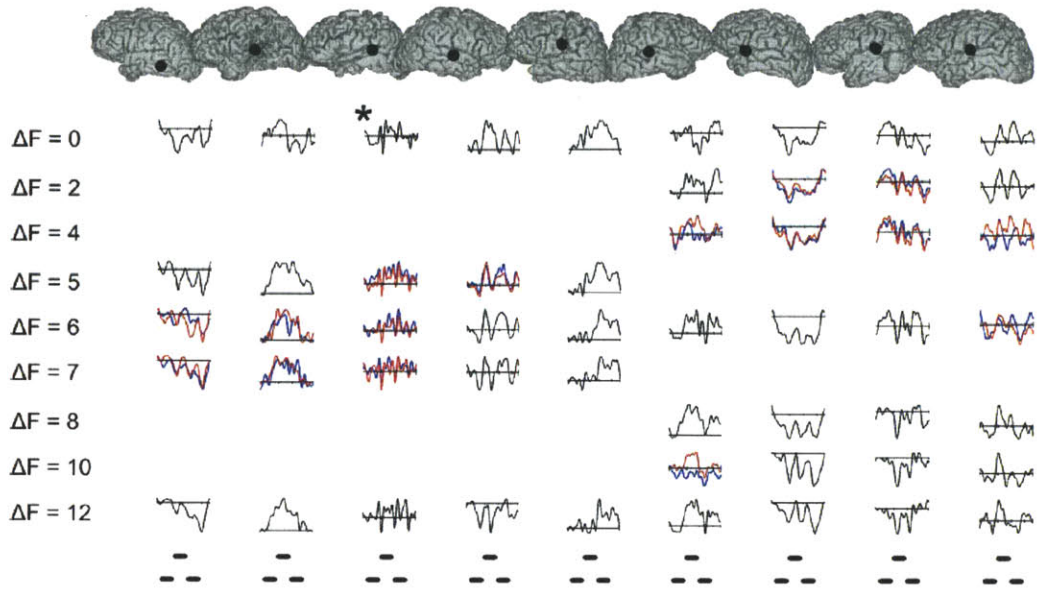

Figure 4.6. Evoked potentials from individual peri-Sylvian electrode sites in each of the 9 subjects.

Blue and red traces for a given frequency separation and subject indicate that the percept for that condition was bistable (*, this patient did not understand the task). Waveforms traced in black indicate that the percept for that condition was unistable. Electrode sites, shown in the top row over each subject's cortical reconstruction, were chosen based on their having the largest RMS power grand-average triplet-locked evoked response in the vicinity of the superior temporal gyrus. The frequency separation (ΔF, semitones) for each set of waveforms is indicated in the left-most column. The timing of individual tones in the triplet is shown in the bottom row.

analysis testing for effects of ΔF, two analyses were carried out; one using only the responses from the second half of each sequence and the other using responses from the entire sequence, expect for the first. Only the results from the second analysis are presented here. The results of this analysis for individual peri-STG sites across all nine patients are shown in Figure 6. The sites, overlaid onto each individual's pial surface as shown in the top row, were chosen based on the fact that each showed a significant correlation with ΔF and was the site with largest triplet-locked RMS power in the vicinity of the pSTG. Responses to sequences that were perceptually bistable (defined as: $0.3 \leq P(\text{2-stream percept}) \leq 0.7$) are shown by the blue (1-stream percepts)

and red (2-stream percepts) traces; otherwise, traces are black. As can be seen, EP morphology was highly variable across individual subjects. Waveforms changed significantly as a function of ΔF as determined by Monte Carlo permutations using Spearman rank correlation as the sample-level statistic (see Methods), but, surprisingly, did not show significant differences based on percept *per se*. Across all the channels in the study, there were individual channels which showed significant differences based on percept, but this effect was inconsistent across the multiple ΔF conditions for which a bistable percept was evoked. In summary, several brain areas both within and outside of the auditory cortex showed evoked responses that significantly correlated with ΔF but not conscious perceptual organization.

4.3.4 Dissimilarity Analysis

In order to further evaluate whether sites showing significant EP-ΔF correlations also showed correlates of perceptual bistability, we carried out a dissimilarity analysis using the grand average triplet-locked response to the 0-semitone condition as the template. Responses from each ΔF condition were binned according to percept as well as collapsed across them and compared to the template by sum-squared error (SSE, see Materials and Methods). Our hypothesis was that responses from conditions with greater ΔF – as well as responses from trials in which the subject reported hearing two streams – would show a larger "dissimilarity index" computed from the SSE between the response of interest and the template. Figure 7 shows the results of this analysis. The value of the dissimilarity index increased as ΔF increased (Spearman's rho = 0.46, $p < 0.0001$) and, across all channels from all patients, showed a marginally significant difference based on percept alone ($W_+ = 3427$, $p = 0.097$) in the expected direction (i.e., greater dissimilarity indices for 2- vs. 1-stream percepts), suggesting a propensity for activity during 2-stream percepts to be more similar to activity evoked by large ΔF conditions. However, a sufficient number of channels (23%) showed the opposite pattern so as to limit the statistical significance of the effect. Individually, across all sites which showed a significant correlation with ΔF (N=44), four channels

showed significant effects of percept on the dissimilarity index in the expected direction, while none showed a significant effect in the opposite direction. Three of those channels were from S4 [G30, G37, and a site over the left posterior STG (not shown)] whose data are shown in Figure 4, and the fourth was from a site over the inferior post-central gyrus in S1 (not shown). None of the four channels which showed significant percept-based differences in the dissimilarity index showed significant differences in the waveforms when evaluated directly.

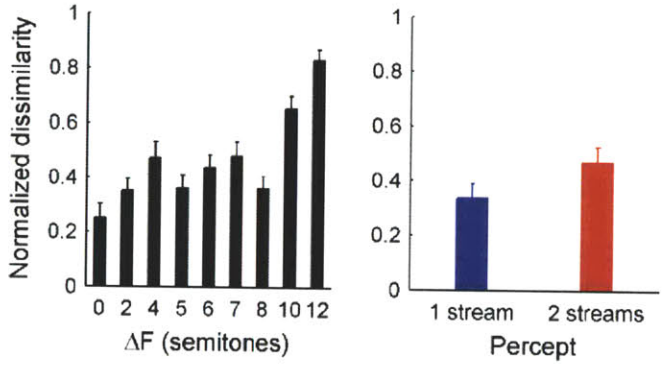

Figure 4.7. Dissimilarity index.

The left panel shows the dissimilarity index as a function of frequency separation collapsed across percept. The right panel shows the dissimilarity index as a function of percept collapsed across delta-F conditions in which the percept was bistable.

A complementary analysis was carried out using the grand-average triplet-locked response collapsed across all conditions as the template (Figure 1 in appendix). *Averaged* evoked responses from each ΔF condition were binned according to percept as well as collapsed across them and compared to the template by sum-squared error. Using this analysis, the dissimilarity index increased as ΔF increased ($\chi^2(8,46) = 200.47, p < 0.0001$) but, across all channels from all patients, did not differ based on percept alone ($W_+ = 1076, p = 0.33$), confirming a significant main effect of ΔF and lack of a significant main effect of percept.

4.3.5 Gamma Power Analysis

Two sets of triplet-locked high-gamma (80-190 Hz) power waveforms were constructed using either (i) wavelet transforms or (ii) analytic signal methods (see Methods). These waveforms were subjected to the same Monte Carlo permutation statistics as the triplet-locked evoked potentials to test for effects of either ΔF or percept. No significant effects were found (Figure 2 in appendix).

4.4 Discussion

Combining a classical behavioral paradigm using long sequences of tones alternating in frequency and direct cortical recordings in humans, the present results demonstrate a widespread set of brain areas - mainly in posterosuperior temporal and peri-rolandic cortex, but also extending to the middle temporal gyrus as well as inferior and middle frontal gyri - putatively involved in auditory streaming. EP amplitude tightly correlated with ΔF, but did not consistently differ based on perceptual organization alone. Waveform morphology was highly variable within and across brain areas, suggestive of their having different roles in auditory stream formation.

4.4.1 Complex meso-scale activity in the auditory cortex during streaming

Results from previous M/EEG (Gutschalk et al., 2005, 2007; Snyder et al., 2006) and fMRI (Gutschalk et al., 2007; Wilson et al., 2007) studies of streaming have suggested either a uniform role for the whole of the auditory cortex in stream formation or that the majority of activity in response to stimuli similar to those used in the present study is localized on the superior temporal plane (either on Heschl's gyrus or just posterior to it). The results from the present study demonstrate that, in addition to there being responses in higher auditory areas (i.e., lateral STG), the activity within a given macroscopic brain area is not uniform, a result that has also been noted by other investigators using evoked responses from iEEG with other classic auditory paradigms (Crone et al., 2001; Howard et al., 2000; Brugge et al., 2003, 2008; Edwards et al., 2005, 2009), especially from presumed higher-order auditory areas such as that on the lateral STG. This can be seen in the single-subject data shown in Figure 4, where the responses in adjacent electrode sites (e.g. G14 and G15 on the pSTG) indicate intra-areal variability in the response to the ABA- triplets.

This discrepancy may be due to several factors. First, the lead fields of the electrodes

used to measure brain activity in the present study are more likely to measure responses from gyral crowns than from sulcal sources such as those located on the superior temporal plane (the area to where non-invasive studies have localized dipoles during streaming), although others have reported iEEG potentials interpreted to arise from sulci (Edwards et al., 2005; Acar et al., 2009; Whitmer et al., 2010). We observed little evidence for sources on the STP in that (i) there were very rarely clear polarity reversals across the lateral fissure and (ii) the earliest peak in the average response to sequence onset was > 50ms, later than the earliest response in the medial portion of the transverse gyrus of Heschl, which occurs at < 25ms (Liegeois-Chauvel et al., 1991). This last point does not preclude the possibility that some of the responses we measured arose from *lateral* portions of the STP, particularly in the N1-latency range (Gutschalk et al., 2005; Snyder et al., 2006). However, to us, this seems unlikely given point (i). Second, the responses we observed from the lateral STG could have radial source orientations, which would not be identified with MEG but could be with EEG. Indeed, Snyder and colleagues (2006) reported radially oriented sources which could have been localized to the STG. Third, although both aforementioned fMRI studies of streaming – as well as others (Deike et al., 2010) – reported activation maps with multiple foci of activation, the complex relationship between auditory-evoked responses and the fMRI BOLD signal (Mukamel et al., 2005; Mayhew et al., 2010; Mulert et al., 2010, 2005; Gutschalk et al., 2010; Steinmann and Gutschalk, 2011) as well as BOLD-fMRI's low temporal resolution precludes a detailed characterization of areal sub-specialization. Fourth, and perhaps most likely, the activity recorded by EEG and, to a lesser extent, MEG, represents a spatially-smoothed version of the true cortical source configuration (Ahlfors et al., 2010; Halgren, 2004), and does not tend to see brain activity having response variability with high spatial frequency, contrary to the locally-generated signals measured by

intracranial EEG.

4.4.2 The role of extra-auditory areas in streaming

The present study is the first to report brain activity from extra-auditory cortical areas with high temporal resolution during the streaming paradigm. As shown in Figures 5 and 6, evoked potentials from several widespread brain areas correlated with ΔF. Waveform morphology was spatially variable both across and *within* macroscopic brain areas (though consistent across trials), even within individual participants, suggesting that (1) areas outside the auditory cortex may play an as-yet undetermined role in streaming and (2) the role of a given macroscopic brain area may not be uniform, known issues of EP variability notwithstanding (e.g. Edwards et al., 2009).

While several authors have posited a role for areas outside the classically-defined auditory pathway in streaming (Snyder and Alain, 2007b; Bidet-Caulet and Bertrand, 2009; Elhilali et al., 2009; Micheyl et al., 2005), nearly all neurophysiological studies of streaming have focused exclusively on the auditory cortex (but see (Cusack, 2005; Kondo and Kashino, 2009; Pressnitzer et al., 2008)). Only two previous studies examined whole-brain activity during the streaming paradigm (Cusack, 2005; Kondo and Kashino, 2009).

Cusack (2005), using a perceptually-bistable sequence of tones similar to those used in the present study, reported increased BOLD activity in the intraparietal sulcus during 2-stream vs. 1-stream percepts, but did not report percept or ΔF-based differences in the auditory cortex. The present study could not assess the intraparietal sulcus given that (i) the sub-dural electrodes used were confined to superficial gyri and (ii) the lead field of sub-dural electrodes is unlikely to measure activity from as deep in the sulcus as the foci reported by Cusack. Studies utilizing methods with high temporal resolution (e.g. MEG, iEEG, or microelectrodes in experimental

animals) focusing on this region could elucidate it's precise role in streaming and auditory perceptual organization more generally [e.g. (Rauschecker and Scott, 2009; Teki et al., 2011)]. Given the results of the present study as well as previous work (Bee and Klump, 2004, 2005; Bee et al., 2010; Fishman et al., 2004, 2001; Micheyl et al., 2005; Gutschalk et al., 2005; Wilson et al., 2007; Snyder et al., 2006), it is unclear why Cusack did not observe a neurophysiological correlate of ΔF in the auditory cortex, though an account based on subtle paradigmatic differences cannot be ruled out.

Kondo and Kashino (2009) used an event-related fMRI paradigm in order to measure brain activity during perceptual switching. Their subjects listened to tone sequences nearly identical to those used in the present study and indicated when the percept switched from one to two streams and *vice versa*. In addition to the auditory cortex, significant switch-related activations were found in the posterior insula, medial geniculate body, supra-marginal gyrus. No explicit contrasts were carried out to test for effects of perceptual organization or ΔF, but the results do highlight the need for further examination of the involvement of areas outside the auditory cortex in streaming.

Our results demonstrate that the cortical areas engaged during the streaming paradigm is much more complex and widespread than has been shown by previous work, and highlights the need for detailed neurophysiological examinations of the streaming paradigm in behavioral animal models.

4.4.3 Failure to observe correlates of bistability

Contrary to the study of the visual system in which there are many reports of brain activity covarying directly with perception (Leopold and Logothetis, 1999; Logothetis, 1998; Sterzer et al., 2009), such observations are scarce in the auditory system (Hillyard et al., 1971;

Cusack, 2005; Gutschalk et al., 2005, 2008). By recording brain activity with high spatiotemporal precision from widespread areas of the human cortex, the present study attempted to identify neural correlates of streaming, *per se*, in the absence of physical stimulus differences. As mentioned above, Cusack (2005) reported increased BOLD activity in the anterior intraparietal sulcus during 2- vs. 1-stream percepts but did not find percept- or ΔF-based differences in the auditory cortex. The latter finding is contrary to what Gutschalk et al. (2005) reported using magnetoencephalography, namely amplitudes of the P_1m and N_1m components evoked by the B-tone in a sequence of ABA- triplets which co-varied with both ΔF and perceptual organization, *per se*. No evidence for activity in the intraparietal sulcus was found in that study, though this could be due to activity in the Cusack study not being precisely time-locked to the stimuli, a condition necessary for the measurement of evoked responses with EEG or MEG. Neither finding – increased activity in the intraparietal sulcus or planum temporale during 2- vs. 1-stream percepts – was replicated by the present study, possibly due to lack of coverage in the areas of activity reported by both Cusack and Gutschalk et al. (intraparietal sulcus, transverse gyrus on the superior temporal plane) or, again, that the electrical activity responsible for the generation of the BOLD effects reported by Cusack was not time-locked to the stimuli.

Possible explanations for why we did not observe robust correlates of perceptual bistability despite widespread cortical sampling (see Figure 3) are many. First, although it seems unlikely to us given the large amount of data suggesting a role for frontal areas in conscious visual perception (Libedinsky and Livingstone, 2011), it could be that the areas reported by Cusack (2005) and Gutschalk et al. (2006) are unique in maintaining representations of auditory perceptual organization and that we simply were unable to examine activity from these areas.

Second, although the possibility that the known issue of trial-to-trial variability in the evoked potentials caused the lack of a significant percept-based finding cannot be ruled out, we find this explanation unlikely given the robust effects of ΔF as well as the relatively flat waveforms in the pre-sequence baseline period we observed. Finally, the neural correlates of auditory streaming could be found (i) in another cortical area not sampled, (ii) in a distributed network of brain areas which couldn't be determined based on the uni-variate analyses used, (iii) on a finer spatial scale than was assessed by the present study, or (iv) in an aspect of neural activity not examined such as sustained potentials or sustained gamma-band activity, though our analysis of evoked gamma-band power showed neither ΔF- or percept-based effects. This is perhaps due to the relatively constant acoustic stimulation used in our paradigm vs. the less frequent stimuli used in previous reports demonstrating large gamma-band effects (Crone et al., 2001, 2006; Edwards et al., 2005).

4.5 References

Acar, Z. A., Worrell, G., and Makeig, S. (2009). Patch-basis electrocortical source imaging in epilepsy. *Conf Proc IEEE Eng Med Biol Soc* 2009, 2930-2933.

Ahlfors, S. P., Han, J., Lin, F.-H., Witzel, T., Belliveau, J. W., Hämäläinen, M. S., and Halgren, E. (2010). Cancellation of EEG and MEG signals generated by extended and distributed sources. *Hum Brain Mapp* 31, 140-149.

Almonte, F., Jirsa, V. K., Large, E. W., and Tuller, B. (2005). Integration and segregation in auditory streaming. *Physica D: Nonlinear Phenomena* 212, 137-159.

Anstis, S., and Saida, S. (1985). Adaptation to auditory streaming of frequencymodulatedtones. *Journal of Experimental Psychology. Human Perception and Performance* 11, 257-271.

Bee, M. A., and Klump, G. M. (2005). Auditory stream segregation in the songbird forebrain: effects of time intervals on responses to interleaved tone sequences. *Brain Behav. Evol* 66, 197-214.

Bee, M. A., and Klump, G. M. (2004). Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. *J. Neurophysiol* 92, 1088-1104.

Bee, M. A., Micheyl, C., Oxenham, A. J., and Klump, G. M. (2010). Neural adaptation to tone sequences in the songbird forebrain: patterns, determinants, and relation to the build-up of auditory streaming. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol* 196, 543-557.

Bell, A. J., and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput* 7, 1129-1159.

Bidet-Caulet, A., and Bertrand, O. (2009). Neurophysiological mechanisms involved in auditory perceptual organization. *Front Neurosci* 3, 182-191.

Bregman, A. S. (1994). *Auditory scene analysis: the perceptual organization of sound*. MIT Press.

Brugge, J. F., Volkov, I. O., Garell, P. C., Reale, R. A., and Howard, M. A. (2003). Functional connections between auditory cortex on Heschl's gyrus and on the lateral superior temporal gyrus in humans. *J. Neurophysiol* 90, 3750-3763.

