Auditory Pathway Responses to Parametrized Vowels in Autism Spectrum Disorders

by

Bennett Bullock
B.A. Classics
Brown University, 1998
M.Sc. Arabic Language, Literature, and Linguistics
Georgetown University, 2005

SUBMITTED TO THE HARVARD/MIT DIVISION OF HEALTH SCIENCES AND TECHNOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN HEALTH SCIENCES AND TECHNOLOGY AT THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY SEPTEMBER 2010

©2010 Bennett Bullock. All rights reserved.
The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Signature of Author: ____________________________
Harvard/MIT Division of Health Sciences and Technology
September 3, 2010

Certified by: ____________________________
Nicholas Lange, Sc.D.
Associate Professor of Psychiatry and Biostatistics
Harvard University Schools of Medicine and Public Health
Director, Neurostatistics Laboratory, McLean Hospital
Thesis Supervisor

Accepted by: ____________________________
Ram Sasisekharan, Ph.D.
Director, Harvard-MIT Division of Health Sciences and Technology
Edward Hood Taplin Professor of Health Sciences & Technology and Biological Engineering
Harvard/MIT Program in Health Biosciences and Technology

S.M. THESIS

Auditory Pathway Responses to Parametrized Vowels in Autism Spectrum Disorders

Author:
Bennett BULLOCK
bbullock@mit.edu

Advisor:
Nicholas LANGE, Sc.D.
Associate Professor of Psychiatry and Biostatistics
Harvard University Schools of Medicine and Public Health
Director, Neurostatistics Laboratory, McLean Hospital

September 3, 2010
ABSTRACT

Autism spectrum disorder (ASD) is characterized by many behavioral symptoms, including delays in social and communicative development. A cluster of symptoms concentrate on speech and language development, especially manipulation of non-verbal information conveyed in prosody. It is largely unknown whether this is due to functional or structural differences in the brain regions involved in auditory and speech processing, although recent studies have shown that ASD individuals do exhibit different activation patterns in various brain regions in response to speech stimuli. This study investigated responses in regions of the auditory pathway to short recorded and synthesized vowel stimuli. These regions were the Inferior Colliculus, the Left Thalamus, the left Posterior Insula, the Auditory Cortex, Wernicke’s area, and Broca’s area. The stimuli were parametrized so as to target different signal processing capabilities associated with each region. They were presented to ASD and typically developing (TD) subjects while the salient regions were subject to a functional magnetic resonance imaging (fMRI). The results suggest that there were not gross differences in how ASD individuals responded from TD individuals in the subcortical regions. Results from the Auditory Cortex, however, showed a significant hemisphere dominance in TD subjects with more temporally complex stimuli that did not appear in ASD subjects. Moreover, the results showed that it was temporally-measured periodicities in the signal that were responsible for this difference. The results also show slightly different activation patterns in cortical regions which could have implications for attentiveness, and semantic and emotional processing. These results suggest that deficiencies in the temporal processing capabilities of the left Auditory Cortex play a major role in ASD speech processing.
ACKNOWLEDGEMENTS

I would like to thank Nicholas Lange, Sc.D., Janet Lainhart, M.D., Jeff Anderson, M.D., Ph.D., for their support and invaluable advice in this study. I would also like to thank Charles Bullock, Marie Bullock, Ph.D., Ryan Bullock, John Rosowski, Ph.D., Ana Luisa Abras, Ph.D., and Coco Hernandez, Ph.D., for their encouragement and support, without whom this project would not have been possible.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Glossary</strong></td>
<td>6</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td>8</td>
</tr>
<tr>
<td>1.1 The Source-Filter Model of Speech</td>
<td>10</td>
</tr>
<tr>
<td>1.1.1 Sound Sources in Speech</td>
<td>10</td>
</tr>
<tr>
<td>1.1.2 The Vocal Tract Filter</td>
<td>11</td>
</tr>
<tr>
<td>1.1.3 Extraction of Information about the Source and Filter</td>
<td>15</td>
</tr>
<tr>
<td>1.2 Neural Representations of Speech</td>
<td>18</td>
</tr>
<tr>
<td>1.2.1 Representations of Sound in the Cochlea</td>
<td>18</td>
</tr>
<tr>
<td>1.2.2 Representation of Sound in the Upper Auditory Pathway</td>
<td>23</td>
</tr>
<tr>
<td>1.2.3 Speech Processing in AC</td>
<td>28</td>
</tr>
<tr>
<td>1.2.4 Effects of Sound and Speech on Other Regions</td>
<td>29</td>
</tr>
<tr>
<td>1.3 ASD and Speech</td>
<td>31</td>
</tr>
<tr>
<td>1.3.1 ASD Symptoms Pertaining to Language Development</td>
<td>32</td>
</tr>
<tr>
<td>1.3.2 ASD and Auditory Processing</td>
<td>33</td>
</tr>
<tr>
<td>1.3.3 Neural Underpinnings of ASD Sound Processing</td>
<td>34</td>
</tr>
<tr>
<td>1.4 Specific Aims</td>
<td>35</td>
</tr>
<tr>
<td><strong>2 Methods</strong></td>
<td>37</td>
</tr>
<tr>
<td>2.1 Subjects and Scanning</td>
<td>37</td>
</tr>
<tr>
<td>2.2 Brain Regions</td>
<td>37</td>
</tr>
<tr>
<td>2.3 Auditory Stimuli</td>
<td>38</td>
</tr>
<tr>
<td><strong>3 Results</strong></td>
<td>45</td>
</tr>
<tr>
<td>3.1 Group 1 (Inferior Colliculus, left Thalamus)</td>
<td>45</td>
</tr>
<tr>
<td>3.1.1 Inferior Colliculus</td>
<td>45</td>
</tr>
<tr>
<td>3.1.2 Left Thalamus</td>
<td>47</td>
</tr>
<tr>
<td>3.2 Group 2 (Auditory Cortex, left Posterior Insula)</td>
<td>47</td>
</tr>
<tr>
<td>3.2.1 Auditory Cortex</td>
<td>47</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>AC</td>
<td>Auditory Cortex</td>
</tr>
<tr>
<td>AI</td>
<td>Primary Auditory Cortex</td>
</tr>
<tr>
<td>AIC</td>
<td>Akaike Information Criterion</td>
</tr>
<tr>
<td>AII</td>
<td>Belt of Auditory Cortex</td>
</tr>
<tr>
<td>AIII</td>
<td>Parabelt of Auditory Cortex</td>
</tr>
<tr>
<td>AM</td>
<td>Amplitude Modulation</td>
</tr>
<tr>
<td>ANF</td>
<td>Auditory Nerve Fiber</td>
</tr>
<tr>
<td>ASD</td>
<td>Autism Spectrum Disorder</td>
</tr>
<tr>
<td>autocorr</td>
<td>Autocorrelation</td>
</tr>
<tr>
<td>BM</td>
<td>Basilar Membrane</td>
</tr>
<tr>
<td>BOLD</td>
<td>Blood Oxygen Level-Dependent</td>
</tr>
<tr>
<td>ERP</td>
<td>Event-Related brain Potentials</td>
</tr>
<tr>
<td>FI</td>
<td>Front angular belt of Insula</td>
</tr>
<tr>
<td>FM</td>
<td>Frequency Modulation</td>
</tr>
<tr>
<td>fMRI</td>
<td>functional Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>fs</td>
<td>Frequency Similarity</td>
</tr>
<tr>
<td>HG</td>
<td>Heschl’s Gyrus</td>
</tr>
<tr>
<td>IC</td>
<td>Inferior Colliculus</td>
</tr>
<tr>
<td>IQ</td>
<td>Intelligence Quotient</td>
</tr>
<tr>
<td>LPC</td>
<td>Linear Predictive Coding</td>
</tr>
<tr>
<td>MEG</td>
<td>Magnetic Encephalography</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>PI</td>
<td>Posterior Insula</td>
</tr>
<tr>
<td>pitchvar</td>
<td>Pitch Variation</td>
</tr>
<tr>
<td>PM</td>
<td>Phase Modulation</td>
</tr>
<tr>
<td>REML</td>
<td>Restricted Maximum Likelihood</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of Interest</td>
</tr>
<tr>
<td>stimit</td>
<td>Stimulus Iteration</td>
</tr>
<tr>
<td>TD</td>
<td>Typically Developing</td>
</tr>
<tr>
<td>TMG</td>
<td>Transcranial Magnetic Simulation</td>
</tr>
<tr>
<td>TR</td>
<td>Time of Repetition</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Autism Spectrum Disorder (ASD) is a brain-based disorder that underlies symptoms largely related to delays in social and linguistic development [48]. It is believed that ASD involves different strategies of sensory processing. Although it sometimes corresponds to mental retardation, ASD does not necessarily affect IQ, and many high-functioning ASD individuals are successful in fields which require a facility with abstractions and extreme attention to detail. For example, the widely known autistic Temple Grandin revolutionized the meat industry by noticing how minute and previously unnoticed details of the slaughterhouse affected the behavior of the animals [25].

Classified as DSM IV, is divided into three major categories - Autism, Asperger syndrome, and Pervasive Developmental Disorders Otherwise Not Specified [48]. Although ASD is defined in terms of behavioral symptoms [48, 5], the consistency of these symptoms over different patients and the apparent heritability of ASD have led to the belief that ASD has a neuropathology with genetic origins. This is supported by the fact that symptoms appear in the same periods of child development, from 6 months onward [61]. Although no genetic cause has been identified, studies of siblings and twins show that autism is highly heritable [1].

Although the neural mechanisms for autism are not known, hypotheses about the neuropathology of autism often focus on cytological differences which have global repercussions for the functionality of the brain [12, 65, 7]. In these hypotheses, many brain regions show strong internal cohesion and functionality but weaker connectivity with other regions. For example, Schmitz and Rezaie [65] postulate a degree of connectivity within the cerebral cortex and weaker connections between this region and the frontal lobe. This would have wide-spread implications for at-
tention, orientation of auditory and visual stimuli, and novelty detection. Other hypotheses are an imbalance in excitatory and inhibitory networks, and an under-development in the mirror neurological system, which activates when an animal observes another animal performing the same action [56].

The neural mechanisms for speech and language are highly distributed across the brain [47] and many symptoms of autism specifically involve speech and language development [5]. Therefore, it is reasonable to investigate how the mechanisms of auditory and speech processing differ between autistic and typically developing (TD) subjects. Magnetic encephalography (MEG) and neuroimaging studies have shown that there are significant differences in ASD responses to entire sentences in various brain regions [4, 36, 60, 8]. There have also been behavioral and neuroimaging studies on how ASD subjects respond to simpler sounds such as pure tones, amplitude-modulated (AM) tones, and frequency-modulated (FM) tones [9, 11]. These studies implicate ASD subject’s ability to process wideband stimuli which have a high amount of variation across time, and that it is a combination of spectral and temporal complexity which elicits ASD-specific responses. To our knowledge, no study has yet quantitatively related ASD responses to temporal complexity in the speech stimulus. By imaging different regions along the auditory pathway in response to stimuli which systematically vary along various source-filter parameters, it may be possible to localize regions where there are significant subject group differences and to relate these differences to these regions’ known signal processing capabilities.

Section 1.1 of this chapter reviews the source-filter model of speech as it pertains to the design of the stimuli of this study. Section 1.2 reviews salient regions on the auditory pathway for investigating autism’s relationship to speech. Section 1.3 reviews autism’s relation to auditory processing, speech, and language. Section 1.4 states the specific aims and hypotheses of this study. In this study, the terms ASD and autistic will be used interchangeably, due to the prevalence of similar language-related symptoms across all categories of ASD [48, 5]. A thorough introduction to the basics of acoustic theory and digital signal processing is included at the end of this study, in the Appendix. This Section covers the basic models for propagation of sound, the characterization of click-like, periodic, and noisy sound sources, mathematical models for convolution of the source and filter, and basic digital signal processing concepts.
1.1 The Source-Filter Model of Speech

The source-filter model of speech models the acoustic interactions of various sound sources with the vocal tract. This model relates a sound source, which a physical event which originates the propagation of acoustic energy, and a filter, which is a physical structure which somehow transforms this acoustic energy. In this model, the glottis serves as the source and the vocal tract serves as a filter. The speech signal is a convolution of this glottal source and vocal tract filter.

1.1.1 Sound Sources in Speech

The sound sources in speech can be broadly characterized as periodic, noisy, or click-like. Periodic sources arise from the glottis. Noisy sound sources are turbulence at an obstruction in the glottis or vocal tract. Click-like sources depend on the sudden release of air after a complete obstruction of air has been released, in either the glottis or the vocal tract.

The glottis is a sound source for vowels and any consonant requiring voicing. It is a cartilagenous structure situated above the epiglottis and the trachea, and below the pharynx. Two strips of elastic tissue, the vocal folds, connect the arytenoids to the thyroid cartilage on the anterior of the glottis. These are two pointed cartilagenous structures on the posterior ridge of the glottis which can rotate inwards or outwards. Because the points connect to the vocal folds, inward rotation of the points spreads the vocal folds apart, while outward rotation brings them in contact with one another. [71]

When the vocal folds are brought together, airflow from the lungs can exert pressure on them. This pressure moves them upward while spreading them apart. After they are spread apart, the tension in the folds brings them downward and back together. Steady airflow ensures that these excursions occur periodically. These periodic excursions create a series of pulses of volume velocity which propagate upwards into the vocal tract. Because the vocal folds have a greater excursion than incursion, the waveform they produce has distortion, and therefore is a harmonic complex in which the frequency of the pulses is the fundamental frequency. Fundamental frequency is controlled by the height of the glottis. The glottis can move up the spinal chord, stretching the vocal folds and increasing tension. This decreases the time it takes for the vocal folds to come together after an excursion, and thus increases the frequency of the excursions [71].
Figure 1.1: A view of the glottis from the back of the oral cavity looking downward, showing the vocal folds and arytenoids. (a) shows the arytenoids in their resting position. (b) shows the points of the arytenoids moving inward and drawing the vocal folds apart. (c) shows the points of the arytenoids moving outward and drawing the vocal folds together. From Stevens [71].

Speech also depends on other sound sources in constrictions of the vocal tract and the glottis under certain configurations. One major class of sound sources is noisy turbulence arising from air flow through constrictions in the vocal tract. The other major type of sound source is a single click-like pulse of volume velocity arising from releasing the complete obstruction of the airflow.

Different phonemes - or discrete units of sound in language - depend on different sound sources. A vowel depends entirely on the glottis, although breathy vowels, such as those found in Gujarati, also depend on turbulence in the glottis. Voiced consonants, such as /z/ or /g/, feature the periodicity generated by the vocal folds. Fricatives, such as /h/, /th/, /dh/, /s/, or /sh/, feature noisy turbulence generated in the glottis or a constriction in the vocal tract. Stops, such as /t/, /d/, /p/, or /k/, result from a released obstruction in the vocal tract, which resembles a click followed by a brief period of turbulence. In affricates, such as /ch/ or /j/, the release after a complete obstruction of airflow is followed by more turbulence in the constriction.

1.1.2 The Vocal Tract Filter

The vocal tract, extending from the glottis to the lips, can be portioned into tube-like cavities by the pharynx and tongue. The vocal tract can increase the amplitude of different frequency components in the sound source, in a process called resonance. In source-filter terms, the glottis and noisy constrictions in the vocal tract serve as sources, while the vocal tract serves as a filter. It is important to note that
the sound source can be any physical event which introduces acoustic energy into a medium and the filter is an environment which alters this energy. For example, if someone is speaking in a room, it is also possible to define the vocal tract excited by the glottis as a source and the room as the filter.

1.1.2.1 Boundary Conditions for Resonance

Certain constrictions in the vocal allow the vocal tract to be modelled as a series of concatenated tubes. A major property of a tube is resonance. This occurs when a tube is exposed to a sinusoid at certain frequency which causes the tube to amplify this sinusoid. This frequency depends on the length of the tubes and whether the ends of the tubes are open or closed. These open or closed openings mandate boundary conditions, and these boundary conditions determine the resonant frequency. The boundary condition for an open end of the tube is that the pressure must be 0 and the volume velocity must be at a maximum. This is because Newton's law is encoded in the wave equations (see Equations 1.3, 1.1, and 1.2).

As was shown above, pressure is analogous to force and acceleration involves particle velocity. Since volume velocity is particle velocity over a cross-sectional area, acceleration implies a change in volume velocity. A point where pressure is 0 is a point where the acoustic medium has fully accelerated, and volume velocity is at its maximum. The enclosure of this medium inside the tube ensures that the wave's energy drives the medium back and fourth in a directed fashion. If this is done at maximum velocity, the maximum amount of energy will enter the tube. If the 0 point for pressure were outside the tube, not all of this energy would enter the tube. In this way, the medium at the open end of the tube can act as a vibrating piston-like sound source.

At a closed end, the medium cannot move, and volume velocity at this point is 0. A point where volume velocity is 0 is also a point where pressure has not accelerated the medium, and is therefore at its maximum value. A wave where pressure is at its maximum at the closed end and minimum at the open end is a wave where the most energy is conserved in the system.

The effects of resonance on a wave are understood by the relation of pressure to volume velocity, \( p = zu \), where \( z \) is an acoustic impedance. Impedance is the loss of energy in the propagation of the wave. At an open end of a resonating tube, if pressure is 0 and volume velocity is at a maximum, then impedance is 0. Therefore, as the wave enters the tube, energy will accumulate in the tube. As
energy accumulates, maximum volume velocity at the open end increases. The piston-like motion of the medium at the open end then acts as a sound source at this resonant frequency, thus amplifying the original stimulus frequency.

1.1.2.2 Types of Resonators

**Tube Resonators.** In a resonating tube of length $L$, which is open at both ends, pressure is 0 and volume velocity is maximum at these ends. A pressure and volume velocity wave of a resonant frequency must have at least a half-wavelength fit into the tube. Therefore, the resonant frequencies will be integer multiples of $\frac{c}{2L}$. In a tube closed at both ends, the conditions will also be satisfied by these resonant frequencies, although the 0 point in pressure will be in the middle of the tube. In a tube open at one end and closed at the other, a wave must have a quarter-wavelength fit into the tube. The resonant frequencies are odd-numbered multiples of $\frac{c}{4L}$.

**Helmholz Resonators.** The other major type of resonance is the Helmholz resonance. This occurs when a tube open at one end is concatenated to a tube open at both ends, and the cross-sectional area of the tube open at both ends is large enough to create a mass-spring effect in the acoustic medium. There are two conditions for the pressure wave. First, its wavelength must be large enough to create
only a minor pressure drop across the resonator. In such a situation, the medium in the tube open at both ends behaves similarly to a rigid-body mass moved by a force, and the medium in the tube closed at one end behaves as a spring which is compressed by the mass. Energy from the mass is transferred to the spring as it compresses, and transferred back to the mass as it extends. Secondly, the wave must accelerate the mass at the same time that the spring has fully extended. A wave which meets these two conditions allows the maximum amount of energy to accumulate in the resonator. The mass at the open end of the tube behaves as a piston-like sound source for this amplified stimulus. The frequency which causes this behavior depends on both the length of the mass-like element, the spring-like element \((L_M \text{ and } L_S, \text{ respectively})\), and the cross-sectional areas of these elements \((A_M \text{ and } A_S, \text{ respectively})\) - 

\[ F_H = \sqrt{\frac{A_M}{A_S L_M L_S}}. \]

**Concatenated Tube Resonators.** Resonators formed from concatenations of tubes have the resonances of the individual tubes, and any Helmholz interactions which the concatenation may produce. Therefore, a Helmholz resonator has the Helmholz resonance and the resonances of the two tubes.

**1.1.2.3 Modeling the Vocal Tract**

**Modeling the Filter.**

**Vocal-Tract Configurations for Vowels.** The cardinal vowels of English can be modeled in terms of these tubes. The model for /i/ is a concatenation of a tube open at both ends and a tube closed at one end. This has a Helmholz resonance and the resonances of the individual tubes. The models for /a/ and /e/ are two tubes closed at one end. /o/ is three tubes closed at both ends, concatenated against one another, and /u/ is two tubes closed at both ends.
Vocal Tract Configurations for Consonants. It is important to note that certain consonants are also steady-state activations of the vocal tract filter. If they are voiced, as in the case of /d/ or /g/, or if there is a noise source in the glottis, as in the case of /h/, the entire vocal tract will be excited, showing the resonances of the vocal tract. However, if the noise source comes from an obstruction in the vocal tract, as in the case of /sh/, /s/, /th/, and /f/, only the portion of the vocal tract anterior to the constriction will resonate. The posterior portion will serve as a source of zeros, or anti-resonances, which attenuate the signal at certain frequencies. The frequencies where the noise band attenuates indicate the length of the back cavity, and therefore the constriction.

Temporal Information in Speech. Clearly, speech is not a series of steady-state sounds, and there is just as much information conveyed by transitions in the vocal tract. Some vowels convey information in the change in the vocal tract configuration. /ay/, /ey/ or /ow/, for example, are diphthongs which are characterized by their change. Stop consonants such as /t/ or /k/ involve a complete obstruction of airflow, followed by a short burst of air. The change in the vocal tract resonances preceding and following the constriction indicates the position of the constriction. This is what is referred to as temporally significant information in speech.

1.1.3 Extraction of Information about the Source and Filter

Speech processing depends on extracting information about the filter and the source - especially regarding pitch. The interaction of a glottal source and a vocal tract filter can be modeled mathematically as a convolution. Extracting a spectrum for both the source and the filter is done by deconvolution.

1.1.3.1 Extraction of Information About the Filter

The neural mechanisms for deconvolution of the source and filter are not known, but a brief review of the engineering solutions demonstrates the complexity of the task.

Smoothing the Spectrum. One of the simplest auditory processing algorithms is to smooth the convolved spectrum. The spectrum of the speech signal is represented as the magnitudes of a discrete set of amplitudes for Fourier coefficients. The magnitude for each coefficient in the spectrum of the filter is the mean of the last $N$ coefficients, the coefficient, and the next $N$ coefficients [73].
Cepstral Analysis. Cepstral analysis (a term formed by reversing the first four letters of the word “spectrum”) performs a Fourier Transform on values computed from the convolved spectrum. First, the magnitudes of the Fourier coefficients of the convolved signal are squared. The logarithm is taken from these coefficients, and a Fourier transform is performed on them. The squared magnitude of these coefficients compose the cepstrum. The equivalents to frequencies in this scheme are termed quefrencies, a play on the word “frequency”. Since the fine structure of the convolution arises from the glottal source, it corresponds with higher que-frequency coefficients in the cepstrum. The filter corresponds to lower quefrencies. Therefore, isolating the lower and higher quefrencies has the effect of separating the source and filter [50].

Linear Predictive Coding. Linear Predictive Coding (LPC) is a mean-squares method. A discretized speech signal \( x(n) \) is compared to an estimation of this signal from the previous \( M \) components and a set of coefficients \( a: \hat{x}(n) = \sum_{i=1}^{M} a_i x(n - i) \). The mean-squared error, \( \epsilon^2(n) = (x(n) - \hat{x}(n))^2 \) is minimized to solve for the coefficients. These coefficients are then used to estimate a transfer function for the filter, the peaks of which are the resonances. [50]

1.1.3.2 Detection of other Features

There are other relevant features of a sound which have a variety of strategies for detection. Although the neural mechanisms of pitch perception are largely unknown, the methods applied in signal processing demonstrate the complexity of the task. There are two major classes of algorithms for these tasks used in signal processing - spectral and temporal.

Pitch. Pitch detection involves making a reliable judgement on fundamental frequency, and there are spectral and temporal approaches to this. Temporal algorithms focus on the signal in the time domain. If all harmonics have the same phase, they will all have maximum and minimum amplitude at the same time as the tone at the fundamental frequency. Therefore, the maxima and minima of each harmonic sum to produce the maxima and minima in the entire signal. They can therefore be used to infer the fundamental frequency. Similar inferences can be made from zero-crossings in the signal. The envelope of the waveform, or its general trajectory over time, is another means of identifying pitch. One of the more sophisticated methods of temporal analysis involves autocorrelation, a measure for the degree of periodicity in the stimulus. The fundamental frequency defined with a time constant \( \tau \) as \( \frac{1}{\tau} \). Because periodicity from the fundamental frequency can
be found in individual harmonics, the time constant which maximizes the integral:
\[
\tau(\tau) = \int x(t)x(t - \tau)dt
\]
(1.1) is also the time constant corresponding to the fundamental frequency [13].