Brugge, J. F., Volkov, I. O., Oya, H., Kawasaki, H., Reale, R. A., Fenoy, A., Steinschneider, M., and Howard, M. A. (2008). Functional localization of auditory cortical fields of human: click-train stimulation. *Hear. Res* 238, 12-24.

Carlyon, R. P., Cusack, R., Foxton, J. M., and Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *J Exp Psychol Hum Percept Perform* 27, 115-127.

Carlyon, R. P. (2004). How the brain separates sounds. *Trends Cogn. Sci. (Regul. Ed.)* 8, 465-471.

Crone, N. E., Boatman, D., Gordon, B., and Hao, L. (2001). Induced electrocorticographic gamma activity during auditory perception. Brazier Award-winning article, 2001. *Clin Neurophysiol* 112, 565-582.

Crone, N. E., Sinai, A., and Korzeniewska, A. (2006). High-frequency gamma oscillations and human brain mapping with electrocorticography. *Prog. Brain Res* 159, 275-295.

Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *J Cogn Neurosci* 17, 641-651.

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9, 179-194.

Deco, G., and Romo, R. (2008). The role of fluctuations in perception. *Trends Neurosci* 31, 591-598.

Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M., and Friston, K. (2008). The dynamic brain: from spiking neurons to neural masses and cortical fields. *PLoS Comput. Biol* 4, e1000092.

Deike, S., Gaschler-Markefski, B., Brechmann, A., and Scheich, H. (2004). Auditory stream segregation relying on timbre involves left auditory cortex. *Neuroreport* 15, 1511-1514.

Deike, S., Scheich, H., and Brechmann, A. (2010). Active stream segregation specifically involves the left human auditory cortex. *Hear. Res* 265, 30-37.

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9-21.

Delorme, A., Sejnowski, T., and Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage* 34, 1443-1449.

Edwards, E., Soltani, M., Deouell, L. Y., Berger, M. S., and Knight, R. T. (2005). High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. *J. Neurophysiol* 94, 4269-4280.

Edwards, E., Soltani, M., Kim, W., Dalal, S. S., Nagarajan, S. S., Berger, M. S., and Knight, R. T. (2009). Comparison of time-frequency responses and the event-related potential to auditory speech stimuli in human cortex. *J. Neurophysiol* 102, 377-386.

Elhilali, M., and Shamma, S. A. (2008). A cocktail party with a cortical twist: how cortical mechanisms contribute to sound segregation. *J. Acoust. Soc. Am* 124, 3751-3771.

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J., and Shamma, S. A. (2009). Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron* 61, 317-329.

Engel, A. K., Moll, C. K. E., Fried, I., and Ojemann, G. A. (2005). Invasive recordings from the human brain: clinical insights and beyond. *Nat. Rev. Neurosci* 6, 35-47.

Ernst, M. D. (2004). Permutation Methods: A Basis for Exact Inference. *Statist. Sci.* 19, 676-685.

Fay, R. R. (1998). Auditory stream segregation in goldfish (Carassius auratus). *Hear. Res* 120, 69-76.

Fay, R. R. (2000). Spectral contrasts underlying auditory stream segregation in goldfish (Carassius auratus). *J. Assoc. Res. Otolaryngol* 1, 120-128.

Fischl, B., Sereno, M. I., and Dale, A. M. (1999a). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9, 195-207.

Fischl, B., Sereno, M. I., Tootell, R. B., and Dale, A. M. (1999b). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp* 8, 272-284.

Fishman, Y. I., Reser, D. H., Arezzo, J. C., and Steinschneider, M. (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear. Res* 151, 167-187.

Fishman, Y. I., Arezzo, J. C., and Steinschneider, M. (2004). Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *J. Acoust. Soc. Am* 116, 1656-1670.

Gigante, G., Mattia, M., Braun, J., and Del Giudice, P. (2009). Bistable perception modeled as competing stochastic integrations at two levels. *PLoS Comput. Biol* 5, e1000430.

Gutschalk, A., Hämäläinen, M. S., and Melcher, J. R. (2010). BOLD responses in human auditory cortex are more closely related to transient MEG responses than to sustained ones. *J. Neurophysiol* 103, 2015-2026.

Gutschalk, A., Micheyl, C., and Oxenham, A. J. (2008). Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biol* 6, e138.

Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., and Oxenham, A. J. (2005). Neuromagnetic correlates of streaming in human auditory cortex. *J. Neurosci* 25, 5382-5388.

Gutschalk, A., Oxenham, A. J., Micheyl, C., Wilson, E. C., and Melcher, J. R. (2007). Human cortical activity during streaming without spectral cues suggests a general neural substrate for auditory stream segregation. *J. Neurosci* 27, 13074-13081.

Halgren, E. (2004). How can intracranial recordings assist MEG source localization? *Neurol Clin Neurophysiol* 2004, 86.

Hillyard, S. A., Squires, K. C., Bauer, J. W., and Lindsay, P. H. (1971). Evoked potential correlates of auditory signal detection. *Science* 172, 1357-1360.

Howard, M. A., Volkov, I. O., Mirsky, R., Garell, P. C., Noh, M. D., Granner, M., Damasio, H., Steinschneider, M., Reale, R. A., Hind, J. E., et al. (2000). Auditory cortex on the human posterior superior temporal gyrus. *J. Comp. Neurol* 416, 79-92.

Itatani, N., and Klump, G. M. (2009). Auditory streaming of amplitude-modulated sounds in the songbird forebrain. *J. Neurophysiol* 101, 3212-3225.

Itatani, N., and Klump, G. M. (2010). Neural correlates of auditory streaming of harmonic complex sounds with different phase relations in the songbird forebrain. *J. Neurophysiol.* Available at: http://www.ncbi.nlm.nih.gov/pubmed/21068270 [Accessed December 23, 2010].

Kanwal, J. S., Medvedev, A. V., and Micheyl, C. (2003). Neurodynamics for auditory stream segregation: tracking sounds in the mustached bat's natural environment. *Network* 14, 413-435.

Kondo, H. M., and Kashino, M. (2009). Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *J. Neurosci* 29, 12695-12701.

Leopold, and Logothetis (1999). Multistable phenomena: changing views in perception. *Trends Cogn. Sci. (Regul. Ed.)* 3, 254-264.

Libedinsky, C., and Livingstone, M. (2011). Role of prefrontal cortex in conscious visual perception. *J. Neurosci* 31, 64-69.

Liegeois-Chauvel, C., Musolino, A., and Chauvel, P. (1991). Localization of the primary auditory area in man. *Brain* 114 (Pt 1A), 139-151.

Logothetis, N. K. (1998). Single units and conscious vision. *Philos. Trans. R. Soc. Lond., B, Biol. Sci* 353, 1801-1818.

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177-190.

Mayhew, S. D., Dirckx, S. G., Niazy, R. K., Iannetti, G. D., and Wise, R. G. (2010). EEG signatures of auditory activity correlate with simultaneously recorded fMRI responses in humans. *Neuroimage* 49, 849-864.

Micheyl, C., Carlyon, R. P., Gutschalk, A., Melcher, J. R., Oxenham, A. J., Rauschecker, J. P., Tian, B., and Courtenay Wilson, E. (2007). The role of auditory cortex in the formation of auditory streams. *Hear. Res* 229, 116-131.

Micheyl, C., Tian, B., Carlyon, R. P., and Rauschecker, J. P. (2005). Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron* 48, 139-148.

Miller, G., and Heise, G. (1950). The Trill Threshold. *The Journal of the Acoustical Society of America* 22, 637-638.

Moreno-Bote, R., Rinzel, J., and Rubin, N. (2007). Noise-induced alternations in an attractor network model of perceptual bistability. *J. Neurophysiol* 98, 1125-1139.

Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., and Malach, R. (2005). Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. *Science* 309, 951-954.

Mulert, C., Leicht, G., Hepp, P., Kirsch, V., Karch, S., Pogarell, O., Reiser, M., Hegerl, U., Jäger, L., Moller, H. J., et al. (2010). Single-trial coupling of the gamma-band response and the corresponding BOLD signal. *Neuroimage* 49, 2238-2247.

Mulert, C., Jäger, L., Propp, S., Karch, S., Störmann, S., Pogarell, O., Möller, H.-J., Juckel, G., and Hegerl, U. (2005). Sound level dependence of the primary auditory cortex: Simultaneous measurement with 61-channel EEG and fMRI. *Neuroimage* 28, 49-58.

Nelken, I., and Bar-Yosef, O. (2008). Neurons and objects: the case of auditory cortex. *Front Neurosci* 2, 107-113.

van Noorden, L. (1975). Temporal coherence in the perception of tone sequences.

Pressnitzer, D., Sayles, M., Micheyl, C., and Winter, I. M. (2008). Perceptual organization of sound begins in the auditory periphery. *Curr. Biol* 18, 1124-1128.

Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci* 12, 718-724.

Schadwinkel, S., and Gutschalk, A. (2010a). Activity Associated with Stream Segregation in Human Auditory Cortex is Similar for Spatial and Pitch Cues. *Cereb Cortex*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20237241 [Accessed September 14, 2010].

Schadwinkel, S., and Gutschalk, A. (2010b). Functional dissociation of transient and sustained fMRI BOLD components in human auditory cortex revealed with a streaming paradigm based on interaural time differences. *Eur. J. Neurosci* 32, 1970-1978.

Schul, J., and Sheridan, R. A. (2006). Auditory stream segregation in an insect. *Neuroscience* 138, 1-4.

Shamma, S. A., and Micheyl, C. (2010). Behind the scenes of auditory perception. *Curr. Opin. Neurobiol* 20, 361-366.

Shamma, S. A., Elhilali, M., and Micheyl, C. (2010). Temporal coherence and attention in

auditory scene analysis. *Trends Neurosci*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21196054 [Accessed January 27, 2011].

Shpiro, A., Moreno-Bote, R., Rubin, N., and Rinzel, J. (2009). Balance between noise and adaptation in competition models of perceptual bistability. *J Comput Neurosci* 27, 37-54.

Snyder, J. S., Alain, C., and Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *J Cogn Neurosci* 18, 1-13.

Snyder, J. S., and Alain, C. (2007a). Sequential auditory scene analysis is preserved in normal aging adults. *Cereb. Cortex* 17, 501-512.

Snyder, J. S., and Alain, C. (2007b). Toward a neurophysiological theory of auditory stream segregation. *Psychol Bull* 133, 780-799.

Steinmann, I., and Gutschalk, A. (2011). Potential fMRI correlates of 40-Hz phase locking in primary auditory cortex, thalamus and midbrain. *Neuroimage* 54, 495-504.

Sterzer, P., Kleinschmidt, A., and Rees, G. (2009). The neural bases of multistable perception. *Trends Cogn. Sci. (Regul. Ed.)* 13, 310-318.

Sussman, E., Ritter, W., and Vaughan, H. G. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology* 36, 22-34.

Teki, S., Chait, M., Kumar, S., von Kriegstein, K., and Griffiths, T. D. (2011). Brain bases for auditory stimulus-driven figure-ground segregation. *J. Neurosci* 31, 164-171.

Whitmer, D., Worrell, G., Stead, M., Lee, I. K., and Makeig, S. (2010). Utility of independent component analysis for interpretation of intracranial EEG. *Front Hum Neurosci* 4, 184.

Wilson, E. C., Melcher, J. R., Micheyl, C., Gutschalk, A., and Oxenham, A. J. (2007). Cortical FMRI activation to sequences of tones alternating in frequency: relationship to perceived rate and streaming. *J. Neurophysiol* 97, 2230-2238.

Winkler, I., Denham, S. L., and Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn. Sci. (Regul. Ed.)* 13, 532-540.

4.6 Appendix

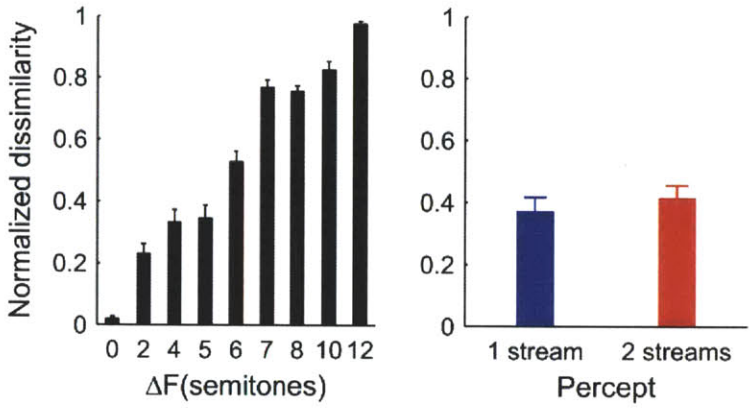

Figure 4.S1. Complementary dissimilarity index.

Dissimilarity index computed using the grand-average EP collapsed across all conditions as the template and the per-condition average EPs as the test waveforms. As in Figure 7, the left panel shows the dissimilarity index as a function of frequency separation collapsed across percept. The right panel shows the dissimilarity index as a function of percept collapsed across delta-F conditions in which the percept was bistable.

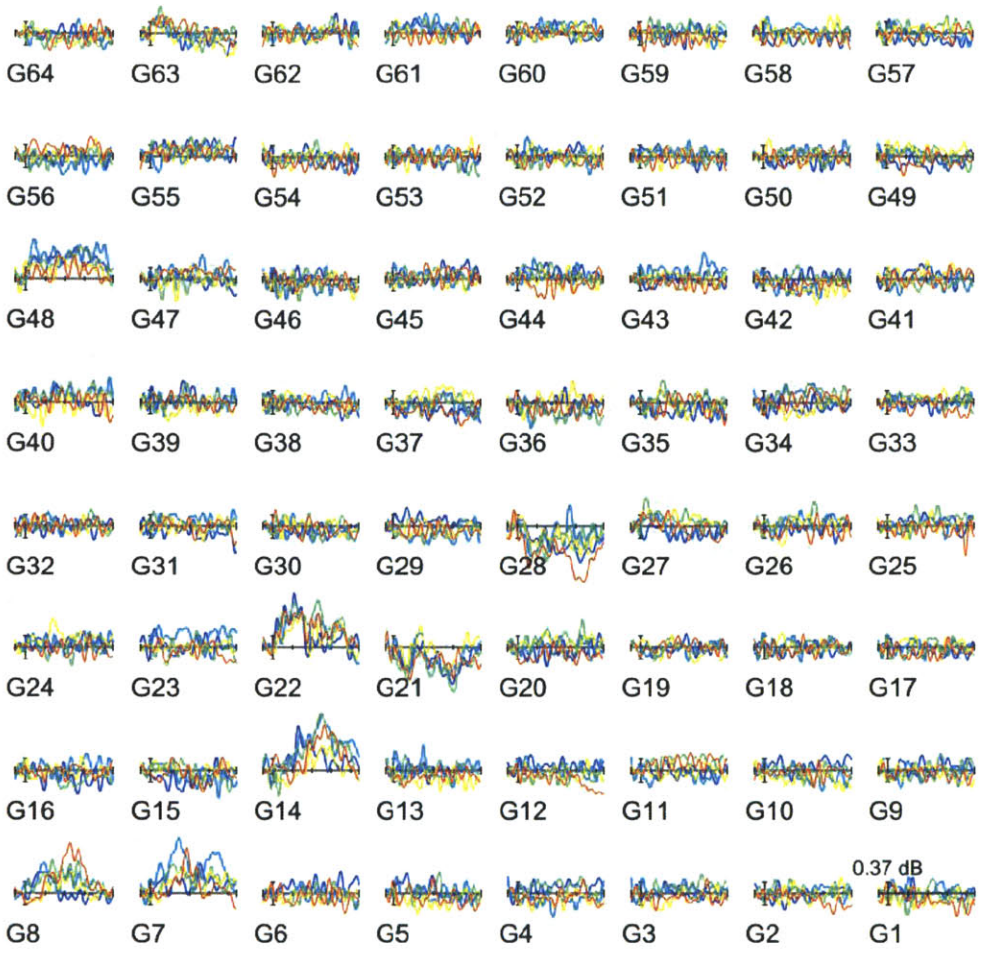

Figure 4.S2. High-gamma power waveforms.

High-gamma (70-190 Hz) power waveforms for the same subject as in Figure 4. The legend is as in Figure 4 except for ordinate units, which are now in dB with respect to the pre-stimulus baseline period.

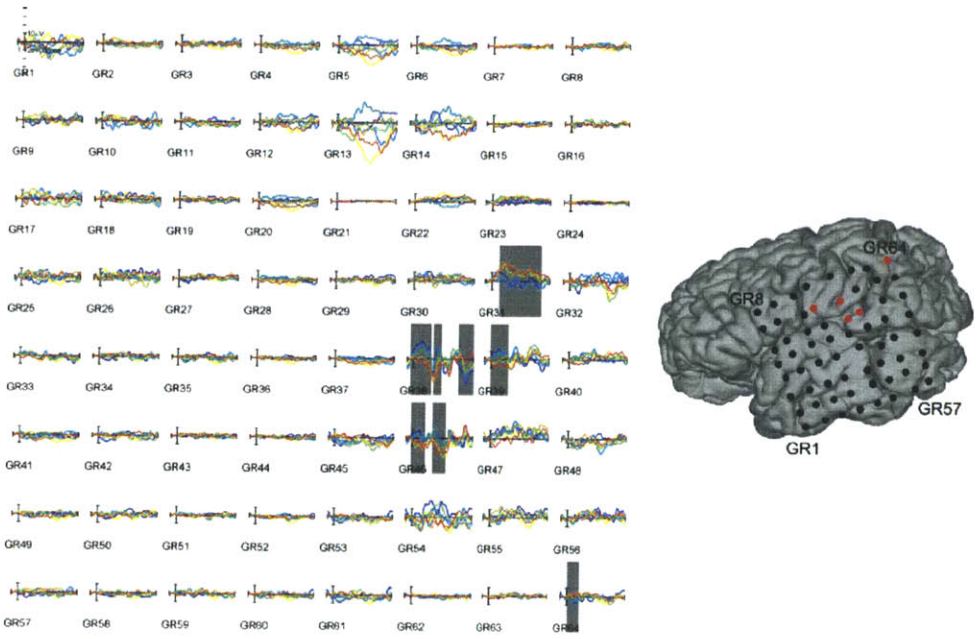

Figure 4.S3. Evoked potentials from S8 across the seven ΔF conditions used for this participant.