Spectral algorithms analyze the spectrum of the speech signal. A common class of algorithms for pitch recognition is comb-fitting, in which a comb-like template, whose “teeth” are at the harmonic frequencies, is fit to the spectrum of the signal. A signal for a harmonic complex will fit each template to a certain degree, and pitch is determined by the template which maximizes this fit [13]. This requires multiple templates for multiple pitches. Another important feature of comb-fitting algorithms is that wideband signals like noise and clicks can also activate these templates.

**Detection of AM and FM.** Speech depends highly on temporal variation. Therefore, mechanisms to detect changes in amplitude and frequency are crucial. These mechanisms can be detected using AM and FM stimuli. Although their amplitude and frequency modulate periodically, the mechanisms which detect this modulation may also be able to detect transient modulations. Because it must track changes in amplitude, the AM-sensitive mechanism would account for any differences in amplitude of vocalization, and transitions from high-amplitude intervals of the signal, such as vowels, to low amplitude portions of the signal, such as certain fricatives or stops. An FM-sensitive mechanism would be needed to track changes in frequency associated with modulations in pitch and vocal tract resonances.

Spectrally, AM can be detected by taking advantage of the equivalence:
\[
\sin(2\pi f_A t) \sin(2\pi ft) = \sin(2\pi ft) + \sin(2\pi (f + f_A)t) + \sin(2\pi (f - f_A)t)
\]
(1.2)
where \( f > f_A \). These three components appear in the spectrum, and it is possible to infer these frequencies. Temporally, if the modulator frequency is known, differences in consecutive peaks of amplitude in the wave can be used to track AM gradients. Another means of detecting AM is to analyze its time-domain “envelope”. An envelope is the gradual pattern of AM, independent of the fine structure of the modulator wave. The envelope can be taken by extracting a Hilbert transform of the wave, defined in Equation 1.5, taking its absolute value, and low-pass filtering the signal. Then, AM can be extracted by performing a Fourier Transform on the envelope.
A widely used method for tracking FM to detect phase modulation. The \( f_{FM} \) sinusoid in Equation 1.11 could also be interpreted as a phase modulation, as long as the amplitude of this term were sufficiently small (< \( \frac{\pi}{2} \)). In situations where \( \beta < \frac{\pi}{2} \), it would be possible to take advantage of the trigonometric identity \( \sin \alpha - \sin \beta \approx \alpha - \beta \). To do this, a sinusoid at the carrier frequency must be taken \( \sin(2\pi ft) \), and then compared with a phase-shifted sinusoid with a phase modulation term \( \cos(2\pi ft + x_{PM}(t)) \). It is also to obtain \( \cos(2\pi ft) \) and \( \sin(2\pi ft + x_{PM}(t)) \).

Applying a trigonometric identity yields:

\[
\sin(2\pi ft) \cos(2\pi ft + x_{PM}(t)) - \cos(2\pi ft) \sin(2\pi ft + x_{PM}(t)) \\
= \sin(2\pi ft - 2\pi ft - x_{PM}(t)) \approx -x_{PM}(t) \\
= \frac{\sin(2\pi ft - 2\pi ft + x_{PM}(t))}{2} - \frac{\sin(2\pi ft + 2\pi ft + x_{PM}(t))}{2} \\
= \frac{\sin(2\pi ft - 2\pi ft + x_{PM}(t))}{2} - \frac{\sin(4\pi ft + x_{PM}(t))}{2} \\
\approx x_{PM}(t)
\]  

Because the second term of the difference is twice the carrier frequency, it can be filtered out, and \( x_{PM}(t) \) can be computed as if there were no second term. Equation 1.3 can also be used to compute amplitude modulation, if the carrier sinusoid in Equation 1.2 is subtracted.

### 1.2 Neural Representations of Speech

In the auditory pathway, there is evidence for both temporal and spectral representations of the acoustic signal. In a spectral representation, there are a series of mechanico-neural channels, each of which represent a stimulus at a different range of frequencies. At rest, these neurons have a spontaneous firing rate. When a channel is activated by a stimulus, the neuron’s firing rate is above the spontaneous rate. The temporal representation preserves the periodicity of the stimulus in the firing patterns of the neurons [23].

#### 1.2.1 Representations of Sound in the Cochlea

The ear is a transducer which represents a signal acoustically, mechanically, electrochemically, and neurally. An acoustic signal enters the ear canal. The signal compresses the tympanic membrane, which is coupled to the bones of the middle ear.
Figure 1.4: A schematic diagram of the cochlea, showing the position of the basilar membrane. The unfolded portion shows the vibration patterns in response to an acoustic stimulus. From [54].

Figure 1.5: A schematic diagram of an unfolded basilar membrane, with a conceptual model as a series of discrete resonators, in which each region responds to a frequency determined by the mass and springiness of the membrane at that region. From [67].
These bones mechanically transduce the signal to the round window of the cochlea. The cochlea is a spiral fluid-filled tube which is internally bisected by the basilar membrane (BM).

1.2.1.1 Transduction in the Cochlea

Acoustic-to-Mechanical Transduction. The BM is an acoustic-to-mechanical transducer. Consecutive regions of this membrane can be modeled as a series of discrete mechanical resonators, as shown in Figure 1.5. The mass and stiffness of one of these regions can be modeled as two springs which connect a mass to opposite sides of the tube. In this way, regions of the BM behave as mass-spring oscillators similar to Helmholtz resonators, and each region has a resonant frequency which derives from the mass and stiffness. From the base of the cochlea to its apex, the mass and stiffness vary such that the BM represents a contiguous range of frequencies, with higher frequencies at the base. The range of the human ear is 20-20000 Hz [23].

The signal enters the fluid of the cochlea at the round window. This fluid acts as an acoustic medium. As the signal displaces the BM, frequency components in the signal activate the corresponding resonant BM regions. Because of the mechanics of the BM, the wave travels from base to apex. At the regions corresponding with frequency components in the signal, the membrane oscillates upwards and downwards at the resonant frequency. The degree of displacement is proportional to the amplitude of the frequency component. In this sense, the entire BM represents a spectrum mechanically, with different regions representing different frequency intervals. This strategy of spatially representing a spectrum is called tonotopy [23].

Mechanical-to-Neural Transduction. On the BM, there are inner hair cells which act as mechanical-to-electrochemical transducers. These are epithelial cells with a bundle of stiff connected tubes, or stereocilia, on their apex. As the membrane displaces, these tubes deflect, opening ion channels which allow potassium to enter the cell. This potassium depolarizes the cytoplasm, causing ion channels selective for calcium to open at the base of the cell. Each inner hair cell is innervated by nerve fibers, the auditory nerve fibers (ANF’s). At each synapse, the inner hair cell contains a structure which binds several vesicles together. Incoming calcium causes this structure to release vesicles into the synapse, as well as the release of vesicles which are not attached to this structure. These vesicles travel across the synapse, causing the auditory nerve fiber to hyperpolarize and fire an action potential. It is important to note that vesicles can also be released spontaneously without depolarization. Therefore, the neuron has a spontaneous firing rate when the hair
cell is at rest, and a higher-than-spontaneous rate when the hair cell is depolarizing [23].

Figure 1.6: A diagram of an inner hair cell. Basilar membrane displacement causes stereocilia to deflect, allowing an influx of potassium ions. These ions promote vesicle release to the afferent nerve fiber. From [71].

1.2.1.2 Spectral and Temporal Representations

Spectral Representations. The cochlea's method of transduction allows for a spectral representation of the signal. Because inner hair cells associate consecutive regions of the BM to consecutive ANF's, there is a tonotopy in which ANF's act as filtering channels for a given frequency interval. An ANF's firing at the spontaneous rate represents a lack of frequency components in this interval, while a higher-than-spontaneous firing rate in an ANF represents the presence of a frequency component [23].

Temporal Representation. The temporal strategy for representing frequency components in a signal is called phase locking. Because a region of the BM oscillates at its resonant frequency, because hair cells depolarize with the displacement of a BM
region, and because vesicles can be released upon displacement, it is possible to reproduce the signal’s periodicity in the ANF. However, the transduction process of the inner hair cell introduces certain temporal limitations, which make it possible for an ANF to fire at a maximum rate of approximately 250-300 spikes/sec. Therefore, it is possible for an ANF to fire with every cycle in a stimulus at or below 250-300 Hz. Although an ANF cannot completely follow the stimulus above this frequency, it will probably spike at some cycles, preserving some of the stimulus’ periodicity. This could appear in a histogram as a fluctuation in the average firing rate at the same frequency of the cycle. For example, a stimulus of 300 Hz or 900 Hz could have an average firing rate at at 150 spikes/sec, although there will not be a spike for every cycle of the sinusoid [23].

1.2.1.3 Lower Auditory Pathway Representation of Other Sound Features

Pitch. Because the ANF can preserve periodicity in the stimulus, pitch can be represented in the ANF. ANF’s at a harmonic of the fundamental frequency under 250 Hz can completely follow the stimulus. With harmonics above this frequency, the spikes can reproduce some of the periodicity of fundamental frequency and some of the periodicity of other harmonics. Pitch can also be inferred from multiple ANF’s. If a single ANF at a harmonic can represent periodicity of lower harmonics and the fundamental frequency, then the frequency at which multiple ANF’s fired simultaneously would be the fundamental frequency. A mechanism designed to detect this could then serve as a way to derive pitch [69].

AM and FM. The spectral and temporal representations of sound in the cochlea can also encode other features of a stimulus as well. An AM stimulus with a frequency can be spectrally represented by tone shown in Equation 1.2. In a temporal representation, the ANF will most likely phase-lock to the AM frequency. This is because the greatest amplitude in an AM stimulus will occur periodically with the AM frequency. Because the ANF’s firing rate is limited to 250 Hz, times of maximum amplitude in the waveform will be most likely to cause the ANF’s to fire. FM has spectral and temporal representations as well. An FM tone could be spectrally represented by alternating increases in firing rates in adjacent groups of ANF’s. Phase-locking in the ANF could also represent FM. In a phase-locked firing pattern, consecutive ANF spikes may repeat consecutive spikes that occur because of changes in frequency.
1.2.1.4 Other Auditory Pathway Structures

As shown in Figure 1.7, the auditory pathway is composed of several stages, many of whose functionality is still not fully understood. One common feature of the lower auditory pathway is the preservation of tonotopy. The ANF's in the spiral ganglion relay the signal to the cochlear nucleus. Like the cochlea, the cochlear nucleus contains a tonotopic representation of the spectrum. The signal is relayed through the cochlear nucleus' projections through lateral lemniscus to the inferior colliculus (IC). From here, the signal travels through the medial geniculate body and the thalamus, into auditory cortex (AC).

The IC is important because all auditory information must pass through this structure to get to higher brain regions. In doing so, the evidence suggests that it acts as a sensory integrator for auditory and other sensory information. Moreover, there is evidence that this region is sensitive to both AM and FM frequencies [57] and to higher-order processes in AC [59].

1.2.2 Representation of Sound in the Upper Auditory Pathway

There is evidence that both the tonotopic and spectral representation of the signal propagate upwards into AC. However, from here other brain regions involved in auditory perception do not demonstrate this representation, suggesting that regions connected to AC are involved in higher-order cognitive processing of the signal.

1.2.2.1 Spectral and Temporal Representations in AC

The lower auditory pathway’s spectral and temporal representations are repeated in AC, whose neurons demonstrate both tonotopy and extreme sensitivity to spike timing that allows it to detect temporal components in the signal precisely. Since changes in a signal’s frequency and amplitude can be encoded in AM and FM, these temporal mechanisms are predominantly specialized to detect these two quantities in a signal.

The Neuroanatomy of AC. Located on the superior temporal gyrus, this region is divided into a core (AI), a belt surrounding the core (AII), and a parabelt (AIII). AI has extensive two-way connectivity with AII, and AIII has connectivity with the contralateral AIII. There is no known connectivity between AI or AII and the contralateral AC. At the anterior is Heschl’s Gyrus (HG) [68].
Figure 1.7: A diagram of the auditory pathway, showing the major structures from the cochlea to auditory cortex. From [33].
Figure 1.8: A diagram of auditory cortex, showing a lateral view of the macaque (a) and human brain (b). In these schematic illustrations, dashed lines indicate where overlying parietal cortex has been removed. From Semple [68].
Spectral and Temporal Representations in the AC. There is evidence that the spectral and temporal representations of a signal have propagated into AC. Single-cell recordings of various mammals show that Al has several rostro-caudal tonotopic gradients which place higher frequencies caudally [68].