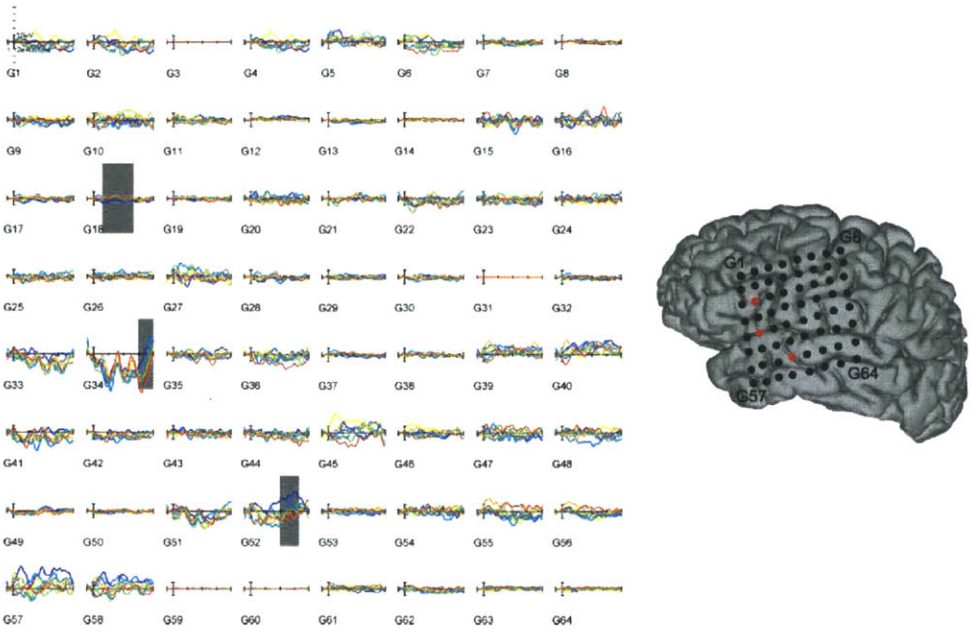

Figure 4.S4. Evoked potentials from S7 across the seven ΔF conditions used for this participant.

Chapter 5: Multiple Timescales of Acoustic Segmentation Recorded Directly from Human Cortex

Andrew R. Dykstra[1,2], Eric Halgren[3], Thomas Thesen[4], Chad E. Carlson[4], Werner Doyle[4], Joseph R. Madsen[5], Emad Eskandar[6], Sydney S. Cash[2]

*This chapter is in preparation for submission to *The Journal of Neuroscience*.

1Program in Speech and Hearing Bioscience and Technology, Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, **2**Cortical Physiology Laboratory, Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, **3**Departments of Radiology and Neurosciences, University of California San Diego, San Diego, CA, **4**Comprehensive Epilepsy Center, New York University School of Medicine, New York, NY, **5**Department of Neurosurgery, Childrens Hospital Boston and Harvard Medical School, Boston, MA, **6**Department of Neurosurgery, Massachusetts General Hospital and Harvard Medical School

ABSTRACT

Auditory perception is organized on multiple timescales, ranging from milliseconds (pitch), tens of milliseconds (flutter), hundreds of milliseconds (acoustic events or rhythm) to much longer timescales which support melody perception, sentence comprehension, and auditory memory. How the auditory system segments the acoustic environment to support organization on these multiple timescales is unclear. In the visual system, recent evidence suggests that cortex organizes it's sensory input into increasingly-longer timescales in a hierarchical manner such that relatively fast timescales are more strongly represented in early unimodal cortex and slower timescales in higher-order association cortex and/or multi- and supra-modal areas. We recorded the intracranial EEG from nine patients with epilepsy while they listened to sequences of alternating-frequency pure tones. Each tone sequence elicited a clear rhythmic percept and lasted between 6.5 and 10 seconds. Simultaneously-recorded evoked responses from widespread brain areas showed organization on multiple temporal scales. Some sites responded only to the onset and offset of sound, while other sites showed either sustained or steady-state responses throughout the duration of the stimulus. In general, sites which showed sustained or steady-state responses were located near the posterior superior temporal gyrus, while sites which showed onset or offset responses were outside this region. The results support the idea of a hierarchical segmentation of acoustic input on progressively longer temporal scales.

5.1 Introduction

The natural acoustic environment is organized on several timescales, ranging from milliseconds for auditory percepts such as pitch to hundreds of milliseconds for rhythm to longer timescales for melody, speech comprehension, and auditory memory. How – and where in the auditory system – this segmentation is performed is not well understood. It has long been proposed that the anatomical hierarchy of the auditory system, from the auditory nerve to the auditory cortex could support the perceptual organization of the acoustic world into progressively-longer temporal scales (Seifritz et al., 2002; Harms et al., 2005; Harms and Melcher, 2002; Kiebel et al., 2008; Hasson et al., 2008; Lerner et al., 2011). Indeed, the discrepancy of timescales of neural adaptation observed between auditory centers located early in the ascending pathway (i.e. the auditory nerve, cochlear nucleus, and inferior colliculus) vs. that which occurs at higher auditory centers (medial geniculate body and the auditory cortex) (Harms and Melcher, 2002), in addition to the degradation of phase locking as one ascends the auditory pathway (Creutzfeldt et al., 1980; Langner and Schreiner, 1988; Schreiner and Langner, 1988; Langner, 1992), has been interpreted as evidence for a neural instantiation of the hierarchical organization of acoustic time (Kiebel et al., 2008; Harms et al., 2005).

In humans, evidence for such discrepancies in the timescales of adaptation has relied exclusively on studies utilizing non-invasive methods, including fMRI and M/EEG, which have focused primarily on the auditory cortex [but see (Harms and Melcher, 2002)]. However, inherent limitations such as, in the case of fMRI, the indirect relationship between neural activity and the BOLD signal and, in the case of M/EEG, the lack of unique inverse solution, render it difficult to use those techniques to examine neural correlates of acoustic temporal segmentation across widespread brain areas simultaneously. Direct evidence for the neural differentiation of

acoustic timescales remains sparse, especially in cortical areas outside the auditory cortex.

Here, we report the results from a series of experiments in which neurosurgical patients with epilepsy actively listened to long (on the order of several seconds) sequences of pure tones while we simultaneously recorded their intracranial electroencephalogram (iEEG). Each participant was engaged in an auditory streaming task and indicated at the end of each acoustic sequence whether they perceived one or two acoustic "streams." Evoked responses, including low-frequency evoked potentials (EPs) and gamma-band activity (GBA), displayed organization on several timescales, ranging from (i) steady-state responses that followed the envelope of individual tones in the sequence, (ii) onset and/or offset responses that occurred only at the beginning or end of a sequence, or (iii) sustained responses in the GBA that persisted throughout the course of a sequence without an accompanying steady-state response. Most steady-state and sustained responses were found in the auditory cortex overlying the superior temporal gyrus (STG), while onset and offset responses without accompanying steady-state or sustained responses were most often found in electrodes outside the auditory cortex (i.e., not over the STG). The results support the idea that the anatomical hierarchy of the auditory system could recapitulate the temporal hierarchy found in natural acoustic signals.

5.2 Materials and Methods

5.2.1 Ethics statement

All procedures were approved by the Institutional Review Boards at Partners Healthcare (MGH and BWH), the New York University (NYU) Langone Medical Center, and the Massachusetts Institute of Technology (MIT) in accordance with NIH guidelines. Written informed consent was obtained from all patients prior to their participation.

5.2.2 Listeners

Twelve patients with intractable epilepsy underwent invasive monitoring in order to localize the epileptogenic zone prior to its surgical removal. Each patient was implanted with an array of sub-dural platinum-iridium electrodes embedded in silastic sheets (2.3mm exposed diameter, 10mm center-to-center spacing; *Ad-tech Medical, Racine, WI*) placed directly on the cortical surface. Prior to implantation, each patient underwent high-resolution T1-weighted MRI. Subsequent to implantation, patients implanted at Massachusetts General Hospital (MGH) and Brigham and Women's Hospital (BWH) underwent postoperative computerized tomography (CT); patients implanted at NYU underwent postoperative MRI. Electrode coordinates obtained from postoperative scans were co-registered with preoperative MRI and overlaid onto the patient's reconstructed cortical surface using Freesurfer (Dale et al., 1999; Fischl et al., 1999) as well as custom software written in MATLAB (*The Mathworks, Framingham, MA*). Electrode coordinates were then projected onto the Freesurfer average brain using a spherical registration between the individual's cortical surface and that of the Freesurfer average (Fischl et al. 1999b). The data from three patients were excluded from analysis due to excessive noise caused by technical malfunction. From the remaining nine patients, we sampled from nearly 700 sites in the left hemisphere and 250 sites in the right hemisphere (Figure 5.1A).

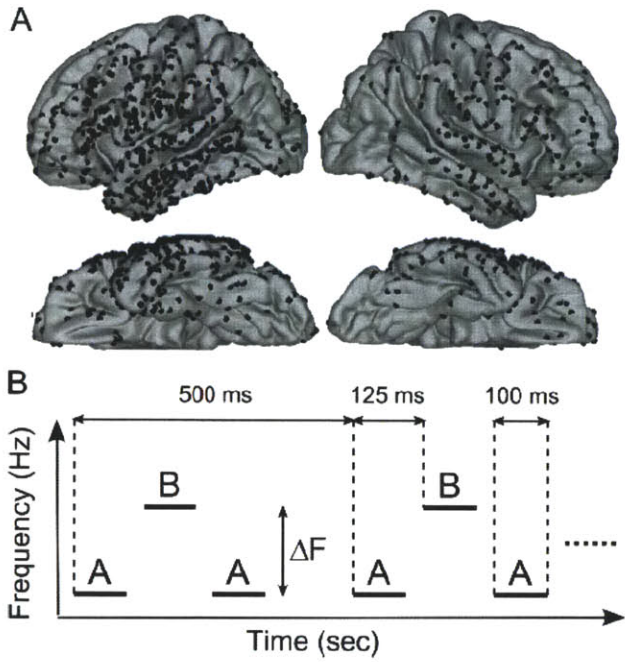

Figure 5.1. Recording sites across all nine patients overlaid onto the Freesurfer average brain (**A**) and a depiction of the stimuli used in the present study (**B**).

In (**A**), each black dot represents a single electrode from one of the nine patients. In total, we sampled from nearly 1000 sites, nearly 700 from the left hemisphere and more than 250 from the right hemisphere. In (**B**), each bar represents a short (100 msec) pure tone and the space between A and B tones is the frequency separation, or ΔF.

5.2.3 Stimuli and procedure

Stimuli were long sequences of pure-tone triplets of the form ABA-ABA-..., where A and B represent individual tones and the dash represents a silent gap (Figure 5.1B). Each tone was 100ms in duration with 10ms raised-cosine on- and off-ramps. The inter-stimulus interval (ISI) between the first A-tone and B-tone, as well as between the B-tone and second A-tone, was 25ms; the ISI between the second A-tone and subsequent triplet was 150ms. Stimulus onset asynchrony (SOA) between successive A-tones was 250ms; SOA between successive B-tones was 500ms; triplet onset asynchrony was also 500ms. Total duration of each sequence varied between 6.5 and 10 seconds depending on the listener (for P1-P5, duration varied between 6.5

and 7.5 seconds; for P6-P9, duration was 10 seconds). The B-tone frequency was fixed at 1kHz. The A-tone frequency varied between 0 and 12 semitones below the B-tone. Listeners P1, P2, P3, P4, and P5 participated in conditions in which the frequency separation was 0, 5, 6, 7, or 12 semitones. Listeners P6, P7, P8, and P9 participated in conditions in which the frequency separation was 0, 2, 4, 6, 8, 10, or 12 semitones. Each patient listened to between 10 and 45 sequences for a given frequency separation. All sounds were generated digitally in MATLAB, stored as wave files, and converted to analog waveforms by the on-board sound card of a laptop equipped with Presentation software (*Neurobehavioral Systems, Albany, CA*). Stimuli were presented at a comfortable listening level via Etymotic ER-2 insert earphones (*Etymotic Research, Inc., Elk Grove Village, IL*), diotically (when possible) or monaurally contralateral to the hemisphere of implantation. Patients were instructed to listen to the sounds and to indicate at the end of each sequence whether they heard a single "stream" comprised of all tones or two "streams," one comprised of A tones and the other of B tones. Responses were made by button press with a response box (*Cedrus Corporation, San Pedro, CA*) interfaced with Presentation via USB. Response windows were theoretically infinite, and the subsequent stimulus began one second after a response to the previous stimulus was entered.

5.2.4 Data acquisition

Intracranial EEG data at MGH and BWH were acquired with standard clinical EEG monitoring equipment (*XLTEK, Natus Medical Inc., San Carlos, CA*) at a sampling rate of 250 Hz (P1) or 500 Hz (P2,P3,P6,P8). At NYU, the intracranial EEG was acquired with a customized system at a sampling rate of 30 kHz (P4,P5,P7,P9). All data were subsequently re-sampled to 500 Hz for analysis. All data were referenced to either an inverted intracranial electrode (i.e. facing the inner skull table) remote from the electrodes of interest

(P1,P2,P3,P6,P8) or a screw bolted to the skull (P4,P5,P7,P9). For each patient, clinically-indicated, high-resolution T1-weighted structural MRI scans were acquired prior to surgery. High-resolution CT (P1,P2,P3,P6,P8) or structural MRI (P4,P5,P7,P9) scans were acquired subsequent to surgery for the purpose of electrode localization.

5.2.5 Data pre-processing

Intracranial EEG data were bandpass filtered offline between 0.5 and 190 Hz and notch filtered at 60 Hz and its harmonics using zero-phase shift FIR filters. Independent component analysis using the *runica* algorithm [69] in EEGLAB [70] was performed on the "raw" data. Components dominated by large artifacts were identified and removed by inspection. The component data were then back-projected in order to remove the artifacts from the original data.

5.2.6 Evoked-potential analysis

The intracranial EEG was epoched relative to the onset of sound sequences (yielding long epochs encompassing the entire sequence and shorter epochs of 0.5 seconds) as well as to the offset of sound sequences (yielding short epochs of 0.5 seconds) and either collapsed across all stimulus/response conditions or binned with respect to either ΔF or perceptual report within a given ΔF. Onset epochs were baseline-corrected with respect to either the 500 ms preceding sequence onset (for long epochs) or the 100 ms preceding sequence onset (for short epochs). Offset epochs were baseline-corrected with respect to the 100 ms preceding sequence offset.

5.2.7 Time-frequency analysis

Time-frequency analysis was performed using a combination of built-in Morlet wavelet functionality in EEGLAB including the *newtimef* function and custom MATLAB scripts. A measure termed the event-related spectral perturbation (ERSP) which compares post-stimulus

amplitudes across the frequency range of interest (here, 5-190 Hz) to that during a pre-stimulus baseline period (-500ms to 0ms, 0 being stimulus onset). ERSP was computed separately for each trial and subsequently averaged for stimulus and/or response condition of interest. The temporal width of the Morlet wavelets increased logarithmically proportional to frequency from three cycles at the lowest frequency examined (5 Hz) to 10 cycles at the highest (190 Hz), yielding approximate temporal resolution of < 500 ms at 8 Hz and < 125 ms in the gamma band. The ERSP for each trial was divided into pre-defined bands including alpha (7-13 Hz), beta (13-30 Hz), low gamma (30-80 Hz), and high gamma (80-190 Hz) and subsequently summed within each band to generate stimulus-locked power waveforms representing the total time-varying power in each band with respect to the pre-stimulus baseline. These waveforms were low-pass filtered using 4th-order zero-phase-shift elliptical filters at 2.5 or 10 Hz for analysis of sustained and onset/offset/steady-state responses, respectively.

5.2.8 Statistical analysis

A modified version of the cluster-based, non-parametric statistical procedure outlined by Maris and Oostenveld (Maris and Oostenveld, 2007) was used to test for effects of ΔF and bistability on EP amplitude and band-limited power waveforms, both time-locked either to the onset or offset of the sequence. Spearman (nonparametric) rank correlation (in the case of a multiple-level factor, e.g., ΔF) and unpaired t-test (in the case of two-level factors, e.g., percept) were used as the sample-level (i.e. individual time point within a single channel) statistics in order to evaluate possible effects of ΔF (five levels for P1-P5 and seven levels for P6-P9) and bistability (always two levels), respectively. Contiguous, statistically-significant samples (defined as $p<0.05$) within a single electrode were used to define the cluster-level statistic, which was computed by summing the sample-level statistics within a cluster. Statistical significance at

the cluster level was determined by computing a Monte Carlo estimate of the permutation distribution of cluster statistics using 1000 re-samples of the original data (Ernst, 2004). For multiple-level factors (ΔF), the estimate of the permutation distribution was performed by 1000 re-samples of the condition labels associated with each level in the factor. Within a single electrode, a cluster was taken to be significant if it fell outside the 95% confidence interval of the permutation distribution for that electrode. The determination of significant clusters was performed independently for each electrode. This method controls the overall false alarm rate within an electrode across time points; no correction for multiple comparisons was performed across electrodes.

5.2.9 Waveshape index

In order to quantify the propensity of a given electrode site to be dominated by either on/off responses or a more sustained response, we defined a waveshape index modified from Harms et al. (2005), varying between -1 and 1, where an index value of -1 indicates that a site is dominated by transient responses occurring at the onset or offset of sound stimuli and an index value of 1 indicates that a site is characterized more by sustained, ongoing responses. The index was computed per channel i as:

$$WI_i = \frac{1}{2}\left(\frac{max(ON_i, OFF_i)}{SS_i + mean(ON_i, OFF_i)} - 0.5\right) SNR_i$$

where ON_i and OFF_i, are the root-mean-squared (RMS) power in the average EP from 0-500 msec post-onset and post-offset, respectively, SS_i is the RMS power in the steady-state average EP, and SNR_i is the RMS power in the average EP from 0-500 msec post-onset divided by the RMS power in the average EP in the 500 msec pre-stimulus baseline period.

5.3 Results

5.3.1 Evoked potentials

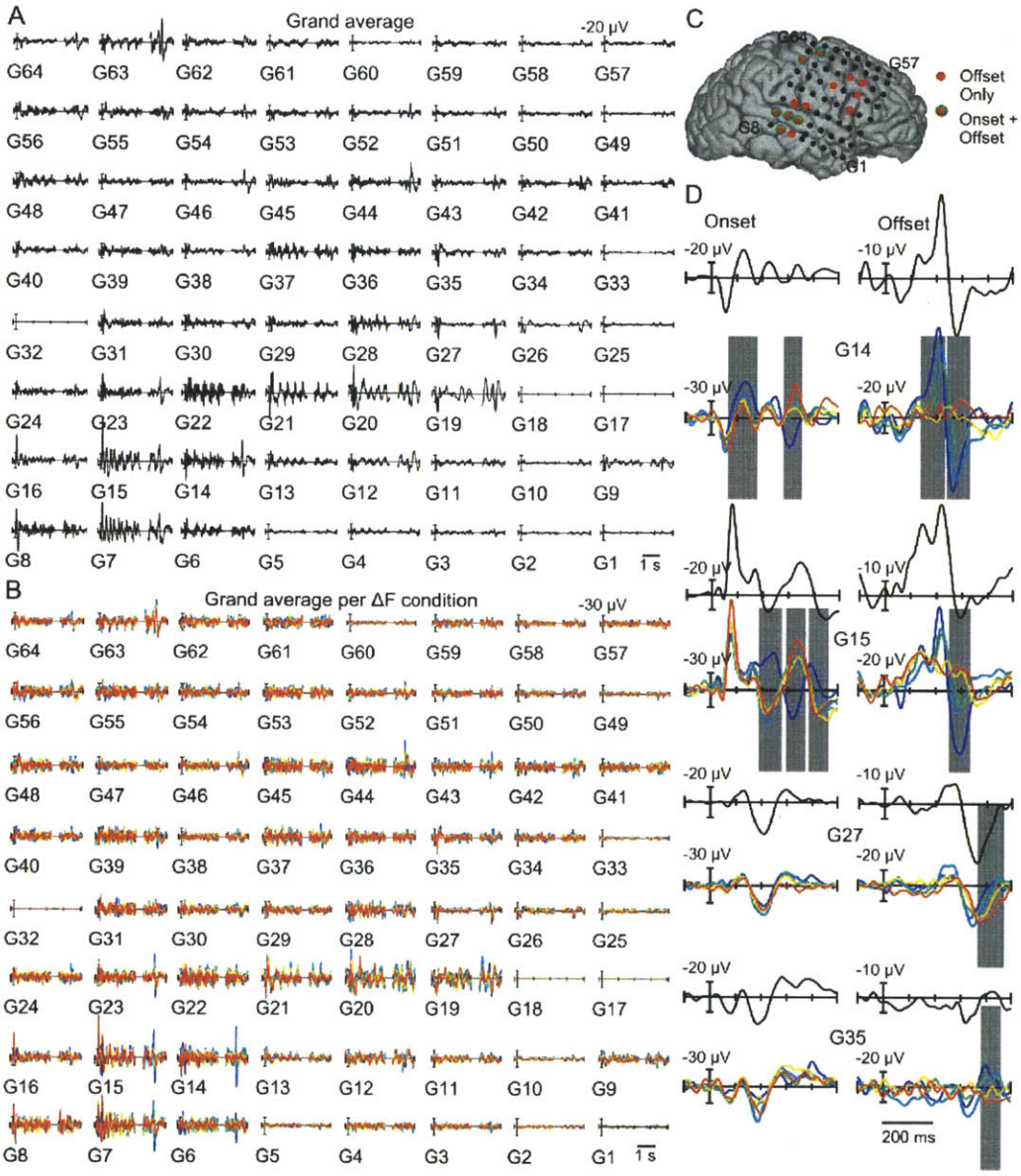

Figure 5.2. Example of evoked potential waveforms from a single subject demonstrating multiple timescales of responses.

(**A**) Grand average EPs collapsed across all stimulus conditions for each of the 64 electrodes shown in (**C**). (**B**) Average EPs per ΔF condition for each of the five conditions used for this participant (0, 5, 6, 7, and 12 semitones). The blue curves are for the 0-semitone condition, and the red curves are for the 12-semitone condition, with 5, 6, and 7 semitones shown as the colors in between

blue and red. (D) Examples of onset and offset responses from individual channels over temporal (G14, G15) and frontal (G27, G35) cortex. Gray shading indicates significant correlation between EP amplitude and ΔF condition. Several sites outside of the auditory cortex only showed correlations in the offset response.

Figure 5.2 shows an example of the evoked potentials from an individual participant (P4) during the duration of the entire sequence, locked both stimulus onset and, due to the sometimes variable length of each sequence, stimulus offset. Panel A shows the grand average evoked potential collapsed across all stimulus conditions. As can be seen from this figure, several brain areas showed responses to sound, with highly variable waveform morphology across individual electrode sites. Most sites near the posterior superior temporal gyrus (e.g. G14, G15, G22) tended to show responses throughout the duration of the sequence (i.e., steady-state responses). In contrast, most of the sites demonstrating exclusively on or off responses were located outside this area (e.g. G27, G35, G44, G46), although there are certainly counter examples in both directions, i.e. sites near the STG whose responses were dominated by on and off components (e.g. G23) and sites away from the STG whose responses displayed steady-state as well as on and off responses (e.g. G63).

Evoked responses were further subdivided based on the frequency separation between the A and B tones (panels B and the color panels in D). Several sites showed EP amplitude in the on and off responses which correlated with frequency separation, though there were several sites which only showed this correlation for off responses (e.g. G27 and G35, shown in panel D). The distribution of these sites for this patient is shown in Figure 5.2B.

5.3.2 Event-related spectral perturbation

Figure 5.3 shows an example of the time-frequency analysis we performed for each

electrode site. This site (G22) was taken from the same patient as in Figure 5.2, and is indicated by the red dot in Figure 5.3B. Figure 5.3A shows the average evoked response time-locked to sound onset and collapsed across all conditions. For each electrode site, the event-related spectral perturbation (ERSP) was computed by computing a wavelet-based time-frequency representation (see Materials and Methods) for every trial and subsequently averaging these representations across trials to calculate an average ERSP, shown in Figure 5.3C. Figure 5.3D shows the average post-stimulus power in the ERSP, which is a baseline-corrected measure, as a function of frequency. This electrode site showed a steady-state response around 8 Hz (the frequency of tone repetition), a widespread alpha/beta decrease extending even into the low gamma range, and a broadband high gamma increase. Power waveforms for these three bands, and their corresponding power spectra, are shown in panels E, F, and G, and H, I, and J, respectively. Note that all three bands of activity showed fluctuations around 2 Hz (the triplet frequency), but only the high gamma band showed fluctuations at 8 Hz (the individual tone frequency).

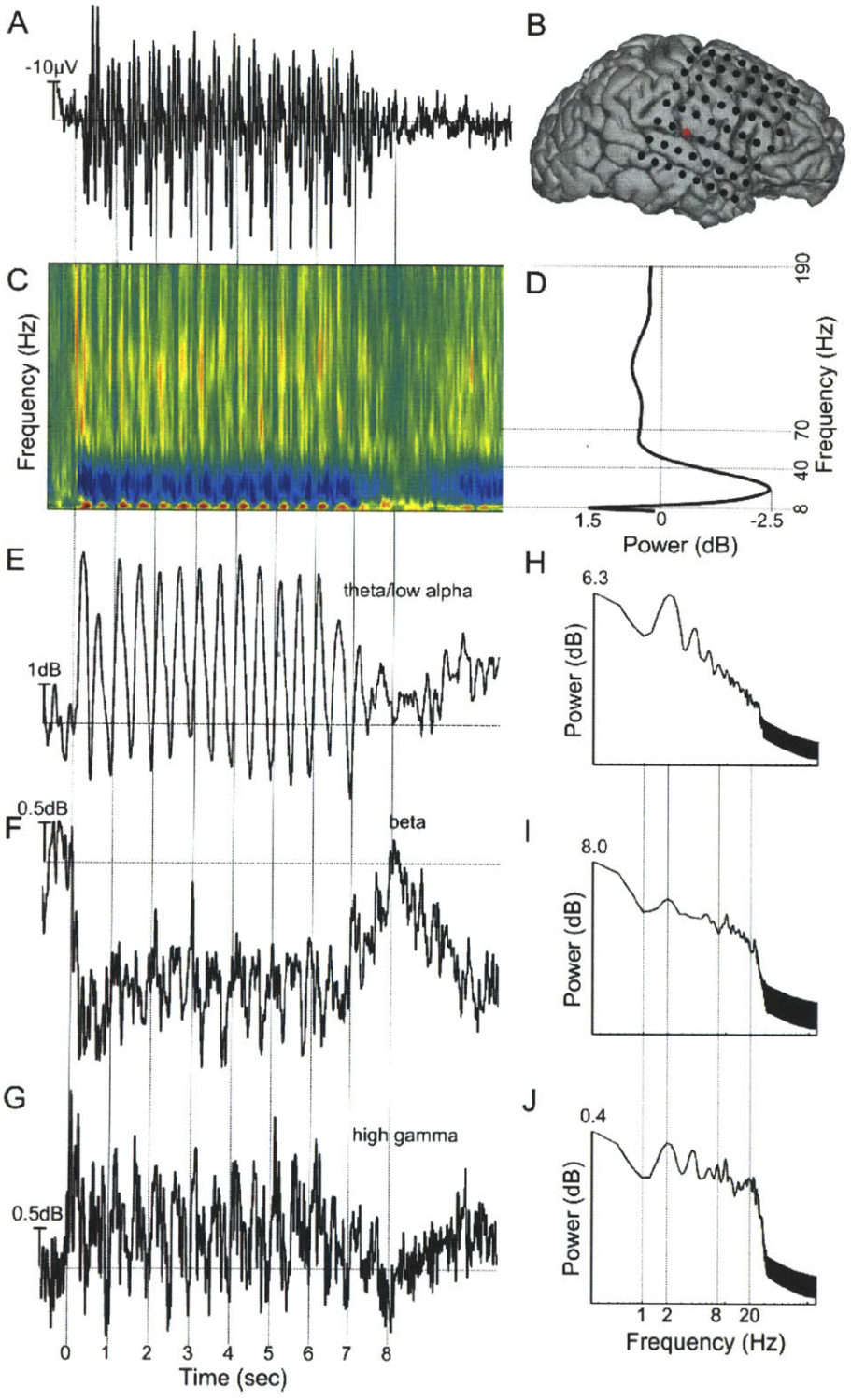

Figure 5.3. Example of the analysis we carried out for each channel.

(**A**) Grand-average EP across all conditions for the channel depicted in red in (**B**).

(**C**) The grand-average event-related spectral perturbation across time and frequency. Red (blue) colors indicates stimulus-driven increases (decreases) in band-limited power. The time averaged power at each frequency for this response is shown in (**D**). In both (**C**) and (**D**) it can be seen that there were three distinct bands of interest (theta/low alpha: 7-9 Hz, beta: 10-30 Hz, and high gamma: 70-190 Hz). The time courses of power modulations in each of these bands, and their associated power spectra, are shown in **E** and **H** (theta/low alpha), **F** and **I** (beta), and **G** and **J** (high gamma).

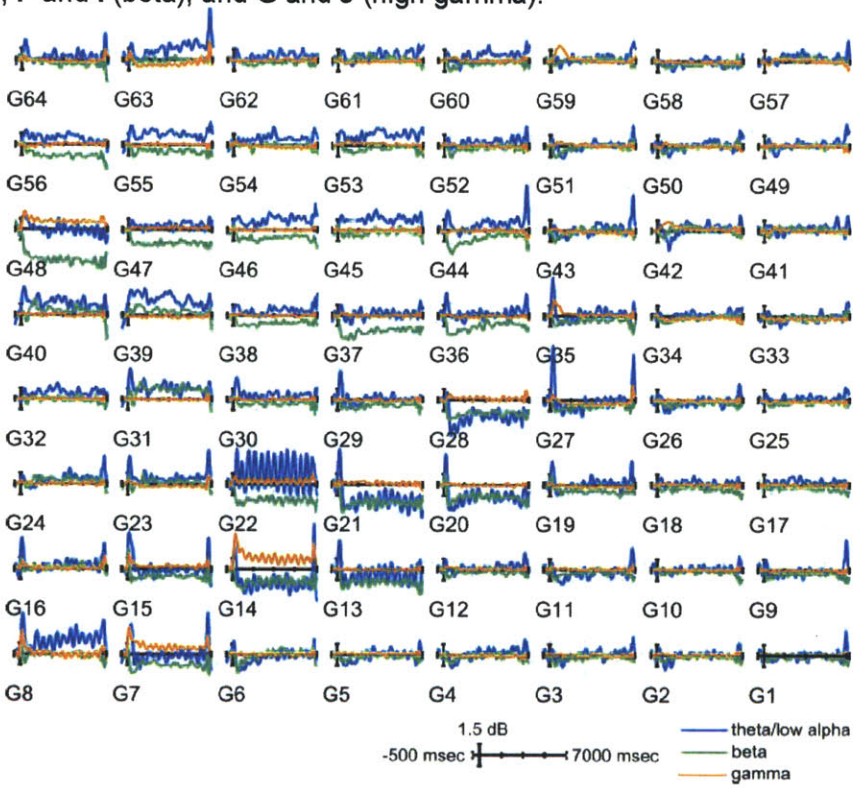

Figure 5.4. Waveforms of the three different frequency bands (theta/low alpha, beta, and high gamma) across all the electrode sites for the same patient as in Fig. 2. Theta/low alpha, beta, and high gamma are shown in blue, green an orange, respectively. It can be seen that each band displayed characteristic spatial patterns, with widespread beta reductions, focal high gamma activity, and alpha patterns which were either dominated by the EP (e.g. G22), or showed true oscillatory increases or decreases across space.

Due to prevalence of task correlates found in high gamma activity, we focused on that spectral region for the subsequent analysis of timescales of acoustic segmentation. Figure 5.4 shows the average power waveforms in several frequency bands (alpha, beta, high gamma) time-locked to stimulus onset and offset for the same patient as in Figure 5.2. As can be seen

from this figure, there were several profiles of high-gamma activity, varying from very sustained (e.g. G14) to highly phasic (e.g. G35). In general, the high-gamma activity corresponded well with the evoked responses in that if an electrode site showed a sustained (phasic) response, the high-gamma activity also was sustained (phasic). However, it can also be seen that the spatial extent of the sustained responses is more focal in the high-gamma activity than in the evoked responses. In contrast, beta activity (12-30 Hz) displayed widespread sustained decreases relative to the pre-stimulus baseline period, and theta/low alpha (7-9 Hz) activity, when not simply reflecting the evoked response (e.g. G22), displayed patterns of sustained increases (G39, G55) and decreases (e.g. G14, G21). Figure 5.5 shows the spatial patterns of activity for the evoked potentials and

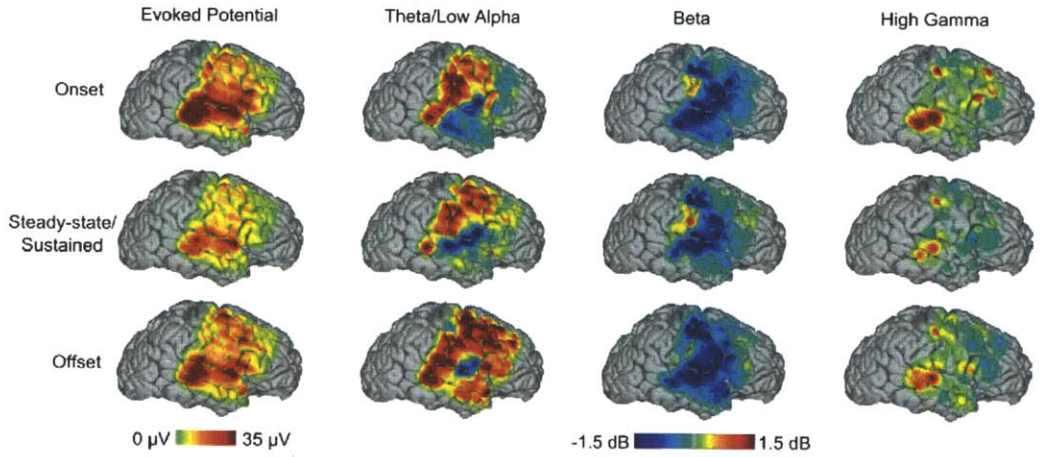

Figure 5.5. Cortical maps of the evoked potentials and individual frequency bands across three different time periods: stimulus onset (top row), steady-state or sustained responses (middle row), and stimulus offset (bottom row). The spatial patterns of sustained vs. phasic responses can be seen clearly in both the evoked potentials and high gamma activity where temporal lobe sites showed the greatest propensity for a robust response throughout the stimulus and frontal sites showed more phasic responses at stimulus onset and offset. In contrast, beta activity showed widespread decreases which persisted throughout the stimulus duration, while theta/low alpha activity showed separate patterns of increases and decreases, some of which were driven by the evoked potentials, especially at stimulus onset and offset.

various frequency bands (theta/low alpha: 7-9 Hz; beta: 12-30 Hz; high gamma: 70-190 Hz) across time within the sequence of acoustic stimulation. Both the evoked responses and high gamma activity showed spatial patterns of phasic vs. sustained responses. Much of the sustained and/or steady-state responses were from sites near the lateral fissure, while sites outside this region showed primarily onset or offset responses. The widespread beta decrease is confirmed by the maps in Figure 5.5. The theta/alpha activity tended to show responses which were either (i) driven by components in the evoked potential (e.g. G22) or sustained increases or decreases which were spatially segregated.

5.3.3 Waveshape index

The tendency of individual electrode sites to show phasic or sustained and/or steady-state responses was quantified by calculating a waveshape index (See Methods) for both the evoked potentials and high-gamma responses. The index was calculated by summing the RMS power in the *ON* and *OFF* responses divided by the sum of the average steady-state (in the case of EPs) or sustained (in the case of high-gamma) response and the max of the *ON* and *OFF* responses, and finally normalized by the signal-to-noise ratio for each channel. The left panels of Figure 5.6 show

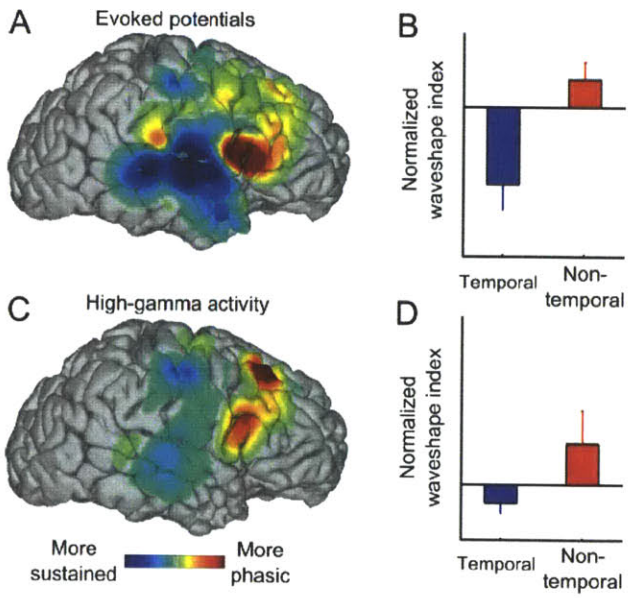

Figure 5.6. Cortical maps of the waveshape index defined for both the evoked potentials (**A**) and high-gamma activity (**C**), and there corresponding averages broken down across brain area (**B** and **D**). Phasic responses were most often observed in frontal areas, while sustained/steady-state responses were most often observed in temporal areas, though this comparison did not reach statistical significance for the high-gamma activity, perhaps due to a strong sustained high-gamma response near the central sulcus.

the waveshape indices for the EPs (panel A) and high-gamma activity (panel C) mapped onto the cortical surface. The right panels show the average waveshape indices for temporal (blue) and extratemporal (red) electrode sites for the EPs (panel B) and high-gamma activity (panel D). The difference between the waveshape index in temporal vs. extra-temporal brain areas was highly significant for the evoked response ($p<0.001$) but failed to reach statistical significance in the high-gamma activity (*n.s.*), as assessed by Mann-Whitney U tests.