There is also evidence that this frequency axis corresponds with a lateral axis which is sensitive to temporal changes in the signal [68]. Cells along this axis demonstrate sensitivity for AM and FM. In recording from Al in cats, Eggermont [17] has demonstrated that AM tones, clicks, gamma tones (tones which rise rapidly and attenuate gradually), and AM noise, and time-reversed gamma tones - all of which have different temporal envelopes - have unique representations in Al. In this study, the uniqueness of these responses depended on the degree of synchronization of cortical nerve firing patterns to the envelope of the waveform. The most reliable synchronization occurred with stimuli with a rapid onset - clicks and gamma tones. Synchronization decreased significantly with time-reversed gamma tones, AM noise, FM tones, and AM tones, respectively. The decreased sensitivity to periodicity in the latter two cases suggests that this mechanism is most responsive to changes in the envelope of the stimulus - which are more apparent in noisy, click-like, and rapid-onset periodic stimuli.

Possible Mechanisms for Temporal Sensitivity. The evidence for the exact mechanism of FM sensitivity presents a complex picture, and there is strong evidence that AM and FM detection rely on the same mechanism. Ding and Simon [14] used MEG to observe responses to FM and AM modulation, finding distinct auditory steady states for AM and AM-FM stimuli, the latter of which shows phase modulation with AM and FM. Luo et al. [44] performed a similar experiment, arriving at a comparable conclusion. Drawing on the fact that phase modulation can encode both AM and FM, as is shown in Equation 1.3, they suggested that one type of neuron uses this method to encode FM and AM at lower AM frequencies. These neurons account for synchronization to the temporal envelope. The authors also noticed that higher AM frequencies produce a different auditory steady state response. This led the authors to suggest that there were neurons which could encode AM and FM through phase modulation and neurons sensitive only to AM.

These results are supported by Elhilali et al. [18], who based on their study on the paradoxical observation that cortical neurons, whose cytoarchitecture seemed specialized for temporal precision, could only encode AM of up to 20 Hz. Recording from ferrets, they demonstrated that cortical neurons show a remarkable temporal precision in response to AM tones, such that they can follow a carrier stim-
ulus of above 200 Hz. However, this response disappears when there is a lower AM amplitude. They attributed this to a cortical inhibitory mechanism or corticothalamic excitatory mechanism which acted as a gateway for an AM signal. When the AM amplitude was sufficient, they noticed that not only could the cortical neuron follow the carrier, but that differences in spike phase could encode FM. The interaction of cortical neurons with this inhibitory mechanism allowed for a temporally precise representation of both AM and FM.

1.2.2.2 Hemispheric Specialization for Temporal Precision

Although both left and right AC have tonotopy and sensitivity to FM and AM, there is evidence for hemispheric differences which result in specialization of the left AC for temporal precision and the right for spectral precision.

fMRI Studies. Functional magnetic resonance imaging (fMRI) studies demonstrate this specialization in humans. Zatorre and Belin [75] presented subjects with tone complexes with differences in the number of tones and frequency transitions. They found that more spectrally complex stimuli disproportionately activated the right AC, while the more temporally complex stimuli disproportionately activated the left AC. This has implications for speech perception. Zatorre [76] showed that the left HG activated with more temporal precision in response to a speech stimulus.

Physiological Indications. While speculating as to the causes of this hemispheric difference, Zatorre and Belin [75] suggested that there are differences in the makeup of the cortical cells themselves. In-vivo volumetric measures from HG from MRI scanning have suggested that there is a greater concentration of white matter in the left Heschl’s gyrus (HG) than the right [52]. Since white matter is myelinated tissue, and greater myelination corresponds with greater conductivity, there is a potential physiological basis for the functional differences between left and right AI.

1.2.2.3 AC and Pitch

A third property of a speech signal which could have temporal and spectral means of representation in AC is pitch, although the exact strategy is not known. There is physiological evidence for brain regions devoted to pitch processing, but the exact role of spectral and temporal forms of pitch processing is unknown [13]. Although it was originally believed that HG contained a mechanism for pitch processing due to the region’s response to dichotic noise [55], a subsequent study by Hall and
Plack [28] showed other regions responding to pitch, such as the temporo-parieto-occipital junction or prefrontal cortex. The evidence strongly suggests the use of a spectral algorithm, but it is not known whether this is performed in conjunction with a temporal algorithm. The pure “template” model faces the difficulty that harmonic templates are shown to arise in pre-natal mammals [13]. Shamma and Klein [69] developed an algorithm which exploits the periodicity coming from the cochlear filters. Their algorithm postulates a series of coincidence detectors. Synchronous spikes coming from cochlear filters at the harmonics indicates a harmonic structure. This allows harmonic templates to be learned from non-harmonic sounds such as physiological noise.

1.2.3 Speech Processing in AC

Although spectral and temporal representations play a role in speech perception, there is ample evidence for the promenence of temporal information over spectral information.

1.2.3.1 Perceptual Evidence for Temporal Information

Research on cochlear implants has demonstrated the prominence of temporal information in speech. Cochlear implants use channels, which sum the amplitude of a series of frequency bands and run a proportional electric current onto the ANF’s corresponding to this frequency band. Friesen et al. [21] have demonstrated that only seven channels are adequate for speech perception in cochlear implants, and that similar results can be achieved by presenting hearing subjects with AM noise whose AM is proportional to the amplitude within a series of channels associated with contiguous frequency bands. Moreover, Shannon et al. [70] have demonstrated that many consonants can be effectively perceived when represented with as few as three AM channels.

1.2.3.2 AM and FM Sensitivity in Speech Processing

AM and FM sensitivity seem to play a key role in perceiving this temporal information. Tallal et al. [72] have shown that subjects with language impairment do not have difficulty perceiving steady-state vowels, but do have difficulty with consonants and diphthongs. Moreover, Drullman et al. [15] show that filtering the temporal envelope from a speech signal seriously affects perception of consonants and diphthongs. Both consonants and diphthongs with a steady-state pitch require a transition within the vocal tract filter. In the spectrum, changes in the vocal tract filter result in changes in the amplitude of the harmonics.
As was stated above, pitch is a major component in speech processing, and perceiving changes in pitch seems to depend on temporal sensitivity. Sayles and Winter [64] studied pitch perception by running a harmonic structure through various channels that preserve both the spectral and temporal indications of pitch, and degrade the temporal information by introducing reverberation. Reverberation interferes with the representation of periodicity in the ANF. The study found that degrading temporal information in the channels seriously degraded pitch perception.

1.2.4 Effects of Sound and Speech on Other Regions

AC is the highest region in the auditory pathway in which there is a known tonotopic representation of the spectrum. This being said, there are regions which respond to auditory stimuli in ways that suggest higher-level tasks such as syntactic parsing, lexical access, semantic computation, and pattern recognition. Determining the functionality of these regions is a very difficult task which is prone to speculation.

1.2.4.1 The Insula

In mammals in which the region has been investigated, the Insula (Brodmann areas 13 and 16) serves a number of sensory regions, including the visceral sensory, motor, motor association, vestibular, and somatosensory areas. Its role in auditory processing has yet to be defined [6].

**Brain Architecture and Connectivity.** Humans possess a larger left than right insula [6]. Immunostaining reveals that it is divided into a frontal agranular belt (FI), a middle dysgranular belt, and a posterior agranular belt (PI) [6]. Its connectivity to the auditory pathway suggests that it processes several levels of representation of the input. Recording auditory evoked potentials from the medial geniculate body shows significant connectivity with the parainsular field and the PI [10]. Moreover, PI receives projections from AIII, and has some limited connectivity to AI and AII [68]. Each insula projects fibers across the corpus callosum in close proximity to fibers from the auditory cortex [68].

**Bilateral Functionality.** There is limited evidence that some of the insula's functionality in auditory processing depends on bilateral interaction. In Hyman and Tranel [34], a patient had two strokes affecting both PI's, the right opercular cortex and the right temporal lobe. Immediately after the first stroke, she had rapidly improving hemiparesis. After the second, she did not react to any sound or speech
stimuli and was totally mute. She eventually recovered but still had receptive aposodia, receptive amusia, poor pitch discrimination, poor tone discrimination, and poor identification of environmental sounds and recognition of famous voices.

The insula seems to activate bilaterally when a subject performs higher-order pattern recognition. Engelien et al. PET-scanned stroke victims and healthy subjects [19]. For the latter, passive listening to sounds activated the auditory cortices bilaterally. One stroke victim had two consecutive strokes in the perisylvian areas that included the insulae bilaterally, resulting in auditory agnosia immediately after the second stroke. Eight years later, he could discriminate all sounds and occasionally identify environmental sounds, but had problems with speech perception. fMRI studies by Kiehl et al. showed greater activation of the insulae with target stimuli than with novel stimuli [37]. Paulesu et al. assessed the role of verbal short-term memory [51]. When subjects were presented with letters and produced rhymes, there was significant bilateral activation of the insulae. The authors concluded that the insulae were involved in sub-vocal processing.

Unilateral Functionality. There is evidence that the left insula is involved in processing rhythm stimuli and temporal aspects of prosody. Platel et al. scanned the cerebral structures involved in the appreciation of music, finding that the left insula was involved in rhythm detection [53]. Rumsey et al. examined the neural pathway involved in phonological versus orthographic strategies of word recognition, sounding out the word versus visually and linguistically analyzing the written word in order to read it [62]. Orthographical and phonological pronunciation tasks activated the left insula. Moreover, the left insula in non-human primates demonstrates sensitivity to vocalization. Remedios et al. have shown that rhesus monkeys exhibit marked activation of the left-lateral insula in response to vocalization [58].

These results seem to indicate that left PI is highly responsive to temporal stimuli, whereas the bilateral functionality of the insula seems more associated with more stationary sounds. However, the insula is so multi-faceted in its sensory functionality that it is difficult to draw any conclusions.

1.2.4.2 Posterior Superior Temporal Gyru (Wernicke's Area)

This area, located in the posterior section of the superior temporal gyrus, in Brodmann area 22, encircles AC along the Sylvian fissure. This area is traditionally believed to be associated with language, since it was discovered that lesions in this region caused receptive aphasia [47]. Patients with this condition tend not have
difficulty with grammar or prosody, but produce meaningless, nonsense sentences and display difficulty comprehending words [47]. Moreover, Harpaz et al. [30] used Transcranial Magnetic Simulation (TMG) to suggest that Wernicke's area in the non-dominant hemisphere is used to resolve lexical ambiguities.

One major issue with any research in Wernicke's area is that its anatomical boundaries, and to some degree its functionality, are highly ambiguous. Wise et al. [74] identify three distinct regions with unique response patterns. One region, directed along the supratemporal cortical plane, responded to both non-linguistic vocalization and speech sounds. The most posterior and medial region was linked to speech production. The third region, in the posterior left superior temporal sulcus, responded to external sources of speech and recall of lists of words in verbal fluency tasks. It is not safe to conclude that Wernicke's is associated primarily with lexical access. Rather, it displays a much wider range of functionalities which may include lexical access.

1.2.4.3 Inferior Frontal Gyrus (Broca's Area)

Broca's area is located near the inferior frontal gyrus and Brodmann Areas 44 and 45, although it is largely localized using activation patterns rather than morphological landmarks [27]. Traditionally, it has been associated with syntax, grammar, computation, and other hierarchical processing tasks [47], an assumption that arises from the observation that patients with lesions in this region often have difficulty producing or comprehending syntactically complex sentences. Although Broca's area has also been associated with working memory, neuroimaging studies by Grewe et al. [26] support the traditional view that this region is associated with hierarchical language processing.

1.3 ASD and Speech

Speech is highly distributed across the brain [47], and many of the symptoms of ASD are particular to speech. Therefore, it is reasonable to hypothesize that interconnectivity among different brain regions, discussed in the Introduction (1) may account for speech-specific symptoms of ASD. According to the American Psychological Association's definition of autism [5], symptoms include (1) delay in development of spoken language, or lack thereof, (2) marked difficulty in sustaining communication for children with speech, (3) stereotyped and repetitive use of language, and (4) lack of make-believe or social imitative play appropriate to developmental level. Moreover, there is evidence that there are significantly different
responses to speech in certain brain regions, and that these differences involve the auditory pathway’s dual strategy of representing a signal temporally and spectrally.

1.3.1 ASD Symptoms Pertaining to Language Development

Between one-third to one-half of all ASD individuals do not develop adequate language abilities to conduct basic communication, and those who do often demonstrate highly idiosyncratic use of speech and language [45]. These idiosyncracies can be classified into three categories, both related to socially salient areas of linguistic development - (1) phonological, (2) semantic/pragmatic, and (3) prosodic.