5.4 Discussion

Utilizing direct cortical recordings from humans engaged in an auditory streaming paradigm, we have demonstrated that a multitude of temporal acoustic scales are present in the evoked responses (and high-gamma activity) to long sequences of tones. These timescales ranged from steady-state responses to individual events within the tone sequence (hundreds of milliseconds), persistent sustained responses as well as on and/or off responses (several seconds). Although not wholly segregated anatomically, responses in peri-Sylvian regions showed a propensity for sustained and/or steady-state responses which persisted throughout the duration of the tone sequence while responses in frontal and parietal cortex tended to respond only at the onset and offset of the tone sequence. This was especially true of of averaged evoked potentials and high-gamma activity, though the latter tended to show more fine-grained patterns, a result that has been consistently demonstrated with intracranial recordings in humans (Crone et al., 2001; Edwards et al., 2005, 2009). Interestingly, the finer-grained patterns of high-gamma activity and – to a lesser extent – the evoked potentials in the auditory cortex tended to break down along the boundary between the putative dorsal and ventral processing streams (Rauschecker and Scott, 2009), with those sites located more dorsally showing more phasic responses and those located more ventrally showing more sustained and/or steady-state responses. In contrast, alpha and beta activity (10-30Hz) showed widespread suppression during acoustic stimulation and did not tend to show easily-segregated anatomical patterns (c.f. Figure 5.5).

5.4.1 Mechanisms underlying varying temporal sensitivity

While the neural mechanisms underlying the varying temporal sensitivity observed across different hierarchical levels of the brain are not well understood, one process that may play a key

role is neural adaptation. Several previous studies, in both experimental animals (Fitzpatrick et al., 1999; Creutzfeldt et al., 1980; Burkard et al., 1999; Lu et al., 2001; Bartlett and Wang, 2007; Wang et al., 2008) and humans (Davis et al., 1966; Fruhstorfer et al., 1970; Fruhstorfer, 1971; Picton et al., 1974; Thornton and Coleman, 1975; Näätänen and Picton, 1987; Harms and Melcher, 2002; Seifritz et al., 2002; Harms et al., 2005; Brugge et al., 2009), are consistent with the idea that neural adaptation in response to successive acoustic stimuli increases along the ascending auditory pathway. The results from the present study extend empirical evidence for this idea further into what are thought to be high-level association areas of the auditory cortex (including lateral STG, STS, and MTG) as well as supra- and multimodal areas in the frontal and parietal lobes. However, several sites showed responses which were much larger at stimulus offset than onset (c.f. Figure 5.2), suggesting that mechanisms in addition to neural adaptation may be operative there. A recent study has also shown support for offset-specific mechanisms in the auditory cortex (Scholl et al., 2010). If similar mechanisms exist in auditory-responsive areas of frontal and parietal cortex, then it could be that in the sites which only showed offset responses, our electrodes were somehow sensitive to the synapses driving off-, but not on-, responses. Alternatively, it is known that the representation of acoustic information undergoes drastic transformations from being primarily temporal-based to primarily rate-based at the level of the auditory cortex (Lu et al., 2001; Bendor and Wang, 2008; Wang, 2007; Schreiner and Winer, 2007; Wang et al., 2008; Sharpee et al., 2011), and it could be that the representation of acoustic information at supra- and multimodal levels of the brain is further transformed (Cohen et al., 2004, 2009; Russ et al., 2008; Lee et al., 2009; Lemus et al., 2009) such that it is difficult to observe sustained responses with relatively gross methods (e.g. iEEG, such as was used in the present study) in these high-level brain areas.

5.4.2 Comparison with fMRI studies

Several previous reports have examined the change in BOLD response properties across anatomical subdivisions of the auditory cortex to varying rates of acoustic stimulation (Harms and Melcher, 2002; Harms et al., 2005; Seifritz et al., 2002) or perceptual organization of alternating-frequency tone sequences (Gutschalk et al., 2007; Wilson et al., 2007; Schadwinkel and Gutschalk, 2010). Taken together, the general conclusions from these studies are that (i) auditory centers situated higher in the ascending auditory pathway (i.e. medial geniculate body, Heschl's gyrus, superior temporal gyrus) demonstrate large changes in BOLD-signal waveshape with changes in either physical or perceived rate of acoustic stimulation which covaries with the expected amount of neural adaptation for a given stimulus configuration and that (ii) for a given rate of acoustic stimulation above a certain critical rate, higher auditory centers show responses which are robust at the onset and offset of sound but weak or non-existent sustained responses. Although purely hemodynamic explanations for this effect cannot be completely ruled out, the authors of those studies interpreted their results to reflect concomitant changes in neural activity responsible for generating the BOLD signal. Our stimuli were most similar to those used by Wilson et al. (2007) and Gutschalk et al. (2007), though it is not straight-forward to compare because the physical rate of stimulation used in our study had components between 2 and 8 Hz which, in the STG, for example, would span the categorical changes seen in the BOLD response waveshape (Harms et al., 2005). Given that (i) transient evoked responses are a major contributor to the fMRI BOLD response in the auditory cortex (Gutschalk et al. 2010) and (ii) the steady-state responses we measured did not show drastic qualitative changes based on our stimulus manipulations (Dykstra et al.), our data – at least that from sites over auditory cortex – might not be expected to result in large changes in fMRI waveshape. In contrast, we did measure

offset responses in the auditory cortex that increased dramatically with increasing frequency separation between the alternating tones in our stimulus in a manner consistent with the changes seen in the fMRI off response with increased physical or perceived presentation rate. If one assumes a consistent relationship between neural activity and the hemodynamic response across brain areas, the responses we measured in areas outside the auditory cortex which were extremely phasic, especially in frontal cortex, would suggest that the fMRI BOLD response in these areas would also be much more phasic, at least across the acoustic parameter space used to date.

5.4.3 Functional role in scene analysis

The present results are consistent with the idea that the anatomical hierarchy that exists in the auditory system recapitulates the temporal hierarchies that exist in natural acoustic signals (Kiebel et al., 2008, 2009a, 2009b; Harms and Melcher, 2002; Hasson et al., 2008; Lerner et al., 2011) and extends empirical support for it into areas outside the auditory cortex including frontal and parietal areas. The functional role of this anatomical recapitulation could be to support auditory scene analysis in decomposing the acoustic environment into different temporal receptive windows in order to register short-term vs. long-term changes in the acoustic environment. This registration process could be critical for auditory functions operative on successively longer timescales (e.g. pitch, rhythm, word, sentence, and melody comprehension).

It has previously been proposed that the neural activity underlying phasic BOLD responses to perceptually contiguous events in the auditory cortex (Harms and Melcher, 2002; Harms et al., 2005; Seifritz et al., 2002) represents the coding of the onsets and offsets of those events. The present results are broadly consistent with this idea, at least to the level of auditory association cortex. The responses in supra-modal areas and – to a lesser extent – even those in

very posterior STG observed in the present study could extend this idea to even longer timescales. Under such a scheme, frontal and parietal areas could support the encoding of qualitative changes in the acoustic scene (Knight, 1984; Halgren et al., 1995a, 1995b, 2010; Baudena et al., 1995; Winkler et al., 2009; Bekinschtein et al., 2009; Näätänen et al., 2010) for the purposes of attentional reorienting. Halgren and colleagues, using a change-detection paradigm which was completely controlled for effects of tone frequency, rarity, and habituation, recently reported direct evidence for this idea. However, this scheme does not account for responses which only occur at the onset or offset of stimulus sequences, a phenomenon clearly present in the current data set. A simple explanation for this might be that on and off responses are driven by different sets of synapses (Scholl et al., 2010) to which our electrodes are differentially sensitive.

Finally, given that we did not compare activation patterns for attended vs. ignored stimuli, it is difficult to make statements regarding what role the fact that our subjects were engaged in an active behavioral task may have played in the responses we measured. However, given recent results (Halgren et al. 2010), we would expect strong modulation of frontal activity based on whether or not the stimuli are attended along with lesser effects of attention in superior temporal areas.

5.5 References

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

Baudena, P., Halgren, E., Heit, G., and Clarke, J. M. (1995). Intracerebral potentials to rare target and distractor auditory and visual stimuli. III. Frontal cortex. *Electroencephalogr Clin Neurophysiol* 94, 251-264.

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., and Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proc. Natl. Acad. Sci. U.S.A* 106, 1672-1677.

Bendor, D., and Wang, X. (2008). Neural response properties of primary, rostral, and rostrotemporal core fields in the auditory cortex of marmoset monkeys. *J. Neurophysiol* 100, 888-906.

Brugge, J. F., Nourski, K. V., Oya, H., Reale, R. A., Kawasaki, H., Steinschneider, M., and Howard, M. A., 3rd (2009). Coding of repetitive transients by auditory cortex on Heschl's gyrus. *J. Neurophysiol* 102, 2358-2374.

Burkard, R. F., Secor, C. A., and Salvi, R. J. (1999). Near-field responses from the round window, inferior colliculus, and auditory cortex of the unanesthetized chinchilla: manipulations of noiseburst level and rate. *J. Acoust. Soc. Am* 106, 304-312.

Cohen, Y. E., Russ, B. E., Davis, S. J., Baker, A. E., Ackelson, A. L., and Nitecki, R. (2009). A functional role for the ventrolateral prefrontal cortex in non-spatial auditory cognition. *Proc. Natl. Acad. Sci. U.S.A* 106, 20045-20050.

Cohen, Y. E., Russ, B. E., Gifford, G. W., 3rd, Kiringoda, R., and MacLean, K. A. (2004). Selectivity for the spatial and nonspatial attributes of auditory stimuli in the ventrolateral prefrontal cortex. *J. Neurosci* 24, 11307-11316.

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

Crone, N. E., Boatman, D., Gordon, B., and Hao, L. (2001). Induced electrocorticographic gamma activity during auditory perception. Brazier Award-winning article, 2001. *Clin Neurophysiol* 112, 565-582.

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9, 179-194.

Davis, H., Mast, T., Yoshie, N., and Zerlin, S. (1966). The slow response of the human cortex to auditory stimuli: recovery process. *Electroencephalogr Clin Neurophysiol* 21, 105-113.

Dykstra, A. R., Halgren, E., Thesen, T., Carlson, C. E., Doyle, W., Madsen, J. R., Eskandar, E.

N., and Cash, S. S. Widespread Brain Areas Engaged during a Classical Auditory Streaming Task Revealed by Intracranial EEG. *Frontiers in Human Neuroscience*. Available at: http://www.frontiersin.org/human_neuroscience/abstract/11869 [Accessed July 18, 2011].

Edwards, E., Soltani, M., Deouell, L. Y., Berger, M. S., and Knight, R. T. (2005). High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. *J. Neurophysiol* 94, 4269-4280.

Edwards, E., Soltani, M., Kim, W., Dalal, S. S., Nagarajan, S. S., Berger, M. S., and Knight, R. T. (2009). Comparison of time-frequency responses and the event-related potential to auditory speech stimuli in human cortex. *J. Neurophysiol* 102, 377-386.

Ernst, M. D. (2004). Permutation Methods: A Basis for Exact Inference. *Statist. Sci.* 19, 676-685.

Fischl, B., Sereno, M. I., and Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9, 195-207.

Fitzpatrick, D. C., Kuwada, S., Kim, D. O., Parham, K., and Batra, R. (1999). Responses of neurons to click-pairs as simulated echoes: auditory nerve to auditory cortex. *J. Acoust. Soc. Am* 106, 3460-3472.

Fruhstorfer, H. (1971). Habituation and dishabituation of the human vertex response. *Electroencephalogr Clin Neurophysiol* 30, 306-312.

Fruhstorfer, H., Soveri, P., and Järvilehto, T. (1970). Short-term habituation of the auditory evoked response in man. *Electroencephalogr Clin Neurophysiol* 28, 153-161.

Gutschalk, A., Oxenham, A. J., Micheyl, C., Wilson, E. C., and Melcher, J. R. (2007). Human cortical activity during streaming without spectral cues suggests a general neural substrate for auditory stream segregation. *J. Neurosci* 27, 13074-13081.

Halgren, E., Baudena, P., Clarke, J. M., Heit, G., Liégeois, C., Chauvel, P., and Musolino, A. (1995a). Intracerebral potentials to rare target and distractor auditory and visual stimuli. I. Superior temporal plane and parietal lobe. *Electroencephalogr Clin Neurophysiol* 94, 191-220.

Halgren, E., Baudena, P., Clarke, J. M., Heit, G., Marinkovic, K., Devaux, B., Vignal, J. P., and Biraben, A. (1995b). Intracerebral potentials to rare target and distractor auditory and visual stimuli. II. Medial, lateral and posterior temporal lobe. *Electroencephalogr Clin Neurophysiol* 94, 229-250.

Halgren, E., Sherfey, J., Irimia, A., Dale, A. M., and Marinkovic, K. (2010). Sequential temporo-fronto-temporal activation during monitoring of the auditory environment for temporal patterns. *Hum Brain Mapp*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20665718 [Accessed February 22, 2011].

Harms, M. P., and Melcher, J. R. (2002). Sound repetition rate in the human auditory pathway: representations in the waveshape and amplitude of fMRI activation. *J. Neurophysiol* 88, 1433-1450.

Harms, M. P., Guinan, J. J., Sigalovsky, I. S., and Melcher, J. R. (2005). Short-term sound temporal envelope characteristics determine multisecond time patterns of activity in human auditory cortex as shown by fMRI. *J. Neurophysiol* 93, 210-222.

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., and Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *J. Neurosci* 28, 2539-2550.

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Comput. Biol* 4, e1000209.

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2009a). Perception and hierarchical dynamics. *Front Neuroinformatics* 3, 20.

Kiebel, S. J., von Kriegstein, K., Daunizeau, J., and Friston, K. J. (2009b). Recognizing sequences of sequences. *PLoS Comput. Biol* 5, e1000464.

Knight, R. T. (1984). Decreased response to novel stimuli after prefrontal lesions in man. *Electroencephalogr Clin Neurophysiol* 59, 9-20.

Langner, G. (1992). Periodicity coding in the auditory system. *Hear. Res* 60, 115-142.

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

Lee, J. H., Russ, B. E., Orr, L. E., and Cohen, Y. E. (2009). Prefrontal activity predicts monkeys' decisions during an auditory category task. *Front Integr Neurosci* 3, 16.

Lemus, L., Hernández, A., and Romo, R. (2009). Neural encoding of auditory discrimination in ventral premotor cortex. *Proc. Natl. Acad. Sci. U.S.A* 106, 14640-14645.

Lerner, Y., Honey, C. J., Silbert, L. J., and Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J. Neurosci* 31, 2906-2915.

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

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177-190.

Näätänen, R., and 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.

Näätänen, R., Kujala, T., and Winkler, I. (2010). Auditory processing that leads to conscious perception: A unique window to central auditory processing opened by the mismatch negativity and related responses. *Psychophysiology*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20880261 [Accessed December 8, 2010].

Picton, T. W., Hillyard, S. A., Krausz, H. I., and Galambos, R. (1974). Human auditory evoked potentials. I: Evaluation of components. *Electroencephalography and Clinical Neurophysiology* 36, 179-190.

Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci* 12, 718-724.

Russ, B. E., Orr, L. E., and Cohen, Y. E. (2008). Prefrontal neurons predict choices during an auditory same-different task. *Curr. Biol* 18, 1483-1488.

Schadwinkel, S., and Gutschalk, A. (2010). Functional dissociation of transient and sustained fMRI BOLD components in human auditory cortex revealed with a streaming paradigm based on interaural time differences. *Eur. J. Neurosci* 32, 1970-1978.

Scholl, B., Gao, X., and Wehr, M. (2010). Nonoverlapping sets of synapses drive on responses and off responses in auditory cortex. *Neuron* 65, 412-421.

Schreiner, C. E., and Langner, G. (1988). Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. *J. Neurophysiol* 60, 1823-1840.

Schreiner, C. E., and Winer, J. A. (2007). Auditory cortex mapmaking: principles, projections, and plasticity. *Neuron* 56, 356-365.

Seifritz, E., Esposito, F., Hennel, F., Mustovic, H., Neuhoff, J. G., Bilecen, D., Tedeschi, G., Scheffler, K., and Di Salle, F. (2002). Spatiotemporal pattern of neural processing in the human auditory cortex. *Science* 297, 1706-1708.

Sharpee, T. O., Atencio, C. A., and Schreiner, C. E. (2011). Hierarchical representations in the auditory cortex. *Curr Opin Neurobiol*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21704508 [Accessed July 11, 2011].

Thornton, A. R., and Coleman, M. J. (1975). The adaptation of cochlear and brainstem auditory evoked potentials in humans. *Electroencephalogr Clin Neurophysiol* 39, 399-406.

Wang, X., Lu, T., Bendor, D., and Bartlett, E. (2008). Neural coding of temporal information in auditory thalamus and cortex. *Neuroscience* 157, 484-494.

Wang, X. (2007). Neural coding strategies in auditory cortex. *Hear. Res* 229, 81-93.

Wilson, E. C., Melcher, J. R., Micheyl, C., Gutschalk, A., and Oxenham, A. J. (2007). Cortical FMRI activation to sequences of tones alternating in frequency: relationship to perceived rate and streaming. *J. Neurophysiol* 97, 2230-2238.

Winkler, I., Denham, S. L., and Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn. Sci. (Regul. Ed.)* 13, 532-540.

Chapter 6: Intracranial Neural Correlates of Auditory Perceptual Awareness

Andrew R. Dykstra[1,2], Eric Halgren[3], Alexander Gutschalk[4], Alexandra J. Golby[5], Emad Eskandar[6], Sydney S. Cash[2]

**Note: This chapter is in preparation for submission as a brief communication to the Journal of Neuroscience*

[1]Program in Speech and Hearing Bioscience and Technology, Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA
[2]Cortical Physiology Laboratory, Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA
[3]Departments of Radiology and Neurosciences, University of California San Diego, San Diego, CA,
[4]Department of Neurology, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany.
[5]Department of Neurosurgery, Brigham and Women's Hospital and Harvard Medical School, Boston, MA
[6]Department of Neurosurgery, Massachusetts General Hospital and Harvard Medical School

Abstract

In complex acoustic environments, our ability to perceive specific sounds of interest is thought to be limited by information-processing bottlenecks in the central auditory system. The neural activity underlying the conscious perception of target sounds in such environments has not been well characterized. We recorded the intracranial EEG from three neurosurgical patients with epilepsy who reported when they began to detect regularly-repeating target tones amongst random multi-tone maskers. Compared to undetected targets, detected targets elicited early (~50-200 ms) activity in the posterior auditory cortex as well as a broad long-latency (~300-600 ms) response that was widespread, including auditory cortex as well as ventrolateral prefrontal and anterolateral temporal cortices. The results demonstrate that the neural activity associated with the perceptual awareness of specific target sounds in complex settings engages diverse brain areas, including early sensory cortex as well as supramodal areas known to be involved in attentional selection and target detection.

6.1 Introduction

The human auditory system must constantly decompose the cacophony of acoustic input it receives into an accurate representation of underlying sound sources. This decomposition, termed auditory scene analysis (Bregman, 1994), is crucial for survival and communication, particularly in noisy settings. Failures of scene analysis, which increase with age (e.g. Alain and McDonald, 2007; Ross et al., 2007, 2010), are not entirely accounted for by limitations of the auditory periphery, and at least some failures are thought to arise from an information-processing bottleneck in the central auditory system (Overath et al., 2007; Gutschalk et al., 2008; Zylberberg et al., 2010). Failures of this nature have been termed informational masking (Pollack, 1975).

A common paradigm used to study informational masking from psychophysical (Neff and Green, 1987; Kidd et al., 1994, 2003, 2008) and neural (Gutschalk et al., 2008) perspectives involves presenting a target stream of tones amidst a random background. The target tones are surrounded by a protected frequency region to prevent "energetic masking" (Neff et al., 1993) occurring at the auditory periphery. Such randomly varying masker sequences, combined with a lack of *a priori* knowledge of the nature of the target, decrease the detectability of the target tones by as much as 40dB relative to fixed stimulus conditions (Kidd et al., 2008) and are thought to more closely approximate real-world environments.

Recent neurophysiological studies have highlighted the importance of the secondary auditory cortex in detecting (Gutschalk et al., 2008) and selectively attending to (Elhilali et al., 2009) a target stream amidst complex maskers. What role other brain areas might play in overcoming informational masking remains an open question. The present study combined direct electrical recordings from the cortical surface of neurosurgical patients with an auditory target detection task in order to further characterize the neural correlates of auditory perceptual

awareness under informational masking. We were particularly interested in extending the neural correlates of auditory perceptual awareness into (i) brain areas outside the auditory cortex and (ii) frequency regions less observable with non-invasive methods. (e.g. gamma-band activity).

Three patients with epilepsy undergoing invasive monitoring for identification of the seizure focus listened to stimulus sequences of random masker tones that sometimes contained regularly-repeating targets (Fig. 1). Participants were instructed to press a button the moment at which they began to hear out the target stream. Evoked responses and high-gamma activity time locked to individual target tones were compared between conditions in which the targets were perceived or unperceived. Targets and maskers alone served as control conditions and were compared to responses for detected and undetected targets. Detected targets elicited early (50-250 msec) activity that was either diminished or absent for undetected targets in posterior aspects of the auditory cortex as well as a late (300-600 msec) widespread response over large parts of frontal and temporal cortex. The results suggest a widespread cortical network involved in overcoming adverse listening situations.

6.2 Materials and Methods

6.2.1 Ethics Statement

All procedures were approved by the Institutional Review Boards at Partners Healthcare (Massachusetts General Hospital and Brigham and Women's Hospital) and the Massachusetts Institute of Technology (MIT) in accordance with NIH guidelines. Written informed consent was obtained from all patients prior to their participation.

6.2.2 Listeners

Five patients with intractable epilepsy underwent invasive monitoring in order to localize the epileptogenic zone prior to its surgical removal. However, two patients exhibited behavioral curves during training which indicated either a lack of understanding of or inability to perform the task and were thus excluded from further study. Each patient was implanted with an array of sub-dural platinum-iridium electrodes embedded in silastic sheets (2.3mm exposed diameter, 10mm center-to-center spacing; *Ad-tech Medical, Racine, WI*) placed directly on the cortical surface. Prior to implantation, each patient underwent high-resolution T1-weighted MRI. Subsequent to implantation, each patient underwent postoperative computerized tomography (CT). Electrode coordinates obtained from postoperative scans were co-registered with preoperative MRI and overlaid onto the patient's reconstructed cortical surface using Freesurfer (Dale et al., 1999; Fischl et al., 1999a) and custom MATLAB (*The Mathworks, Framingham, MA*) scripts (A. Dykstra et al., *under review*). Electrode coordinates were then projected onto the Freesurfer average brain using a spherical registration between the individual's cortical surface and that of the Freesurfer average (Fischl et al., 1999b).

6.2.3 Stimuli and procedure

Stimuli were 7.2-second sequences of tones comprised of a masker and on some trials, a target stream (Fig. 1) (Kidd et al., 1994, 2003; Micheyl et al., 2007; Gutschalk et al., 2008). Every tone in the sequence was 100 ms in duration including 10-ms raised cosine on and off ramps, and was chosen from frequency bands with the following set of center frequencies: 0.239, 0.286, 0.343, 0.409, 0.489, 0.585, 0.699, 0.836, 1, 1.196, 1.430, 1.710, 2.045, 2.445, 2.924, 3.497, 4.181, or 5.000 kHz. The masker was comprised of tones placed randomly in time and frequency within each band with an average within-band stimulus-onset asynchrony (SOA) of 800 ms (range: 100 – 1,500 ms). Within each band, the exact frequency of any given tone was within an estimated equivalent rectangular bandwidth (ERB), where ERB = 24.7*(4.37*fc + 1), where fc is the center frequency of a given band, in kHz. When present, the target stream was comprised of eight identical tones (chosen from the following subset of six frequencies: 0.489, 0.699, 1, 1,430, 2,045, or 2.924 kHz) with an SOA of 800 ms that always began 800 ms after the onset of the masker. Note that – given the random nature of the masker – this does not indicate that the onset of the first target, when present, was always exactly 800 ms after the onset of the first tone in the masker, but rather 800 ms after the possibility of the first occurrence of a tone in the masker sequence. In order to mitigate the occurrence of energetic masking of the target stream, two bands on each side of the target stream were omitted from the masker such that the masker was comprised of the remaining 13 frequency bands, for trials both with and without the target stream being present. The target stream was presented on 2/3 of the total number of trials presented to each participant.

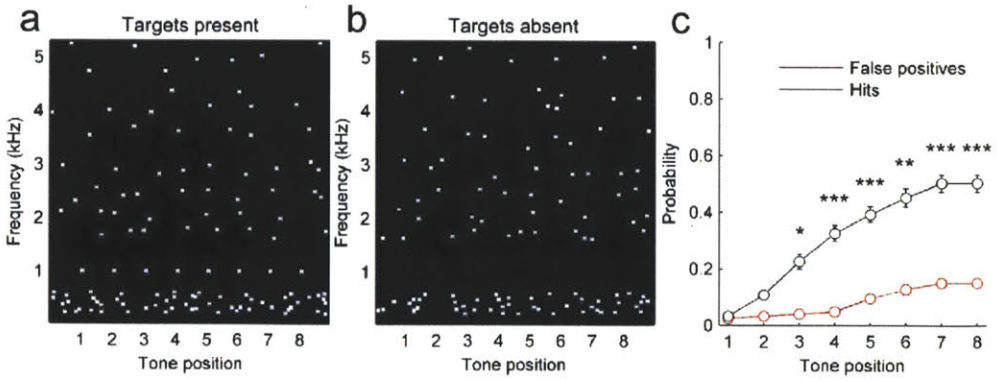

Figure 6.1. Schematic of an example stimulus sequence and behavioral results.

(**a**) Spectrogram of the jittered multi-tone masker stimulus used in the present study with regularly-repeating target tones at 1 kHz. (**b**) Same stimulus as in (**a**) with the targets absent. (**c**) Behavioral results averaged across the three patients. The false alarm rate across time since sequence onset is shown in red; hit rate is shown in black. Error bars denote ±1 s.e. of the mean (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.005$). It can be seen that although the target tones are easily segregated visually (c.f. panels **a** and **b**), they were not as easy to detect in our acoustic stimulus sequences.

All sound stimuli were generated in MATLAB (*The Mathworks Inc., Framingham, MA*), stored as .wav files, and converted to analog waveforms by the on-board sound card of a laptop computer equipped with Presentation software (*Neurobehavioral Systems, Albany, CA*). Stimuli were delivered to participants via Etymotic ER-2 insert earphones (*Etymotic Research, Inc., Elk Grove Village, IL)* diotically at a comfortable listening level. Participants were instructed to attend to the sound stimuli and to indicate via button box (*Cedrus Corporation, San Pedro, CA*) the moment at which they began to hear the repeating target stream, and to press the button again if they no longer heard the target stream. Participants were informed of the fact that the target stream would not be present on every trial but were not told the probability of its occurrence. The start of a new sequence began, on average, 1,600 ms after the preceding sequence's termination.

Each experiment was divided into three blocks. In each block, 36 target+masker (T+M;

Fig. 1a) stimuli and 18 masker-only (M; Fig. 1b) stimuli were presented, followed by 18 presentations of control stimuli which were comprised solely of target streams (T). Per block, this yielded 6 repetitions of a T+M condition for each target frequency, where the exact configuration of the masker stream varied for each T+M trial, 3 repetitions of each masker-only condition (where each condition was defined by the frequency of the target stream had it been present), and 3 repetitions of each T condition. The target-to-masker level ratio was initially set to 0 dB (i.e., targets were the same level as individual masker tones) and, if necessary, was lowered to -6 dB in subsequent blocks in an attempt to maintain approximately the same behavioral performance across blocks.

6.2.4 Data acquisition

Intracranial EEG (iEEG) data at MGH and BWH were acquired with standard clinical EEG monitoring equipment (*XLTEK, Natus Medical Inc., San Carlos, CA*) at a sampling rate of 500 Hz. All data were referenced to an inverted intracranial electrode (i.e. facing the inner skull table) remote from the electrodes of interest. For each patient, clinically-indicated, high-resolution T1-weighted structural MRI scans were acquired prior to surgery. High-resolution CT scans were acquired subsequent to surgery for the purpose of electrode localization.

6.2.5 Data pre-processing

Intracranial EEG data were bandpass filtered offline between 1 and 190 Hz and notch filtered at 60 Hz and its harmonics using zero-phase shift IIR filters. Independent component analysis using the *runica* algorithm (Bell and Sejnowski, 1995) in EEGLAB (Delorme and Makeig, 2004) was performed on the "raw" data. Components dominated by large artifacts were identified by inspection and projected out of the data.

The iEEG was epoched relative to the onset of individual tones within the target stream

(for T+M conditions) and, as a control condition, to the onset of individual target-tone positions had they been present (for M conditions). For comparison, we also epoched the iEEG relative to the onset of individual tones in the target-only (T) condition. Epoched waveforms were baseline corrected to the 100 ms preceding tone onset (in T+M and T conditions) and to the 100 ms preceding virtual target tone onset (in M conditions). Because the epochs in the masker-only condition are time-locked relative to a non-existent tone onset (virtual targets), the evoked response should average to zero (Gutschalk et al., 2008). Epochs containing large epileptoform artifacts were rejected were rejected by visual inspection. Epochs in the T+M condition were binned according to whether or not the target tones were detected by the listener. An individual target tone was defined as "detected" if it fell after the participant indicated by button press that they perceived the target stream. Because the task of each listener was to detect a *repeating* target tone, the two individual target tones that preceded a button press were also placed in the "detected" bin. All remaining tones were placed in the "undetected" bin.

6.2.6 Statistical analysis

A modified version of the cluster-based, non-parametric statistical procedure outlined by Maris and Oostenveld (Maris and Oostenveld, 2007) was used to test for effects of target detection on target locked EP amplitude. Unpaired t-tests were used as the sample-level (i.e. individual time points within a single channel) statistic. Contiguous, statistically-significant samples (defined as $p<0.05$) within a single electrode were used to define the cluster-level statistic, which was computed by summing the sample-level statistics within a cluster. Statistical significance at the cluster level was determined by computing a Monte Carlo estimate of the permutation distribution of cluster statistics using 1000 re-samples of the original data (Ernst, 2004). Within a single electrode, a cluster was taken to be significant if it fell outside the 95%

confidence interval of the permutation distribution for that electrode. The determination of significant clusters was performed independently for each electrode. This method controls the overall false alarm rate within an electrode across time points; no correction for multiple comparisons was performed across electrodes. In order to control for possible confounds of target-tone frequency in the "detected" vs. "undetected" comparison, some target tones at each target-tone frequency were thrown out, making the number of tones in the "detected" and "undetected" bins equal at each frequency.

6.2.7 High-gamma power

Waveforms of high-gamma power were constructed using Hilbert transform-based methods. Each epoch was first band-pass filtered between 70 and 190 Hz using zero-phase shift IIR filters. Power waveforms were computed by squaring the absolute value of the Hilbert transform of the band-pass filtered waveforms. This yielded high-gamma power waveforms with the same temporal resolution as the evoked potentials (2 ms). These waveforms were then baseline-corrected in the same manner as the evoked potentials, namely by subtracting the mean power in each trial computed across the 100 ms preceding target-tone onset. Finally, the gamma-power waveforms were low-pass filtered at 20 Hz. The same statistical procedures described above for evoked potentials were applied to the high-gamma waveforms.

6.3 Results

6.3.1 Behavior

Fig. 1 shows schematics of the stimuli used in our study (panels a and b) as well as the behavioral results averaged across the three participants (c). As can be seen from Fig. 1c, hit rates for detecting regularly-repeating target tones increased throughout the presentation of the stimulus sequence and plateaued near 50%. Though false-alarm rates also increased with time since stimulus onset, they remained low overall and never exceeded 15%. A two-way anova confirmed main effects of hit vs. false-alarm rates ($F=74.4$, $p<0.0001$) and time since stimulus onset ($F=9.1$, $p<0.0001$) as well as a hit-vs-false-alarm-rate by time-since-onset interaction ($F=2.9$, $p<0.01$). Subsequent paired t-tests indicated that hit and false alarm rates began to differ significantly three target tones into the sequence (Fig. 1c).

6.3.2 Evoked potentials

Evoked responses were binned and averaged for each of four stimulus/perceptual conditions: (i) target tones presented in isolation, (ii) masker tones presented in isolation, time-locked to the onset of virtual target tones, (iii) detected and (iv) undetected target tones in T+M sequences. The responses to target tones alone and masker tones alone served as templates with which to compare the responses to detected and undetected target tones presented during T+M conditions.

Fig. 2 shows averaged evoked responses from electrode sites over the pSTG in each of the three patients. Panels a, c, and e show the 3D cortical reconstruction from each patient; black dots denote electrode sites from which the responses shown in b, d, and f, were obtained, respectively. The left panels of b, d, and f show the responses to the target tones (yellow) and masker tones (blue) alone. As expected since the masker-only epochs were time-locked to

virtual tone onsets (see Methods), the averaged response for these epochs (shown in blue) is nearly flat. In contrast, the averaged evoked responses to targets presented in isolation (shown in yellow) were robust and

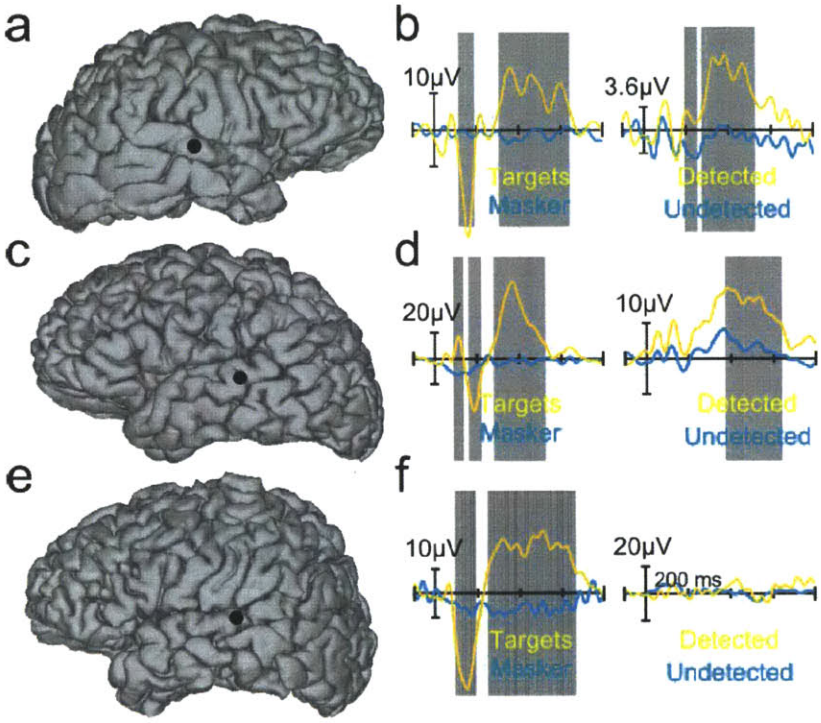

Figure 6.2. Evoked responses from electrode sites over the pSTG in each of the three patients.

Panels a, c, and e show each patient's 3D cortical reconstruction with a black dot indicating the site of the electrode for responses shown in b, d and f. The left panels of b, d, and f show the responses to the targets alone (yellow) and masker alone (blue), where the masker responses were created by time-locking to virtual target tones; the right panels show the responses to detected (yellow) and undetected (blue) targets. Gray shading indicates statistically-significant differences between yellow and blue waveforms in each of the six panels.

showed a stereotypical pattern across all three patients in sites over the pSTG characterized first by a large negative response in the N1, N1c, or P2 latency range (peaking at 152, 162, and 174 msec in patients 1, 2, and 3, respectively) followed by a broad long-latency response peaking between 300 and 600 msec in all three patients.

In 2/3 patients (Fig. 2a and 2c), this same long-latency response also differentiated between detected (yellow) and undetected (blue) targets (right-hand panels in Fig. 2b and 2d) between approximately 200 and 600 msec, with slight latency differences between patients. In the remaining patient (Fig. 2e), there was no significant difference in evoked-potential amplitude between the detected and undetected target conditions (Fig. 2f, right-hand panel). In all three patients, the response to the undetected targets (blue) was relatively flat, similar to the response to the masker-alone condition.