1.3.1.1 Phonological

Although many high-functioning ASD individuals show an ability to learn the phonetics of their native language, the learning process exhibits significant differences. ASD children show a delayed onset of babbling which is followed at ages 2-3 with a reduced frequency of other communicative forms such as consonants, words, word combinations, and gestures [39]. It is significant that these consonants and words depend heavily on temporal information.

1.3.1.2 Semantic/Pragmatic

At every stage of linguistic development, verbal ASD children display difficulties in learning to behave in terms of their role in the conversation. They tend not to coordinate their gestures with their sentences, or arbitrarily associate particular gestures with particular words. When they acquire rudimentary language skills, they often repeat what is being said to them (echolalia), or reverse their pronouns [39]. At around 8-10 years of age, many high-functioning ASD individuals display abnormally large vocabularies, although they have difficulty contextualizing words in a discourse [39]. In later stages of life, these difficulties manifest themselves as difficulties with non-verbal communication, interpreting inference, and reading the emotional tone of a conversation [48].

1.3.1.3 Prosodic Development

Prosody covers aspects of speech which mostly involve pitch and amplitude modulation. Pitch contour in language can communicate many different types of information. In the case of tone languages and pitch-accent languages such as Mandarin or Swedish, it identifies the word itself. Different languages often have set pitch patterns which can define the major portions of a sentence. For example, Farsi defines different pitch contours for the subject and predicate of a declarative sentence,
and a rising intonation in English indicates a question. Pitch contours can be used to emphasize information in the sentence, such as in “Mary washed the car” versus “Mary washed the car”. Finally, pitch contours can communicate non-verbal information such as sarcasm, irony, or enthusiasm.

One linguistic feature of ASD is abnormal prosody, which includes monotonic or machine-like intonation, repetitive intonations which are unrelated to the content or context of a sentence, difficulties in use of pitch and control of volume, deficiencies in vocal quality, and aberrant stress patterns [46]. It is difficult to classify these prosodic differences in terms of one feature, since there are many parameters along which a prosodic pattern can vary. For example, some autistic children have been shown to have a larger-than-normal or smaller-than-normal pitch range in their speech, but place stress and intone their sentences normally, whereas others may have a normal pitch range but aberrant stress patterns [66].

Difficulties with prosodic development are apparent even in the babbling stage, when ASD children tend not to synchronize their vocalization with those around them [46]. Le Normand et al. [46] collected spontaneous speech samples of autistic children and noticed a particular difficulty with declarative and interrogative sentences. This being said, prosodic development does not seem entirely unrelated to their environment. Paccia and Curcio [49] have conducted several behavioral studies which demonstrated that many ASD children, in echolalia, sometimes invert or systematically transform the pitch contour of the speaker, suggesting that they have some capability of prosodic parsing.

These results serve as a caveat to any attempt to pinpoint ASD issues with prosody and auditory processing on any single brain region. ASD children’s difficulties in learning temporally complex speech could indicate some fundamental difference in the mechanisms of temporal auditory processing, or it could indicate that the difference may arise from how to interpret and meaningfully react to the speech, which would implicate higher-order cognitive processing.

1.3.2 ASD and Auditory Processing

Despite the above-mentioned caveats, it is still useful to relate ASD prosodic patterns and processing pitch and other temporally complex features of a stimulus. Interestingly, ASD differences in prosody do not seem to implicate directly the mechanisms of pitch-processing itself. Bonnel et al. [9] showed that ASD individuals have superior performance in pitch discrimination tasks. However, there
is evidence that ASD individuals seem to have difficulty with stimuli with a high amount of spectral and temporal complexity. Samson et al. [63] showed that ASD individuals demonstrate lower event-related brain potentials (ERP) and poorer performance on behavioral tasks with temporally and spectrally complex stimuli, such as speech, than with simple stimuli such as pure tones or FM tones. However, it is not safe to assume that ASD subjects process simple sounds in the same way that typically developing (TD) individuals do. Using MEG, Roberts et al. [60] showed that ASD subjects have a delayed M100 response to pure tone stimuli. Since the M100 response arises from coordinated activity in the auditory cortices, a lower response to the same stimulus in an ASD subject suggest that the ability to coordinate these mechanisms differs significantly from TD subjects.

It is in changes in more temporally complex stimuli that ASD individuals may display the most difficulty. Ceponiene et al. [11] show that ASD subjects show ERP-related responses to pure tones and FM tones that are comparable to those of TD’s, but decreased responses to vowel transitions. This may have a similar neuropathology to the above-mentioned ASD difficulty with learning consonants, which have higher temporal variation than vowels, as discussed in Section 1.1.2.3.

1.3.3 Neural Underpinnings of ASD Sound Processing

It is important to reconcile behavioral observations regarding ASD speech with observations of morphological and neurological differences in brain regions.

1.3.3.1 Morphological Differences

Courchesne et al. [12] and Adrien et al. [2] point out that children with ASD have increased brain volume compared to TD children and an increase in white matter, which may arise from delayed and prolonged development of myelination during childhood. An excess or deficiency of myelination could have significant effects on the connectivity of different brain regions. Many of the brain regions implicated in ASD are responsible for connectivity between other brain regions. The corpus callosum, which connects the left and right hemispheres of the brain, is the region with the highest concentration of white matter, and this region has been found to be underdeveloped in some ASD individuals [29, 16]. Other structures implicated in ASD include the amygdala, arcuate fasiculus, amygdala, the hippocampus, cerebral cortex, and basal ganglia [3].

It is unclear whether there is a single neuropathology to ASD or whether it arises from a confluence of pathologies. It is also unclear whether structural differences
are similar among all ASD individuals. For example, the widely known ASD Kim Peek had no corpus callosum, but he is known for his prodigious memory and ability to perform difficult calculations.

1.3.3.2 Neuroimaging ASD Responses to Sound

fMRI studies have shown that there are statistically significant differences in how various brain regions respond to sensory information. Gomot et al. [24] scanned ASD and TD subjects while they performed a novelty detection test for auditory stimuli. They found that ASD subjects performed better on this task, and had more activation in their premotor-prefrontal and left inferior parietal regions.

1.3.3.3 Neuroimaging ASD Responses to Speech

Particularly relevant are neuroimaging studies of ASD subjects in response to speech stimuli. Boddaert et al. [8] scanned ASD and TD subjects as they were presented with temporally complex speech-like stimuli in which a synthesized vowel continually changed. They found that both ASD and TD subjects responded dominantly in left AC, but this response was muted, with statistical significance, among ASD subjects. If this observation is accurate, then it may also have consequences in higher regions of the auditory pathway. Anderson et al. [4] scanned TD and ASD subjects while presenting them with speech stimuli. Whereas the insula of TD subjects responded to speech predominantly on the left, there was significantly less activity in the left insula for ASD subjects. Moreover, Lange et al. [42] and Lange et al. [43] showed different activation patterns in TD subjects in the insula and other brain regions when exposed to speech and song stimuli. Whereas the insula in TD subjects tended to respond on the left to speech and bilaterally to song, ASD subjects tended to respond on the left to song and bilaterally to speech. Even on the higher levels of auditory processing, differences have been found. Just et al. [36] presented ASD subjects with speech stimuli and found that they had significantly lower activation in left Broca’s area, and higher activation in Wernicke’s area. Moreover, these authors found that the response timings were different than in TD subjects, suggesting abnormalities in synchronization between different brain regions.

1.4 Specific Aims

The above-mentioned evidence suggests that ASD subjects have different strategies of auditory processing compared to TD subjects, and that there is an even greater difference with speech. However, all of the above-mentioned studies used
either pure tone or long speech stimuli with high temporal variation. My advisor Dr. Lange designed and conducted the fMRI study. I recorded the auditory stimuli and processed them to produce the sounds that Dr. Lange's subjects heard, framed the Specific Aims of this thesis, and conducted the following independent analysis of the study data under Dr. Lange's guidance. My thesis investigates how different source-filter parameters in extremely short vowel stimuli affect responses of various brain regions in TD and ASD subjects. Statistical interactions between source-filter parameters and subject groups, if any, are crucial. The study has three specific aims and hypotheses, each tied to a particular interaction of interest:

**Aim 1:** To determine if there are significant differences in the processing of non-verbal information between ASD and TD subjects, as determined by AM, autocorrelation, FM, and pitch modulation.

**Hypothesis 1:** Because ASD subjects have demonstrated difficulty with temporally complex stimuli, stimuli which have the most FM, AM, and pitch modulation will also correlate with the responses most unique to ASD subjects.

**Aim 2:** To determine if there are significant differences between ASD and TD subjects in brain responses to stochastic events in speech, such as those resulting from transient physical events in the glottis. Because such events are highly transient and difficult to quantify. This can be done by comparing responses to a recorded human source with responses to a synthesized glottal source that has a waveform which is deterministically modeled, and therefore does not have the more nuanced features of a real human voice.

**Hypothesis 2:** Because of apparent ASD difficulties with transient components in complex stimuli, the TD and ASD groups respond differently, on average, to deterministically synthesized glottal sources vs. human glottal sources.

**Aim 3:** To determine if there are brain response differences in phonemic information processing between the groups, as determined by the parametric configuration of the vocal tract filter.

**Hypothesis 3:** There are no significant brain response differences in the processing of phonemic information between the two groups.
Chapter 2

Methods

In this study TD and ASD subjects were presented with speech stimuli while being scanned using fMRI.

2.1 Subjects and Scanning

All subjects had healthy hearing and language proficiency. There were $N = 5$ ASD subjects and $N = 4$ TD subjects. All ASD subjects were clinically tested for autism according to the World Health Organization’s International Classification of Diseases criteria [48], under the direction of a board-certified psychiatrist. Verbal IQ was assessed with the Weschler Abbreviated Scale of Intelligence, and language abilities were assessed with the Clinical Evaluation of Language Fundamentals. Whole-brain images were acquired on a 3-Tesla Magneton Trio (Siemens, Erlangen, Germany) scanner. Blood oxygen level-dependent (BOLD) echo-planar images (TR=2 s, TE=2.8 ms) were taken. The values are taken at the peak of the hemodynamic response function 6 seconds after the presentation of the stimulus. Statistical analysis was performed in the R programming language (version 2.11.1).

2.2 Brain Regions

The brain regions in question are known for their role in both speech and auditory processing. They were organized into three groups, according to their known role in speech processing.

**Group 1.** These regions are left and right Inferior Colliculus (IC) and the left thalamus. This group involves low-level auditory processing. This is characterized
mainly by sub-cortical location, and by the fact that the auditory processing does not significantly involve higher-order cognitive processes.

**Group 2.** These are left and right AC and left PI. This group consists of regions where higher-order cognitive processes are known to play a role, but where some semblance of purely acoustic processing is still preserved.

**Group 3.** These are left and right Supratemporal Gyrus, or Wernicke’s area, and left and right Posterior and Inferior Frontal Gyrus, or Broca’s area. This group of regions are understood to focus predominantly on semantic processing of language.

### 2.3 Auditory Stimuli

The stimuli varied systematically along several source and filter parameters. The subject pressed a button if they thought the stimulus was recorded and not synthesized. The button-press also kept the listener attentive to the stimulus and thus ensured the involvement of higher-order attentive input in the scanned brain regions. It was decided that stimuli of 200 ms should be used. This time was chosen partially for scanning purposes, but also because it is on the same order of magnitude as the 100 ms input-output time constant for the central nervous system [23]. A stimulus which was any shorter may not be perceivable as speech, but a stimulus which was longer might commence a complex series of neural feedback loops which would complicate the response. 200 ms allowed the experiment to target the basic signal processing mechanisms in the regions of interest (ROI’s). The stimuli were presented in random temporal order, each preceded by short tone bursts to focus the listener. The whole brain of each subject was scanned while the stimulus was played.

It was necessary to ensure that the speech stimuli had source characteristics that varied according to both source and filter. Each stimulus was a cardinal vowel of English, representing a configuration of the vocal tract filter - /a/, /e/, /i/, /o/, /u/. Each stimulus had a pitch contour that was meant to convey easily recognizable non-verbal information conveyed by the five emotions sad, happy, disgust, fear, and anger. In addition, each stimulus was either a recording of a vowel or a synthesized vowel. Each recorded sound was low-pass filtered at 2600 Hz. Low-pass filtering eliminated high-frequency components as a cue to the recorded sounds, and to ensure that the recordings only contained the vocal tract resonances needed to identify pitch, periodicity, and the vowel. For synthesized vowels, the vocal tract filter and pitch contour were extracted from a recorded vowel using Praat (version
5.10.3), and entered into a Klatt synthesizer [38], also implemented in Praat. A Klatt synthesizer performs a time-domain convolution (Equation 1.14) of a deterministic sequence of glottal pulses with a series of cascaded digital resonators, whose frequency-domain transfer functions are described in Equation 1.17. Only the first three vocal tract resonances were synthesized. A deterministically synthesized pulse train with the pitch contour was then convolved with the filter. This pulse train lacked many of the more stochastic elements of a human voice. The high-frequency cutoff for the synthesized stimuli was approximately 3200 Hz, although it varied from stimulus to stimulus.

Each stimulus can therefore be described in terms of several discrete and continuous parameters. “Factorized” parameters are integer parameters which the R programming language regards as separate, binary values. These values are:

1. **group** is a binary variable describing the subject, 1 for ASD and 0 for TD.

2. **emotion** is a factorized index variable to describe one of the five emotions. Because of its moderate amount of pitch variation, **anger** was chosen as the base emotion for the factorization.

3. **trial** is a parameter indicating the order of the trials per subject.

4. **stimulusIteration** is an integer referring to whether the stimulus is the first, second, or third iteration.

5. **vowel** is a factorized index variable to describe one of the five vowels. Because it contained the most evenly distributed formants, **a** was chosen as the baseline.

6. **origin** is a binary variable indicating whether the sound was human or synthesized - 0 for human and 1 for synthesized.

7. In one portion of the experiment, statistical analysis was run on an ROI in both hemispheres. In this particular case, a binary variable **hemi** is introduced, which is 1 if the region in question is left-lateral and 0 if it is right-lateral.

Alongside these discrete parameters are continuous scalars which describe the waveform of the stimulus.

1. **pitchvar** is the variance of 16 pitch points in the stimulus, measured by Wavesurfer (version 1.8.5).
2. autocorr is derived from an autocorrelation function which is the discrete counterpart to Equation 1.1:

\[
\tau(t, \tau) = \sum_{n=1}^{n=N} x(t - \tau)x(t)
\]  

(2.1)

where \( \tau \) is the number of lags, measured in samples of the .wav waveform, which had a sampling rate of 16000 Hz. This parameter is the average of this function applied with \( \tau = 1, 2...20 \). The purpose of autocorr is to measure periodicity in the signal. However, this periodicity will not focus on known periodicities such as the glottal pulse. 1-20 lags correspond with 1.25 ms, which would be sensitive to periodicities at 800-16000 Hz. This pitch range excludes the fundamental frequency and lower harmonics (which are accounted for by pitchvar), but makes the metric very sensitive to higher harmonics, speech-related noise (beginning at around 1500 Hz), and click-like transients. A steady-state sum of waves would have yield a high value for this variable, while a more transient signal would have a lower value. As shown in Tables 2.7 and 2.8, the synthesized sounds have significantly higher autocorrelation values than the recorded sounds, which are subject to stochastic processes characteristic of recorded speech.

3. am, standing for AM, is a variable that measures the variance of the envelope of the signal. To derive this, a Hilbert transform was applied to a time-domain waveform, and its absolute values were low-pass filtered at 200 Hz. The variance was then taken from this value. It is important to note that this is not a measure of “loudness”, since all signals had comparable amplitudes of around -10 dB in the .wav files. It is intended to describe the degree to which AM-sensitive mechanisms in the brain are activated. The samples for all .wav files had a value of less than ±31681.

4. fs, for "frequency similarity", is the average cosine similarity between consecutive spectral slices of the signal. To obtain this value, an FFT power spectrum, Hamming-windowed at 30 ms with 10% overlap, was taken from the waveform. The cosine distance between two vectors \( v_1 \) and \( v_2 \) is a metric

\[
cos(\theta) = \frac{v_1 \cdot v_2}{|v_1||v_2|},
\]

which is 1 when each element of the vector is equal, and 0 when the vectors are orthogonal. When applied to consecutive spectra in the signal, this metric allows one to quantify the degree of frequency change within the signal. The average of these values is capable of quantifying the stationarity in stationarity signal. This value therefore is a unit-less metric between 1 and 0.
Tables 2.1, 2.2, 2.3, 2.4, 2.7, 2.8, 2.5, and 2.6 show these values for each vowel and emotion, along with their averages for each vowel and emotion.

**Biostatistical Analysis.** The results were analyzed using the linear mixed effects model [40, 41]:

\[ y_i = X_i \beta + Z b_i + \epsilon_i \]  

(2.2)

where \( i = 1, 2, k \). All p-values are not corrected for multiple comparisons. There are \( r \) ROI’s, \( T \) trials. \( y_i \) is the \( M \times 1 \) vector of BOLD responses at each ROI for each trial for the \( i \)-th subject, where \( M = r \times T \). These responses were originally expressed as percentages of the mean BOLD response for the whole brain, although for interpretive purposes 100 was subtracted from the values. As a result, values above the mean were positive and values below were negative. \( X_i \) is an \( M \times 11 \) matrix whose columns represent experimental design features describing the stimuli and the subject and whose rows represent the \( T \) trials for each ROI. \( \beta \) is a \( 8 \times 1 \) vector of fixed effects regression coefficients to be estimated. \( Z \) is an \( M \times 2 \) matrix of ones and the times of each trial. \( b_i \) is a vector of length 2 of random effects, being a random intercept and a random slope for each subject. This includes a random intercept and a random slope for the trial variable. \( \epsilon \) is an \( M \times 1 \) vector of residual errors whose mean square the fitting algorithm minimizes. The columns of \( X_i \) will correspond to the above-described variables. Moreover, all interactions among these variables were also explored.

Every model included group and trial as a fixed effect. Then, models were trained on an exhaustive list of combinations of variables. Each combination included 1-4 terms, and the terms could either be independent variables or statistical interactions of two variables. Each model included an independent group term. The Maximum Likelihood criterion was used. Then, the Akaike Information Criterion (AIC) was computed. The model with the lowest AIC value was selected as the simplest and best possible model, and the variance components were then re-fitted by Restricted Maximum Likelihood (REML) [31, 32].

41
Table 2.1: AM measurements for recorded emotions and vowels. Averages for emotions are listed on the rightmost column, while averages for vowels are listed on the bottom row.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>e</th>
<th>i</th>
<th>o</th>
<th>u</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>0.008</td>
<td>0.012</td>
<td>0.007</td>
<td>0.007</td>
<td>0.010</td>
<td>0.012</td>
</tr>
<tr>
<td>Disgust</td>
<td>0.008</td>
<td>0.007</td>
<td>0.010</td>
<td>0.014</td>
<td>0.010</td>
<td>0.0115</td>
</tr>
<tr>
<td>Fear</td>
<td>0.021</td>
<td>0.006</td>
<td>0.010</td>
<td>0.016</td>
<td>0.013</td>
<td>0.0131</td>
</tr>
<tr>
<td>Happy</td>
<td>0.013</td>
<td>0.012</td>
<td>0.011</td>
<td>0.014</td>
<td>0.006</td>
<td>0.0125</td>
</tr>
<tr>
<td>Sad</td>
<td>0.012</td>
<td>0.023</td>
<td>0.008</td>
<td>0.015</td>
<td>0.015</td>
<td>0.0137</td>
</tr>
<tr>
<td>average</td>
<td>0.012</td>
<td>0.012</td>
<td>0.009</td>
<td>0.013</td>
<td>0.011</td>
<td>0.012</td>
</tr>
</tbody>
</table>

Table 2.2: AM measurements for synthesized emotions and vowels. Averages for emotions are listed on the rightmost column, while averages for vowels are listed on the bottom row.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>e</th>
<th>i</th>
<th>o</th>
<th>u</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>0.013</td>
<td>0.021</td>
<td>0.012</td>
<td>0.005</td>
<td>0.024</td>
<td>0.0160</td>
</tr>
<tr>
<td>Disgust</td>
<td>0.014</td>
<td>0.022</td>
<td>0.020</td>
<td>0.026</td>
<td>0.023</td>
<td>0.0162</td>
</tr>
<tr>
<td>Fear</td>
<td>0.018</td>
<td>0.008</td>
<td>0.004</td>
<td>0.024</td>
<td>0.005</td>
<td>0.0134</td>
</tr>
<tr>
<td>Happy</td>
<td>0.007</td>
<td>0.027</td>
<td>0.021</td>
<td>0.020</td>
<td>0.004</td>
<td>0.0146</td>
</tr>
<tr>
<td>Sad</td>
<td>0.007</td>
<td>0.022</td>
<td>0.010</td>
<td>0.012</td>
<td>0.013</td>
<td>0.0140</td>
</tr>
<tr>
<td>average</td>
<td>0.0121</td>
<td>0.0200</td>
<td>0.0133</td>
<td>0.0174</td>
<td>0.0136</td>
<td>0.0148</td>
</tr>
</tbody>
</table>

Table 2.3: FS measurements for recorded emotions and vowels. Averages for emotions are listed on the rightmost column, while averages for vowels are listed on the bottom row.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>e</th>
<th>i</th>
<th>o</th>
<th>u</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>0.94</td>
<td>0.95</td>
<td>0.94</td>
<td>0.91</td>
<td>0.97</td>
<td>0.9361</td>
</tr>
<tr>
<td>Disgust</td>
<td>0.87</td>
<td>0.89</td>
<td>0.95</td>
<td>0.97</td>
<td>0.96</td>
<td>0.9388</td>
</tr>
<tr>
<td>Fear</td>
<td>0.94</td>
<td>0.92</td>
<td>0.94</td>
<td>0.96</td>
<td>0.93</td>
<td>0.9330</td>
</tr>
<tr>
<td>Happy</td>
<td>0.94</td>
<td>0.96</td>
<td>0.97</td>
<td>0.97</td>
<td>0.91</td>
<td>0.9312</td>
</tr>
<tr>
<td>Sad</td>
<td>0.91</td>
<td>0.92</td>
<td>0.9</td>
<td>0.93</td>
<td>0.9</td>
<td>0.9337</td>
</tr>
<tr>
<td>average</td>
<td>0.920</td>
<td>0.927</td>
<td>0.942</td>
<td>0.944</td>
<td>0.934</td>
<td>0.935</td>
</tr>
</tbody>
</table>
Table 2.4: FS measurements for synthesized emotions and vowels. Averages for emotions are listed on the rightmost column, while averages for vowels are listed on the bottom row.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>e</th>
<th>i</th>
<th>o</th>
<th>u</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>0.95</td>
<td>0.96</td>
<td>0.98</td>
<td>0.93</td>
<td>0.96</td>
<td>0.9424</td>
</tr>
<tr>
<td>Disgust</td>
<td>0.94</td>
<td>0.95</td>
<td>0.95</td>
<td>0.96</td>
<td>0.96</td>
<td>0.9399</td>
</tr>
<tr>
<td>Fear</td>
<td>0.93</td>
<td>0.91</td>
<td>0.91</td>
<td>0.93</td>
<td>0.9</td>
<td>0.9390</td>
</tr>
<tr>
<td>Happy</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.96</td>
<td>0.94</td>
<td>0.9466</td>
</tr>
<tr>
<td>Sad</td>
<td>0.95</td>
<td>0.94</td>
<td>0.96</td>
<td>0.96</td>
<td>0.94</td>
<td>0.9449</td>
</tr>
<tr>
<td>average</td>
<td>0.9315</td>
<td>0.9342</td>
<td>0.9453</td>
<td>0.9468</td>
<td>0.9376</td>
<td>0.9382</td>
</tr>
</tbody>
</table>