Fig. 3 shows topographical maps of average potential from patient 3 (the only patient in which we had broad frontal lobe coverage and there was a difference in detected-vs-undetected responses over posterior auditory cortex) for two latency ranges, 100-200 msec (two left-most columns) and 300-600 msec (two right-most columns) and each of the four behavioral conditions

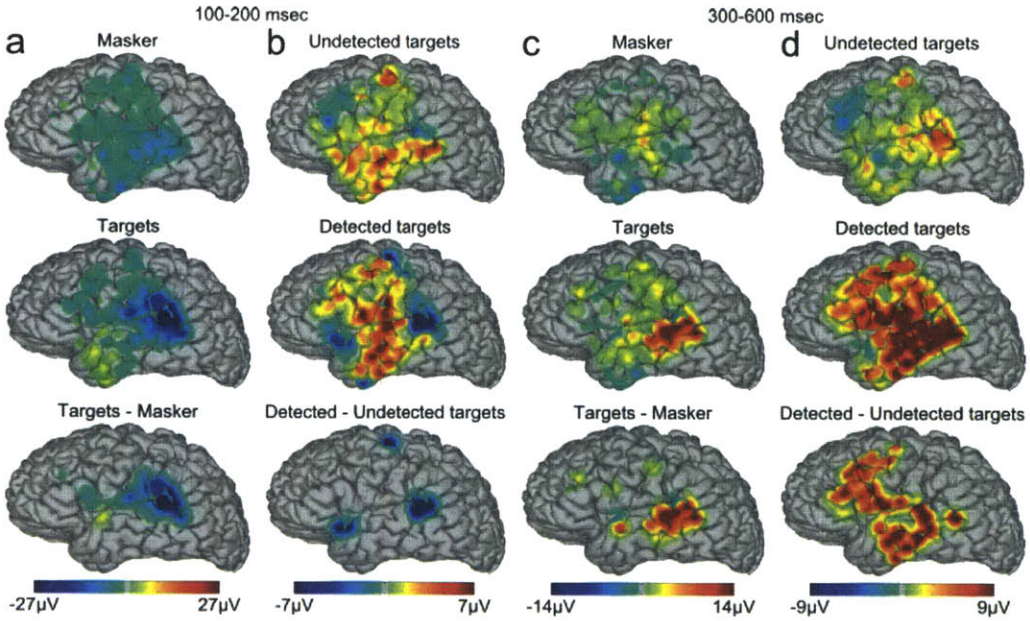

Figure 6.3. Topographies of evoked responses for patient 3.

The left six panels show the average of the evoked responses between 100 and 200 msec for targets alone and masker alone (a), detected and undetected targets (b), and their respective subtraction waveforms. The right six panels (c and d) show the average of the evoked responses between 300 and 600 msec

for the same conditions.

(top two rows) in addition to their subtraction (bottom row). All panels in the bottom row were thresholded to show only the activity from electrode sites which showed statistically-significant differences between conditions in the corresponding latency range. The primary differentiation of detected vs. undetected targets in the N1-latency range is over the posterior auditory cortex (lower panel of b). This was similar to the differentiation seen between targets-alone and masker-alone (lower panel of a). In contrast, the longer-latency response differentiated between detected and undetected targets in much more widespread brain areas than for targets-alone vs. masker-alone in the same latency range (compare lower panels of c and d), particularly ventrolateral prefrontal cortex.

6.3.3 High-gamma activity

High-gamma power waveforms were constructed using Hilbert-transform methods (squared analytic amplitude of band-pass filtered raw data – see Methods). Resultant waveforms were

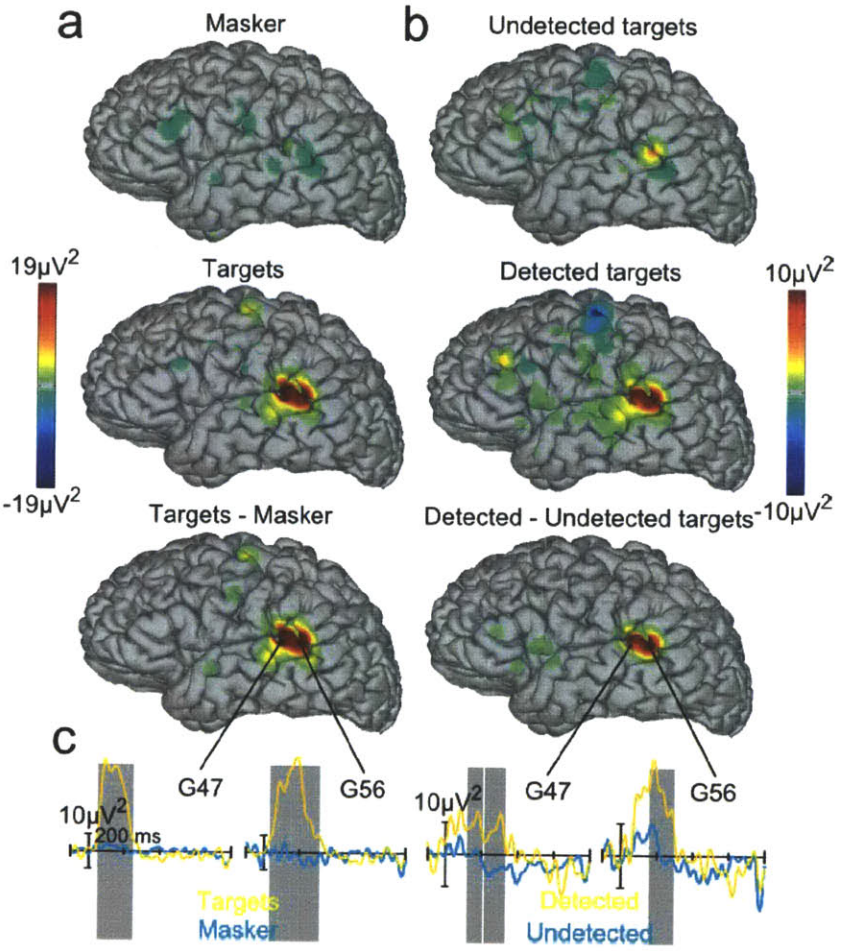

Figure 6.4. High-gamma activity for patient 3.

Topographies of high-gamma activity between 50 and 250 msec for targets alone and masker alone (a), detected and undetected targets (b), and their respective subtraction waveforms (lower panels of a and b). (c) High-gamma responses for each condition (left: yellow – targets alone, blue – masker alone; right: yellow – detected targets, blue – undetected targets) from the two sites showing the largest response.

binned and averaged according to the same procedures as for evoked responses. Figs. 4a and 4b show topographical plots of average high-gamma activity between 50 and 250 msec in patient 3 for targets and masker-alone, undetected and detected targets, and their respective subtractions after thresholding for statistical significance. As is shown by Fig. 4, high-gamma activity was much more focal than evoked responses, confined almost entirely to the posterior auditory cortex

near the sites which showed maximal evoked responses in the N1-latency range (c.f. Fig. 3a and Fig. 3b). Fig. 4c shows the full time courses of the two electrodes which showed by far the largest response in the high-gamma band for both targets vs. masker alone (two left-most panels) and detected vs. undetected targets (two right-most panels). As in the evoked responses, the responses to masker-alone and undetected-target conditions were relatively flat, more so in the masker-alone condition. In contrast, targets-alone and detected-target conditions elicited robust responses peaking between 50 and 250 msec. The onset of the high-gamma response was early than any of the evoked responses, beginning as early as 35 msec and 60 msec in the targets-alone and detected-target conditions. The site shown in Figs. 2a and 2b (patient 2) also showed robust high-gamma responses in the same 50-250 msec latency range, mainly to the targets-alone and detected-target conditions, but did not reach statistical significance for detected-vs-undetected targets (data not shown), perhaps due to a lower signal-to-noise ratio in the high-gamma band for that patient.

6.4 Discussion

Utilizing intracranial recordings in neurosurgical patients, the present study demonstrated robust effects of auditory perceptual awareness in (i) early (50-200 ms) evoked responses and high-gamma activity in posterolateral portions of auditory cortex and (ii) a widespread long-latency (300-600 ms) response that spread to ventrolateral prefrontal and anterolateral temporal cortex. To our knowledge, this is the first demonstration of differences in either high-gamma activity or activity outside the auditory cortex covarying with listeners' perceptual awareness of an acoustic stimulus. High-gamma activity to both isolated targets and detected targets was confined to the posterior regions of the STG [the posterolateral superior temporal (PLST) area described previously (Howard et al., 2000; Brugge et al., 2008)]. The long-latency activity observed in supramodal areas during the detection task was not seen to target tones presented in isolation, suggesting that the activity in these areas was behaviorally gated (Fritz et al. 2010) and that it reflected the imposition of a perceptual template of the to-be-detected target stream onto earlier processing stages located in auditory cortex.

6.4.1 Early responses in auditory cortex

Only one previous study used stimuli similar to those used here to examine the neural correlates of auditory perceptual awareness (Gutschalk et al., 2008). That study reported a broad long-latency (50-250 msec) response to detected targets – unmeasurable for undetected targets – localized to the posterior superior temporal plane (pSTP) near Heschl's gyrus, approximately 1.5 cm from the lateral surface. Given that (i) we did not observe polarity reversals across the lateral fissure (c.f. Fig. 3a) and (ii) the locus of response we measured was more posterior than the dipole sources reported in that study, it is unlikely that we measured activity from the same region. Rather, the early (100-200 msec) responses we measured were likely generated by the

region of the pSTG just underneath the electrode with radial dipole orientations, similar to what has been shown for the N1c by non-invasive EEG experiments (Picton et al., 1999; Snyder et al., 2006). Gutschalk et al. (2008) may not have measured activity from this region due to MEG's insensitivity to radial sources (Hämäläinen, 1992). Thus, the activity we measured in sites over the pSTG constitute novel responses during this task, and could possibly be confirmed by non-invasive studies utilizing a technique sensitive to radially-oriented sources such as EEG. Although this response significantly differentiated between detected and undetected targets in patient 3 (c.f. Fig. 3b), the difference was not as robust as that reported by Gutschalk et al. (2008) and was not present in the other two patients.

Another novel finding of the present study were the early high-gamma responses in some of the same sites that showed robust early evoked potentials. This early high-gamma activity strongly covaried with the listener's perceptual awareness of the target tones and, unlike the evoked responses from the same gross brain areas, were highly focal. Interestingly, the high-gamma activity measured here did not extend into other brain areas and longer latencies as the evoked responses did. The focal nature of intracranial high-gamma activity, especially when compared with evoked potentials measured intracranially (Edwards et al., 2009; Dykstra et al., 2011),.is a consistent finding across several labs and auditory paradigms, and may reflect active neural processes in the immediate vicinity of the electrode. Furthermore, the early gamma activity found in the present study may reflect neural processes necessary in order for an acoustic stimulus to reach conscious awareness, antecedent to the long-latency responses observed in the auditory cortex and supramodal areas.

6.4.2 Widespread long-latency responses

A consistent finding across all three patients was the presence of a broad, long-latency

response over the STG in response to targets presented in isolation (Fig. 2 and Fig. 3c). In two of the three patients, a similar but much more widespread response strongly different covaried with listeners' perceptual awareness of the targets when presented amongst the random multi-tone masker (Fig. 3d). This is the first evidence of such a broad, long-latency, and widespread response to detected targets during this task, and could reflect either (i) brain areas downstream from the pSTP sites thought to be initially involved in the detection of the target tones (Gutschalk et al., 2008). or (ii) that there is an ongoing interaction between bottom-up and top-down detection processes when a listener is searching for a target stimulus in noise. Gutschalk et al. (2008) may have failed to detect this long-latency response based on the methods used, specifically that single-dipole models were used to examine activity arising from the auditory cortex and the fact that MEG could be insensitive to the long-latency activity we measured from the pSTG.

The difference in the extent of the response between targets in isolation and targets amongst the random masker stimulus could be due to the active nature of the task during T+M conditions (Fritz et al., 2007, 2010). Specifically, when our listeners were actively engaged in detecting target tones which were difficult to perceive (Fig. 1c), the same target tones as presented in isolation in the targets-alone condition elicited robust activity in ventrolateral prefrontal cortex and anterolateral temporal areas. These responses are reminiscent of the P3b in that they were long-latency, broad, widespread, prominent in ventrolateral prefrontal and anterolateral temporal cortex, and only present when the subjects were engaged in an active task which required sustained overt attention to the stimuli (Halgren et al., 1998).

Lastly, the fact that patient 1 showed robust responses to the targets presented in isolation but virtually no response to the detected targets in the T+M condition is strange but may be

related to attentional factors. Of the three patients, patient 1 demonstrated the lowest overall sensitivity for detecting target tones and, qualitative observations during the experimental session indicated that the patient may have suffered momentary attentional lapses including falling asleep intermittently. If the patient payed sufficient initial attention to detect the target tones but subsequently ignored the remainder of each sequence, this might explain the null result obtained in this case.

6.5 References

Alain C, McDonald KL (2007) Age-related differences in neuromagnetic brain activity underlying concurrent sound perception. J Neurosci 27:1308-1314.

Bell AJ, Sejnowski TJ (1995) An information-maximization approach to blind separation and blind deconvolution. Neural Comput 7:1129-1159.

Bregman AS (1994) Auditory scene analysis: the perceptual organization of sound. MIT Press.

Brugge JF, Volkov IO, Oya H, Kawasaki H, Reale RA, Fenoy A, Steinschneider M, Howard MA (2008) Functional localization of auditory cortical fields of human: click-train stimulation. Hear Res 238:12-24.

Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9:179-194.

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9-21.

Dykstra AR, Halgren E, Thesen T, Carlson CE, Doyle W, Madsen JR, Eskandar EN, Cash SS (n.d.) Widespread Brain Areas Engaged during a Classical Auditory Streaming Task Revealed by Intracranial EEG. Frontiers in Human Neuroscience Available at: http://www.frontiersin.org/human_neuroscience/abstract/11869 [Accessed July 18, 2011].

Edwards E, Soltani M, Kim W, Dalal SS, Nagarajan SS, Berger MS, Knight RT (2009) Comparison of time-frequency responses and the event-related potential to auditory speech stimuli in human cortex. J Neurophysiol 102:377-386.

Elhilali M, Xiang J, Shamma SA, Simon JZ (2009) Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. PLoS Biol 7:e1000129.

Ernst MD (2004) Permutation Methods: A Basis for Exact Inference. Statist Sci 19:676-685.

Fischl B, Sereno MI, Dale AM (1999a) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9:195-207.

Fischl B, Sereno MI, Tootell RB, Dale AM (1999b) High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum Brain Mapp 8:272-284.

Fritz JB, David SV, Radtke-Schuller S, Yin P, Shamma SA (2010) Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. Nat Neurosci 13:1011-1019.

Fritz JB, Elhilali M, David SV, Shamma SA (2007) Auditory attention--focusing the searchlight on sound. Curr Opin Neurobiol 17:437-455.

Gutschalk A, Micheyl C, Oxenham AJ (2008) Neural correlates of auditory perceptual awareness under informational masking. PLoS Biol 6:e138.

Halgren E, Marinkovic K, Chauvel P (1998) Generators of the late cognitive potentials in auditory and visual oddball tasks. Electroencephalogr Clin Neurophysiol 106:156-164.

Hämäläinen MS (1992) Magnetoencephalography: a tool for functional brain imaging. Brain Topogr 5:95-102.

Howard MA, Volkov IO, Mirsky R, Garell PC, Noh MD, Granner M, Damasio H, Steinschneider M, Reale RA, Hind JE, Brugge JF (2000) Auditory cortex on the human posterior superior temporal gyrus. J Comp Neurol 416:79-92.

Kidd G, Mason CR, Deliwala PS, Woods WS, Colburn HS (1994) Reducing informational masking by sound segregation. J Acoust Soc Am 95:3475-3480.

Kidd G, Mason CR, Richards VM (2003) Multiple bursts, multiple looks, and stream coherence in the release from informational masking. J Acoust Soc Am 114:2835-2845.

Kidd G, Mason CR, Richards VM, Gallun FJ, Durlach NI (2008) Informational Masking In Auditory Perception of Sound Sources New York, NY: Springer, p. 143-190.

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177-190.

Micheyl C, Shamma S, Oxenham AJ (2007) Hearing out repeating elements in randomly varying multitone sequences: a case of streaming In Hearing - from basic research to application Berlin: Springer, p. 267-274.

Neff DL, Dethlefs TM, Jesteadt W (1993) Informational masking for multicomponent maskers with spectral gaps. J Acoust Soc Am 94:3112-3126.

Neff DL, Green DM (1987) Masking produced by spectral uncertainty with multicomponent maskers. Percept Psychophys 41:409-415.

Overath T, Cusack R, Kumar S, von Kriegstein K, Warren JD, Grube M, Carlyon RP, Griffiths TD (2007) An information theoretic characterisation of auditory encoding. PLoS Biol 5:e288.

Picton TW, Alain C, Woods DL, John MS, Scherg M, Valdes-Sosa P, Bosch-Bayard J, Trujillo NJ (1999) Intracerebral sources of human auditory-evoked potentials. Audiol Neurootol 4:64-79.

Pollack I (1975) Auditory informational masking. J Acoust Soc Am 57:S5.

Ross B, Fujioka T, Tremblay KL, Picton TW (2007) Aging in binaural hearing begins in mid-life: evidence from cortical auditory-evoked responses to changes in interaural phase. J

Neurosci 27:11172-11178.

Ross B, Schneider B, Snyder JS, Alain C (2010) Biological markers of auditory gap detection in young, middle-aged, and older adults. PLoS ONE 5:e10101.

Snyder JS, Alain C, Picton TW (2006) Effects of attention on neuroelectric correlates of auditory stream segregation. J Cogn Neurosci 18:1-13.

Zylberberg A, Fernández Slezak D, Roelfsema PR, Dehaene S, Sigman M (2010) The brain's router: a cortical network model of serial processing in the primate brain. PLoS Comput Biol 6:e1000765.

Chapter 7: General Discussion

This thesis described the results from two experiments in which electrical recordings were made directly from the cortical surface neurosurgical patients with epilepsy. Each experiment was designed to measure neural correlates of subjective auditory perceptual organization without confounding changes in physical stimuli. By using intracranial recordings in awake behaving humans, the experiments in this thesis circumvented limitations inherent in previous work. Specifically, active behavioral paradigms were combined with neural recordings yielding high spatiotemporal resolution and broad coverage.

The first experiment used a classic behavioral streaming paradigm and compared neuronal activity between conditions in which physical stimuli were held constant but perceptual organization changed dramatically (Chapter 4). Three main conclusions were made from the results of this experiment: (i) the brain areas that are engaged during the behavioral streaming paradigm are much more widespread than previously shown, (ii) the activity within a given macroscopic brain area is not uniform and (iii) the neural correlates of auditory perceptual organization during the streaming task are likely to be diffuse in nature, found on a finer spatial scale or in a brain area not sampled by the recordings made for this thesis. A separate analysis of the same data set revealed that the timescale on which a brain area responded varied widely between sites showing a sustained and/or steady-state response persisting throughout the duration of each stimulus sequence and those showing responses only at the onset or offset of each sequence (Chapter 5).