Table 2.5: Pitch variation measurements for recorded emotions and vowels. Averages for emotions are listed on the rightmost column, while averages for vowels are listed on the bottom row.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>e</th>
<th>I</th>
<th>o</th>
<th>u</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>526.34</td>
<td>299.78</td>
<td>92.52</td>
<td>333.3</td>
<td>57.51</td>
<td>415.8540</td>
</tr>
<tr>
<td>Disgust</td>
<td>1356.42</td>
<td>475.94</td>
<td>651.55</td>
<td>1036.82</td>
<td>121.78</td>
<td>509.6660</td>
</tr>
<tr>
<td>Fear</td>
<td>290.6</td>
<td>446.94</td>
<td>28.05</td>
<td>516.47</td>
<td>3.79</td>
<td>397.2667</td>
</tr>
<tr>
<td>Happy</td>
<td>1420.93</td>
<td>333.49</td>
<td>190.07</td>
<td>681.39</td>
<td>90.75</td>
<td>457.0228</td>
</tr>
<tr>
<td>Sad</td>
<td>462.6</td>
<td>544.2</td>
<td>413.15</td>
<td>507.71</td>
<td>28.86</td>
<td>363.2108</td>
</tr>
<tr>
<td>average</td>
<td>811.378</td>
<td>420.070</td>
<td>275.068</td>
<td>615.138</td>
<td>60.538</td>
<td>428.604</td>
</tr>
</tbody>
</table>

Table 2.6: Pitch variation measurements for synthesized emotions and vowels. Averages for emotions are listed on the rightmost column, while averages for vowels are listed on the bottom row.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>e</th>
<th>i</th>
<th>o</th>
<th>u</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>526.34</td>
<td>299.78</td>
<td>92.52</td>
<td>333.3</td>
<td>57.51</td>
<td>415.8540</td>
</tr>
<tr>
<td>Disgust</td>
<td>1356.42</td>
<td>475.94</td>
<td>651.55</td>
<td>1036.82</td>
<td>121.78</td>
<td>509.6660</td>
</tr>
<tr>
<td>Fear</td>
<td>290.6</td>
<td>446.94</td>
<td>28.05</td>
<td>516.47</td>
<td>3.79</td>
<td>397.2667</td>
</tr>
<tr>
<td>Happy</td>
<td>1420.93</td>
<td>333.49</td>
<td>190.07</td>
<td>681.39</td>
<td>90.75</td>
<td>457.0228</td>
</tr>
<tr>
<td>Sad</td>
<td>462.6</td>
<td>544.2</td>
<td>413.15</td>
<td>507.71</td>
<td>28.86</td>
<td>363.2108</td>
</tr>
<tr>
<td>average</td>
<td>811.378</td>
<td>420.070</td>
<td>275.068</td>
<td>615.138</td>
<td>60.538</td>
<td>433.2095</td>
</tr>
</tbody>
</table>
Table 2.7: Autocorrelation measurements for recorded emotions and vowels. Averages for emotions are listed on the rightmost column, while averages for vowels are listed on the bottom row.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>e</th>
<th>i</th>
<th>o</th>
<th>u</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>0.01</td>
<td>-0.08</td>
<td>0.05</td>
<td>0.03</td>
<td>-0.12</td>
<td>0.0540</td>
</tr>
<tr>
<td>Disgust</td>
<td>-0.04</td>
<td>-0.02</td>
<td>-0.13</td>
<td>-0.12</td>
<td>-0.15</td>
<td>0.0253</td>
</tr>
<tr>
<td>Fear</td>
<td>0.26</td>
<td>0.21</td>
<td>0.21</td>
<td>0.27</td>
<td>0.43</td>
<td>0.0607</td>
</tr>
<tr>
<td>Happy</td>
<td>-0.02</td>
<td>-0.14</td>
<td>-0.16</td>
<td>-0.15</td>
<td>-0.07</td>
<td>-0.0268</td>
</tr>
<tr>
<td>Sad</td>
<td>0.07</td>
<td>0.15</td>
<td>-0.03</td>
<td>-0.09</td>
<td>-0.03</td>
<td>0.1099</td>
</tr>
<tr>
<td>average</td>
<td>0.056</td>
<td>0.024</td>
<td>-0.012</td>
<td>-0.012</td>
<td>0.012</td>
<td>0.045</td>
</tr>
</tbody>
</table>

Table 2.8: Autocorrelation measurements for synthesized emotions and vowels. Averages for emotions are listed on the rightmost column, while averages for vowels are listed on the bottom row.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>e</th>
<th>i</th>
<th>o</th>
<th>u</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>0.13</td>
<td>0.42</td>
<td>0.3</td>
<td>0.13</td>
<td>0.53</td>
<td>0.5380</td>
</tr>
<tr>
<td>Disgust</td>
<td>0.56</td>
<td>0.49</td>
<td>0.55</td>
<td>0.56</td>
<td>0.56</td>
<td>0.6360</td>
</tr>
<tr>
<td>Fear</td>
<td>0.86</td>
<td>0.66</td>
<td>0.63</td>
<td>0.86</td>
<td>0.83</td>
<td>0.6513</td>
</tr>
<tr>
<td>Happy</td>
<td>0.56</td>
<td>0.65</td>
<td>0.61</td>
<td>0.56</td>
<td>0.6</td>
<td>0.5820</td>
</tr>
<tr>
<td>Sad</td>
<td>0.54</td>
<td>0.64</td>
<td>0.53</td>
<td>0.54</td>
<td>0.7</td>
<td>0.5750</td>
</tr>
<tr>
<td>average</td>
<td>0.530</td>
<td>0.572</td>
<td>0.524</td>
<td>0.530</td>
<td>0.644</td>
<td>0.2955</td>
</tr>
</tbody>
</table>
Chapter 3

Results

The following tables report fixed-effect coefficients with their null-hypothesis probabilities. In the tables, the "::" indicates an interaction between two parameters. Only parameters with statistical significance are taken into consideration. The parameters of the model and their standard error are reported as percentages above or below the fixed intercept. If $v$ is the value and $v_i$ is the value for the intercept, these values are written as $\frac{100v}{v_i}$. Where there are data from the ROI on both hemispheres of the brain, the models trained on the left, right, and combined hemispheres are reported. Parameter estimates for the model are presented as percentages of the intercept. A negative sign on the coefficient indicates that the value is subtracting from the response, while a positive sign indicates that it is adding to the fMRI response.

3.1 Group 1 (Inferior Colliculus, left Thalamus)

3.1.1 Inferior Colliculus

3.1.1.1 Left Hemisphere

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std.Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>2.4</td>
<td>127.11</td>
<td>-0.35</td>
<td>0.97</td>
</tr>
<tr>
<td>fs</td>
<td>-0.2</td>
<td>0.1</td>
<td>-1.85</td>
<td>0.005</td>
</tr>
<tr>
<td>stimit</td>
<td>-30.48</td>
<td>33.53</td>
<td>-4.25</td>
<td>0.09</td>
</tr>
<tr>
<td>trial</td>
<td>-7.27</td>
<td>3.63</td>
<td>-3.8</td>
<td>0.005</td>
</tr>
<tr>
<td>stimit:trial</td>
<td>0.12</td>
<td>0.03</td>
<td>8</td>
<td>0.005</td>
</tr>
</tbody>
</table>
There are no significant group effects for this region, where group effects are a statistically significant difference between ASD and TD subjects. However, there are significant effects for trial and stim. The negative values for these effects suggest that responses decrease with time. However, interestingly, the group interaction of these variables, trial:stim, suggests that, although responses decrease over time, there is an increase for consecutive stimulus iterations. As was mentioned above, IC is sensitive to cortical attentive processes [59], and a decrease in attentiveness could be the result of the stimulus losing its novelty over the trial.

### 3.1.1.2 Right Hemisphere

Table 3.2: Results for right IC

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std.Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>-38.49</td>
<td>125.73</td>
<td>-0.35</td>
<td>0.73</td>
</tr>
<tr>
<td>fs</td>
<td>-47.16</td>
<td>29.41</td>
<td>-1.85</td>
<td>0.07</td>
</tr>
<tr>
<td>stim</td>
<td>-11.76</td>
<td>3.18</td>
<td>-4.25</td>
<td>&lt; 0.005</td>
</tr>
<tr>
<td>trial</td>
<td>-0.28</td>
<td>0.08</td>
<td>-3.8</td>
<td>&lt; 0.005</td>
</tr>
<tr>
<td>stim:trial</td>
<td>0.19</td>
<td>0.03</td>
<td>8</td>
<td>&lt; 0.005</td>
</tr>
</tbody>
</table>

These results are similar results to left IC, suggesting that this ROI’s response to both trial and stim are similar.

### 3.1.1.3 Combined Hemispheres

Table 3.3: Results for combined IC

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std.Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>-7.26</td>
<td>132.24</td>
<td>-0.13</td>
<td>0.9</td>
</tr>
<tr>
<td>hemi</td>
<td>-1.41</td>
<td>3.23</td>
<td>-1</td>
<td>0.32</td>
</tr>
<tr>
<td>trial</td>
<td>-0.17</td>
<td>0.16</td>
<td>-2.39</td>
<td>0.02</td>
</tr>
<tr>
<td>stim</td>
<td>-6.6</td>
<td>7.39</td>
<td>-2.04</td>
<td>0.04</td>
</tr>
<tr>
<td>trial:stim</td>
<td>0.11</td>
<td>0.06</td>
<td>3.98</td>
<td>&lt; 0.005</td>
</tr>
</tbody>
</table>

The results for trial and stim are similar to the results for left and right IC. There are no significant effects for hemisphere.
Table 3.4: Results for left Thalamus

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std.Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>-249.76</td>
<td>102.92</td>
<td>-1.66</td>
<td>0.14</td>
</tr>
<tr>
<td>trial</td>
<td>-0.65</td>
<td>0.09</td>
<td>-4.87</td>
<td>&lt;0.005</td>
</tr>
<tr>
<td>stimit</td>
<td>-18.07</td>
<td>2.69</td>
<td>-4.59</td>
<td>0.005</td>
</tr>
<tr>
<td>fs</td>
<td>233.82</td>
<td>60.12</td>
<td>2.65</td>
<td>0.01</td>
</tr>
<tr>
<td>am</td>
<td>19150.69</td>
<td>4987.85</td>
<td>2.62</td>
<td>0.01</td>
</tr>
<tr>
<td>trial:stimit</td>
<td>0.3</td>
<td>0.02</td>
<td>9.04</td>
<td>&lt;0.005</td>
</tr>
<tr>
<td>fs:am</td>
<td>-20526.95</td>
<td>5285.3</td>
<td>-2.65</td>
<td>0.01</td>
</tr>
</tbody>
</table>

3.1.2 Left Thalamus

The results for trial, stimit, and trial:stimit in the left thalamus are similar to right and left IC. Here, am is significant, suggesting that more AM corresponds to more activity. The result for fs suggests that greater frequency similarity in the signal corresponds to greater activity. The interaction am:fs, however, suggests that a high similarity in frequency and a high am results in less activity.

The similarity of the results for trial and stimit to those of the same variables for IC suggests that this is the effect of a similar attentive process. Both the thalamus and IC are subject to higher attentive processes, and it would make sense that the effects of the same processes relay through the thalamus to IC. As has been discussed previously, the thalamus does not appear to contain FM and AM-sensitive neurons, and the results for am and fs, if they are significant, are most likely due to AM and FM processing mechanisms in IC. If this is true, then the lack of such results in IC may be an artifact of the scanner, being less sensitive to smaller structures.

3.2 Group 2(Auditory Cortex, left Posterior Insula)

3.2.1 Auditory Cortex

In AC, there were significant group differences which were even more pronounced in ASD subjects. Left AC was particularly sensitive to the stimulus parameter autocorr, and even moreso among ASD subjects.
Table 3.5: Results for left AC

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std.Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>-120.17</td>
<td>133.34</td>
<td>-1.13</td>
<td>0.29</td>
</tr>
<tr>
<td>trial</td>
<td>0.02</td>
<td>0.03</td>
<td>0.88</td>
<td>0.38</td>
</tr>
<tr>
<td>am</td>
<td>-187.35</td>
<td>90.27</td>
<td>-2.61</td>
<td>0.01</td>
</tr>
<tr>
<td>autocorr</td>
<td>-4.32</td>
<td>2.61</td>
<td>-2.08</td>
<td>0.04</td>
</tr>
<tr>
<td>group:autocorr</td>
<td>9.16</td>
<td>3.43</td>
<td>3.36</td>
<td>&lt; 0.005</td>
</tr>
</tbody>
</table>

3.2.1.1 Left Hemisphere

The *am* coefficient suggests that modulation in amplitude is correlated with causes significant decreases in activity in left AC, such that more change in amplitude causes less activity. The *autocorr* coefficient shows that increased autocorrelation in the signal is also correlated to a decrease in activity. However, this is counterbalanced by an interaction between *group* and *autocorr*, where increased autocorrelation is correlated with more activity in ASD subjects than TD subjects. However, it is important to note that the effect of *am* is two orders of magnitude greater than either *autocorr* and *group:autocorr*, suggesting that these latter effects are very subtle. Therefore, any conclusions drawn about these latter two variables must weigh the small difference and the small sample size against their statistical significance.