The second experiment used a jittered version of the multi-tone masking stimuli and compared neuronal activity between conditions in which regularly-repeating target tones were

detected vs. Undetected (Chapter 6). In addition to relatively early responses in sites over the pSTG showing differences between detected and undetected targets, a broad, widespread, long-latency response was also found to differ greatly between detected and undetected targets, including involvement of lateral temporal and frontal areas not seen to easily-detected targets in the targets-alone stimuli. In contrast, differences between detected and undetected targets in high-gamma activity were highly focal (confined to the pSTG) and comparatively early.

A third part of this thesis described an *ad hoc* procedure for localizing intracranial electrodes with respect to individual neuroanatomy and coregistering electrode ensembles across multiple patients into a common space (Chapter 3). The method accounts for the brain shift known to be caused by implantation of intracranial electrodes by constructing a smoothed version of the three-dimensional pial surface and making minimal assumptions about the nature of the parenchymal shift. Electrode ensembles from individual patients were coregistered using spherical registration methods known to produce better alignment of structural and functional brain areas.

7.1 Widespread brain areas engaged during classical ASA tasks

Several previous authors have hypothesized involvement of supra-and multimodal areas in ASA (Bidet-Caulet and Bertrand, 2009). The present series of experiments are the first to demonstrate engagement of brain areas outside the auditory cortex during ASA paradigms with high temporal resolution. In particular, both paradigms, elicited activity in frontal cortex and multi- or supra-modal lateral temporal areas not seen in any previous studies using the streaming or multi-tone masking paradigm. Superficially, this demonstrates that the cortical networks that are engaged during these paradigms are much more widespread than previously shown and, speculatively, that these areas may be involved in accomplishing these tasks as well as in

overcoming naturally-occurring adverse listening conditions.

7.2 Variability in the responses

The neural activity we measured in response to relatively simple stimuli often varied tremendously from one patient to the next and even from one electrode to the next *within* a patient. Such intra- and inter-subject variability is a common finding in peri-cortical EP studies (Halgren et al., 1995a, 1995b; Baudena et al., 1995; Steinschneider et al., 1999, 2011; Crone et al., 2001; Edwards et al., 2005, 2009; Korzyukov et al., 2007; Rosburg et al., 2009) and, although the fact that these recordings are made in abnormal brains cannot be ruled out as a possible explanation, such variability has also been found in recordings from tumor patients for whom broad neural reorganization is not thought to occur (Edwards et al., 2009). This highlights the fact that non-invasive methods such as EEG and MEG provide a highly smeared version of true cortical source configurations that may contain much higher spatial frequencies, with the caveat that there have been few studies that have compared extracranially-measured evoked fields/potentials with intracranial potentials in the same subject or set of subjects [but see (Dalal et al., 2009; Krusienski and Shih, 2010)]. More studies of this nature would help elucidate the relationship, in individuals, between commonly-obtained extracranial data and rarely-obtained intracranial data. Until then, researchers using extracranial techniques should proceed with caution when interpreting the spatial configuration and extent of cortical sources that produce the fields/potentials observed extracranially.

7.3 The role of the auditory cortex

The experiments described in this thesis were only capable of recording activity from what are thought to be secondary or tertiary auditory areas, i.e. lateral STG and lateral HG, and not from what is thought to be core auditory cortex, i.e. medial HG (Kaas et al., 1999; Kaas and

Hackett, 2000; Sigalovsky et al., 2006; Woods et al., 2009, 2010; Hackett, 2011). Previous M/EEG studies of either streaming (Gutschalk et al., 2005, 2007; Snyder et al., 2006, 2009; Snyder and Alain, 2007; Schadwinkel and Gutschalk, 2010) or multi-tone masking (Gutschalk et al., 2008) have consistently localized dipoles to mainly the superior temporal plane between 1 and 2 cm from the lateral surface of the temporal lobe. The present results showed ΔF-correlated responses in electrodes placed over the lateral STG which likely arose locally and not from patches of tissue to which previous dipoles were located (c.f. Chapters 4 and 6). Thus, with regard to streaming, our results are probably more related to previous fMRI work demonstrating BOLD activity in more lateral portions of the superior temporal plane (Deike et al., 2004, 2010; Wilson et al., 2007; Gutschalk et al., 2007; Schadwinkel and Gutschalk, 2010). Already, those results hinted at possible sub-areal organization of streaming-related responses in that the statistical parametric maps were patchy rather than a single circumscribed region, although this could also be due to variability in the foci of activity across individuals. The present results also showed sub-areal differentiation of responses in that waveform morphology was not necessarily similar across adjacent sites over the STG. Taken together, the present data as well as the previous fMRI studies [and one EEG study - (Snyder et al., 2006)] suggest multiple sub-functions of the cortical tissue within a given macroscopic brain area, in this case the posterior lateral superior temporal area described previously (Howard et al., 2000; Brugge et al., 2008). We also measured activity arising from this region in the informational-masking experiment, something that has not been shown previously. Both early and late evoked responses as well as early gamma activity from this region strongly covaried with perceptual awareness, adding to the early responses shown to covary with percept that presumably arose from more medial aspects of the STP (Gutschalk et al., 2008).

7.4 The role of extra auditory cortical areas

In addition to the activity we observed over the PLST, we also observed much more widespread responses, particularly in frontal cortex. Such involvement of frontal areas in the two tasks we used has not been previously shown, and is likely only when the subjects were engaged in an *active* task, which they always were during the streaming experiments but only sometimes were during the informational masking experiments. In the latter, frontal and other supra-modal areas (i.e. anterolateral temporal cortex) were only active during the conditions in which the subject was actively searching for the target stream. Given that we did not examine any passive-listening conditions during the streaming task, we cannot say for sure whether the involvement of the frontal areas seen there is due to active engagement with the stimuli, but recent studies suggest that this is a reasonable hypothesis (Fritz et al., 2007, 2010).

7.5 Correlates of bistability in one task but not the other?

While target tones in the informational masking paradigm elicited very different neural responses depending on whether or not they were perceived, ABA- triplets did not elicit different neural responses depending on whether the subject was hearing one or two streams. The reasons for this are unclear, but could possibly be due to the nature of each task. In the streaming paradigm, listeners are asked to differentiate between two possible perceptual organizations while in the informational-masking paradigm, listeners are asked to simply indicate the presence of a target stream. These two tasks – generating a percept or not vs. distinguishing between two generated percepts – could be supported by different neuroanatomical networks and mechanisms. Qualitatively speaking, the experience during the informational-masking paradigm is one where each target tone – when detected – is processed individually, although they do bind together

across time due to the constant SOA between successive target tones. In contrast, the difference between 1- and 2-stream percepts in the streaming paradigm is experienced more on longer times scales rather than on the individual triplet level. Speculatively, the fact that most of our analysis of the responses generated during the streaming paradigm focused on triplet-level timescales may be one reason why we did not observe robust effects of perceptual organization and those that have been reported were small and variable (Cusack, 2005; Gutschalk et al., 2005). The text below speculates further about how one might approach identifying correlates of perceptual organization in the streaming paradigm.

As mentioned above, the difference between the experience of 1- and 2-stream percepts during the streaming paradigm subjectively occurs on longer time scales. The mean duration between switches from 1- to 2-stream percepts, and *vice versa*, is approximately 10 seconds (Pressnitzer and Hupé, 2006). These perceptual states must be accompanied by brain attractor states associated with 1- vs. 2-stream percepts acting on a similar time scale. The nature of these states (i.e. their dimensionality and state variables) is unknown, but it is conceivable that they wouldn't reveal themselves in time-locked analysis and possibly even that they are hidden from the common measurement techniques used to record human brain activity (MEG, EEG, fMRI). Transitions from 1-stream to 2-stream perceptual states (or vice versa) could be engendered by at least two factors in the underlying neural states: (i) noise in the stimulus encoding process, which begins at least as early as the cochlear nucleus or (ii) stochastic noise in the state variables. If the noise engendering bistable perception occurs primarily at the level of the state variables, it might produce a dissociation between how stimuli are processed on a triplet-level time-scale and the state transitions which lead to bistable perception, rendering neural correlates of bistable perception in the streaming paradigm difficult to measure using such time-locked analyses.

However, *transitions* between states produces a perceptual change of scene which may evoke large transient neural activity to which common recording techniques become sensitive (Kondo and Kashino, 2009; Schadwinkel and Gutschalk, 2010). Giving a "go" cue during ongoing streaming stimuli could engender an endogenous sensory-perceptual-motor transformation that would differ based on how the scene is perceptually organized at that instant in time. It may be possible to dissociate motor activity from that due only to perceptual organization by applying different time-locking procedures, specifically one to the motor response which could be subtracted from each individual trial time-locked to the go cue. However, it is my view that in order to directly access the underlying brain states accounting for bistable perception in the streaming task, detailed microphysiological recordings, perhaps in many areas of the brain, may be necessary. Such experiments could be done with either humans implanted with microelectrodes while undergoing neurosurgical planning (Howard et al., 1996; Mukamel et al., 2005, 2010; Bitterman et al., 2008; Brugge et al., 2009; Nourski et al., 2009; Griffiths et al., 2010) or with experimental animals trained to report their subjective auditory experience, similar to what has become commonplace in visual studies (Logothetis, 1998; Leopold and Logothetis, 1999).

7.6 Conclusions

The present results substantially extend and constrain theories about where and how ASA takes place in the brain. Specifically, we identified widespread neural networks engaged during both streaming and release from informational masking, and demonstrated that the neural correlates of perceptual bistability may be unique to each task used to study them. The neural correlates of perceptual bistability during the streaming paradigm are likely to be found on a

finer spatial scale than was assessed, and may require the use of trained experimental animals or neurosurgical patients implanted with microelectrodes in the auditory cortex and probably elsewhere, perhaps even simultaneously. Given that ours was the first study to demonstrate widespread frontal involvement in streaming and informational-masking tasks, he precise role of the frontal cortex in each of these tasks, as well as in ASA more generally, remains to be understood; whole-brain fMRI and M/EEG studies as well as microelectrode studies in experimental animals and humans may prove useful in this effort.

7.7 References

Baudena P, Halgren E, Heit G, Clarke JM (1995) Intracerebral potentials to rare target and distractor auditory and visual stimuli. III. Frontal cortex. Electroencephalogr Clin Neurophysiol 94:251-264.

Bidet-Caulet A, Bertrand O (2009) Neurophysiological mechanisms involved in auditory perceptual organization. Front Neurosci 3:182-191.

Bitterman Y, Mukamel R, Malach R, Fried I, Nelken I (2008) Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. Nature 451:197-201.

Brugge JF, Nourski KV, Oya H, Reale RA, Kawasaki H, Steinschneider M, Howard MA 3rd (2009) Coding of repetitive transients by auditory cortex on Heschl's gyrus. J Neurophysiol 102:2358-2374.

Brugge JF, Volkov IO, Oya H, Kawasaki H, Reale RA, Fenoy A, Steinschneider M, Howard MA (2008) Functional localization of auditory cortical fields of human: click-train stimulation. Hear Res 238:12-24.

Crone NE, Boatman D, Gordon B, Hao L (2001) Induced electrocorticographic gamma activity during auditory perception. Brazier Award-winning article, 2001. Clin Neurophysiol 112:565-582.

Cusack R (2005) The intraparietal sulcus and perceptual organization. J Cogn Neurosci 17:641-651.

Dalal SS, Baillet S, Adam C, Ducorps A, Schwartz D, Jerbi K, Bertrand O, Garnero L, Martinerie J, Lachaux J-P (2009) Simultaneous MEG and intracranial EEG recordings during attentive reading. Neuroimage 45:1289-1304.

Deike S, Gaschler-Markefski B, Brechmann A, Scheich H (2004) Auditory stream segregation relying on timbre involves left auditory cortex. Neuroreport 15:1511-1514.

Deike S, Scheich H, Brechmann A (2010) Active stream segregation specifically involves the left human auditory cortex. Hear Res 265:30-37.

Edwards E, Soltani M, Deouell LY, Berger MS, Knight RT (2005) High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. J Neurophysiol 94:4269-4280.

Edwards E, Soltani M, Kim W, Dalal SS, Nagarajan SS, Berger MS, Knight RT (2009) Comparison of time-frequency responses and the event-related potential to auditory speech stimuli in human cortex. J Neurophysiol 102:377-386.

Fritz JB, David SV, Radtke-Schuller S, Yin P, Shamma SA (2010) Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. Nat Neurosci 13:1011-1019.

Fritz JB, Elhilali M, David SV, Shamma SA (2007) Auditory attention--focusing the searchlight on sound. Curr Opin Neurobiol 17:437-455.

Griffiths TD, Kumar S, Sedley W, Nourski KV, Kawasaki H, Oya H, Patterson RD, Brugge JF, Howard MA (2010) Direct recordings of pitch responses from human auditory cortex. Curr Biol 20:1128-1132.

Gutschalk A, Micheyl C, Melcher JR, Rupp A, Scherg M, Oxenham AJ (2005) Neuromagnetic correlates of streaming in human auditory cortex. J Neurosci 25:5382-5388.

Gutschalk A, Micheyl C, Oxenham AJ (2008) Neural correlates of auditory perceptual awareness under informational masking. PLoS Biol 6:e138.

Gutschalk A, Oxenham AJ, Micheyl C, Wilson EC, Melcher JR (2007) Human cortical activity during streaming without spectral cues suggests a general neural substrate for auditory stream segregation. J Neurosci 27:13074-13081.

Hackett TA (2011) Information flow in the auditory cortical network. Hear Res 271:133-146.

Halgren E, Baudena P, Clarke JM, Heit G, Liégeois C, Chauvel P, Musolino A (1995a) Intracerebral potentials to rare target and distractor auditory and visual stimuli. I. Superior temporal plane and parietal lobe. Electroencephalogr Clin Neurophysiol 94:191-220.

Halgren E, Baudena P, Clarke JM, Heit G, Marinkovic K, Devaux B, Vignal JP, Biraben A (1995b) Intracerebral potentials to rare target and distractor auditory and visual stimuli. II. Medial, lateral and posterior temporal lobe. Electroencephalogr Clin Neurophysiol 94:229-250.

Howard MA 3rd, Volkov IO, Abbas PJ, Damasio H, Ollendieck MC, Granner MA (1996) A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. Brain Res 724:260-264.

Howard MA, Volkov IO, Mirsky R, Garell PC, Noh MD, Granner M, Damasio H, Steinschneider M, Reale RA, Hind JE, Brugge JF (2000) Auditory cortex on the human posterior superior temporal gyrus. J Comp Neurol 416:79-92.

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

Kaas JH, Hackett TA, Tramo MJ (1999) Auditory processing in primate cerebral cortex. Curr Opin Neurobiol 9:164-170.

Kondo HM, Kashino M (2009) Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. J Neurosci 29:12695-12701.

Korzyukov O, Pflieger ME, Wagner M, Bowyer SM, Rosburg T, Sundaresan K, Elger CE, Boutros NN (2007) Generators of the intracranial P50 response in auditory sensory gating. Neuroimage 35:814-826.

Krusienski DJ, Shih JJ (2010) A case study on the relation between electroencephalographic and electrocorticographic event-related potentials. Conf Proc IEEE Eng Med Biol Soc 2010:6019-6022.

Leopold, Logothetis (1999) Multistable phenomena: changing views in perception. Trends Cogn Sci (Regul Ed) 3:254-264.

Logothetis NK (1998) Single units and conscious vision. Philos Trans R Soc Lond , B, Biol Sci 353:1801-1818.

Mukamel R, Gelbard H, Arieli A, Hasson U, Fried I, Malach R (2005) Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. Science 309:951-954.

Mukamel R, Nir Y, Harel M, Arieli A, Malach R, Fried I (2010) Invariance of firing rate and field potential dynamics to stimulus modulation rate in human auditory cortex. Hum Brain Mapp Available at: http://www.ncbi.nlm.nih.gov.ezp-prod1.hul.harvard.edu/pubmed/20665720 [Accessed October 28, 2010].

Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Howard MA 3rd, Brugge JF (2009) Temporal envelope of time-compressed speech represented in the human auditory cortex. J Neurosci 29:15564-15574.

Pressnitzer D, Hupé J-M (2006) Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. Curr Biol 16:1351-1357.

Rosburg T, Trautner P, Fell J, Moxon KA, Elger CE, Boutros NN (2009) Sensory gating in intracranial recordings--the role of phase locking. Neuroimage 44:1041-1049.

Schadwinkel S, Gutschalk A (2010) Activity Associated with Stream Segregation in Human Auditory Cortex is Similar for Spatial and Pitch Cues. Cereb Cortex Available at: http://www.ncbi.nlm.nih.gov/pubmed/20237241 [Accessed September 14, 2010].

Sigalovsky IS, Fischl B, Melcher JR (2006) Mapping an intrinsic MR property of gray matter in auditory cortex of living humans: a possible marker for primary cortex and hemispheric differences. Neuroimage 32:1524-1537.

Snyder JS, Alain C (2007) Sequential auditory scene analysis is preserved in normal aging adults. Cereb Cortex 17:501-512.

Snyder JS, Alain C, Picton TW (2006) Effects of attention on neuroelectric correlates of auditory stream segregation. J Cogn Neurosci 18:1-13.

Snyder JS, Holder WT, Weintraub DM, Carter OL, Alain C (2009) Effects of prior stimulus and prior perception on neural correlates of auditory stream segregation. Psychophysiology 46:1208-1215.

Steinschneider M, Volkov IO, Noh MD, Garell PC, Howard MA (1999) Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. J Neurophysiol 82:2346-2357.

Steinschneider M, Nourski KV, Kawasaki H, Oya H, Brugge JF, Howard MA 3rd (2011) Intracranial Study of Speech-Elicited Activity on the Human Posterolateral Superior Temporal Gyrus. Cereb Cortex Available at: http://www.ncbi.nlm.nih.gov/pubmed/21368087 [Accessed July 14, 2011].

Wilson EC, Melcher JR, Micheyl C, Gutschalk A, Oxenham AJ (2007) Cortical FMRI activation to sequences of tones alternating in frequency: relationship to perceived rate and streaming. J Neurophysiol 97:2230-2238.

Woods DL, Herron TJ, Cate AD, Yund EW, Stecker GC, Rinne T, Kang X (2010) Functional properties of human auditory cortical fields. Front Syst Neurosci 4:155.

Woods DL, Stecker GC, Rinne T, Herron TJ, Cate AD, Yund EW, Liao I, Kang X (2009) Functional maps of human auditory cortex: effects of acoustic features and attention. PLoS ONE 4:e5183.