Two hypotheses arise for how a reduction in *am* may cause a decrease in activity in left AC. Because *am* is a variance which measures the change in amplitude in a signal, it measures the rapidity of a sound attenuation. As different frequency components of a sound attenuate, there could be a decrease in activity in their corresponding tonotopic regions. Alternately, it is possible that signals with high AM activate AM-sensitive cortical neurons, and the scanner is sensing their subsequent deactivation. There is no effect for *group* or is there any interaction between *group* and *am*, which supports the observations of Ceponiene et al. [11] that the basic mechanisms of AM processing in ASD left AC are similar to those in TD. This supports the view that this is the result of temporally precise AM-sensitive cortical neurons in left AC.

In the analysis of *autocorr*, it is again important to emphasize that the coefficients are an order of magnitude smaller than the other statistically significant coefficients. This small order of magnitude, combined with the small sample size of the subjects, makes any results concerning this variable in LAC particularly tentative. This being said, it is still the case that *autocorr* corresponds with less
activation in both TD and ASD subjects. This value is sensitive to both periodicity and transient events in the stimulus, with a more periodic signal yielding a higher value, and a more transient signal yielding a lower value. Since a periodic signal is more stationary than a transient signal, it may produce less activity in left AC. This could be why the autocorr coefficient is negative. However, the positive sign of the group:autocorr shows that the opposite effect seems to occur with ASD subjects. Ordinary speech contains many transient components, placing significant demands on our temporal processing abilities. In this hypothesis, when a more periodic signal enters left AC of an ASD subject, there is an increase in activity, since this is a sound that the ASD left AC can more effectively process. This activation may be due to some attentive process or a mechanism internal to left AC.

3.2.1.2 Right Hemisphere

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std.Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>-71.2</td>
<td>134.02</td>
<td>-1.24</td>
<td>0.25</td>
</tr>
<tr>
<td>trial</td>
<td>0.03</td>
<td>0.24</td>
<td>0.3</td>
<td>0.76</td>
</tr>
<tr>
<td>e</td>
<td>-3.26</td>
<td>4.06</td>
<td>-1.88</td>
<td>0.06</td>
</tr>
<tr>
<td>i</td>
<td>2.57</td>
<td>4.06</td>
<td>1.48</td>
<td>0.14</td>
</tr>
<tr>
<td>o</td>
<td>0.94</td>
<td>4.09</td>
<td>0.54</td>
<td>0.59</td>
</tr>
<tr>
<td>u</td>
<td>-0.42</td>
<td>4.08</td>
<td>-0.24</td>
<td>0.81</td>
</tr>
<tr>
<td>stim</td>
<td>-3.99</td>
<td>6.18</td>
<td>-1.51</td>
<td>0.13</td>
</tr>
<tr>
<td>group:stim</td>
<td>-2.64</td>
<td>8.22</td>
<td>-0.75</td>
<td>0.45</td>
</tr>
</tbody>
</table>

The only conclusions that can be drawn from the results in table 3.6 is that there is no similar response in right AC to temporal properties of the stimulus. This finding is consistent with with the specialization of right AC for spectral precision [76]. Moreover, the fact that am is not statistically significant suggests that the values for left AC corresponded not to the degree of tonotopic activation, but rather to AM-sensitive cortical neurons.

3.2.1.3 Combined Hemispheres

These results show that there are significant differences in activation of the brain in different hemispheres for TD subjects when presented with the speech stimuli. The hemi coefficient shows more activity in left AC, which is expected for temporally complex speech stimuli. The hemi:group coefficient, however, indicates an overall
Table 3.7: Results for combined AC

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std.Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>-67.37</td>
<td>133.07</td>
<td>-1.11</td>
<td>0.3</td>
</tr>
<tr>
<td>trial</td>
<td>0.05</td>
<td>0.3</td>
<td>0.34</td>
<td>0.74</td>
</tr>
<tr>
<td>hemi</td>
<td>216.03</td>
<td>13.65</td>
<td>34.81</td>
<td>&lt; 0.005</td>
</tr>
<tr>
<td>stimit</td>
<td>-3.67</td>
<td>15.23</td>
<td>-0.53</td>
<td>0.6</td>
</tr>
<tr>
<td>group:hemi</td>
<td>-64.47</td>
<td>18.31</td>
<td>-7.74</td>
<td>&gt; 0.005</td>
</tr>
</tbody>
</table>

decrease in activity for ASD subjects in left AC. This is consistent with the conclusions of Boddaert et al. [8] who noticed decreased activity in left AC in response to speech stimuli in ASD subjects.

3.2.2 Left Posterior Insula

Table 3.8: Results for left PI

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std.Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>50.46</td>
<td>106.42</td>
<td>0.85</td>
<td>0.42</td>
</tr>
<tr>
<td>trial</td>
<td>-0.17</td>
<td>0.07</td>
<td>-4.38</td>
<td>&lt; 0.005</td>
</tr>
<tr>
<td>stimit</td>
<td>-6.28</td>
<td>2.7</td>
<td>-4.16</td>
<td>&lt; 0.005</td>
</tr>
<tr>
<td>fs</td>
<td>35.99</td>
<td>60.43</td>
<td>1.07</td>
<td>0.29</td>
</tr>
<tr>
<td>am</td>
<td>5661.87</td>
<td>5013.61</td>
<td>2.02</td>
<td>0.04</td>
</tr>
<tr>
<td>trial:stimit</td>
<td>0.09</td>
<td>0.02</td>
<td>7.06</td>
<td>&lt; 0.005</td>
</tr>
<tr>
<td>fs:am</td>
<td>-6026.16</td>
<td>5312.59</td>
<td>-2.03</td>
<td>0.04</td>
</tr>
</tbody>
</table>

There are no significant group effects in this ROI. The pattern for trial, stimit, and trial:stimit follows that observed in the left and right IC, where trial and stimit were correlated with a decrease in activity, but trial:stimit was correlated with an offset in this decrease. Here, this offset is an order of magnitude smaller than the coefficients for either trial and stimit, suggesting it is a very subtle effect. There is a significant effect for am. However, the group interaction fs:am suggests that when high am corresponds with high fs, activity decreases. This interaction coefficient is on the same order of magnitude of the coefficient for am suggesting that the cancelling effect of a high fs value is almost total.

If the small sample size allows for any conclusions, the lack of a significant group effect in these results suggest that the results of Lange et al. [42] and Anderson et al. [4] do not apply for such short stimuli. The overall decrease in response
with trial and stimit suggests that all subjects seem to be decreasing in attention. However, the small offset with trial:stimit may be due to the subjects’ recognizing the stimulus. As stated above, the insula demonstrates certain unique responses to the human voice, and is widely believed to assign emotional valence to sensory stimuli. Therefore, recognition of the stimulus may have elicited a cognitive or emotional response in the subjects.

As has been discussed in Section 1.2.3.2, speech has AM and FM components. Therefore, knowing the above-mentioned sensitivity to speech in this ROI, the results for am suggest that the ROI’s sensitivity to this variable may be speech-specific. The fs:am interaction suggests that, although am may be activating a speech-like response in the ROI, a signal with low frequency modulation. That is, a signal that is less speech-like can cancel this activation. The similar order of magnitude of the am and fs:am coefficients suggests that both high AM and high FM are prerequisites for left PI to respond to a speech-like stimuli. The absence of group differences in these results contrasts to the decreased activity in Lange et al. [42] and Anderson et al. [4] depends on a longer stimulus.

3.3 Group 3 (Wernicke’s area, Broca’s area)

3.3.1 Wernicke’s area

3.3.1.1 Left Hemisphere

<table>
<thead>
<tr>
<th>Table 3.9: Results for left Wernicke’s area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
</tr>
<tr>
<td>group</td>
</tr>
<tr>
<td>trial</td>
</tr>
<tr>
<td>disgust</td>
</tr>
<tr>
<td>fear</td>
</tr>
<tr>
<td>happy</td>
</tr>
<tr>
<td>sad</td>
</tr>
<tr>
<td>am</td>
</tr>
<tr>
<td>stimit</td>
</tr>
<tr>
<td>group:stimit</td>
</tr>
</tbody>
</table>

The near-significant value for group \( (p < 0.06) \) suggests lower activation for ASD subjects than TD subjects. The disgust stimulus is correlated with a significantly higher response. A higher response is also correlated with am. Moreover,
unlike in lower regions, trial is not statistically significant, although stimit is correlated with a higher response. Moreover, there is an interaction between group and stimit, where ASD subjects are correlated with a higher response.

The most striking feature of the result in Table 3.9 is that there is no significant interaction between trial and group. If a correlation of trial with the response suggests the effect of some attentive process, then these results suggest that right Wernicke’s area is not subject to this process. However, there does seem to be a decrease in activity in response to stimit, suggesting that subjects may devote more of this ROI’s processing power to the stimulus. However, the group:stimit interaction suggests that this effect is not as prominent in ASD subjects, and that the stimuli maintain their novelty. The coefficient for this effect is on the same order of magnitude as the coefficient for stimit, suggesting that ASD subjects are not affected by the effects of stimit.

The correlation of am to a higher response suggests some sensitivity in left Wernicke’s area to stimulus-specific parameters. Although there is no interaction, Tables 2.2 shows that the disgust emotion has one of the highest AM values for the synthesized stimuli. Perhaps these results’ sensitivity to this emotion reflect sensitivity to AM. This places a significant caveat to the conclusion that subjects are in some way recognizing the disgust emotion as a lexical item - a meaningful symbolic unit of sound, which in the case of language is a word. As was stated above, Wernicke’s area is not only sensitive to lexical access tasks, but to vocalizations as well [74].

3.3.1.2 Right Hemisphere

In Table 3.10, trial and stimit show a similar interaction that they show in lower regions. fs is correlated with greater activity in this region. Moreover, the emotions disgust, happy, and sad are correlated with greater activity. However, lower fs values in the region are also correlate with lower responses when these emotions are presented.

The significance of fs could be sensitivity to transient vocalization mentioned by Wise [74]. However, it may also relate to the lack of significance for the fear emotion. Table 2.4 shows that this has the lowest average fs for any of the synthesized emotions. However, this does not entail that low fs signifies low response to the emotion. The interactions between fs and the emotions suggest that higher fs corresponds with less activity correlated with the emotion.
Table 3.10: Results for right Wernicke’s area

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std.Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>-84.2</td>
<td>118.86</td>
<td>-1.73</td>
<td>0.13</td>
</tr>
<tr>
<td>trial</td>
<td>-0.08</td>
<td>0.09</td>
<td>-2.1</td>
<td>0.04</td>
</tr>
<tr>
<td>disgust</td>
<td>83.59</td>
<td>66.9</td>
<td>3.05</td>
<td>&lt; 0.005</td>
</tr>
<tr>
<td>fear</td>
<td>22.29</td>
<td>65.66</td>
<td>0.83</td>
<td>0.41</td>
</tr>
<tr>
<td>happy</td>
<td>68.51</td>
<td>71.91</td>
<td>2.32</td>
<td>0.02</td>
</tr>
<tr>
<td>sad</td>
<td>74.03</td>
<td>63.49</td>
<td>2.84</td>
<td>&lt; 0.005</td>
</tr>
<tr>
<td>fs</td>
<td>50.6</td>
<td>49.56</td>
<td>2.49</td>
<td>0.01</td>
</tr>
<tr>
<td>stimit</td>
<td>-5.47</td>
<td>2.38</td>
<td>-5.61</td>
<td>&lt; 0.005</td>
</tr>
<tr>
<td>disgust:fs</td>
<td>-88.59</td>
<td>71.66</td>
<td>-3.01</td>
<td>&lt; 0.005</td>
</tr>
<tr>
<td>fear:fs</td>
<td>-24.53</td>
<td>69.96</td>
<td>-0.85</td>
<td>0.39</td>
</tr>
<tr>
<td>happy:fs</td>
<td>-73.08</td>
<td>76.39</td>
<td>-2.33</td>
<td>0.02</td>
</tr>
<tr>
<td>sad:fs</td>
<td>-78.9</td>
<td>67.93</td>
<td>-2.83</td>
<td>&lt; 0.005</td>
</tr>
<tr>
<td>trial:stimit</td>
<td>0.06</td>
<td>0.02</td>
<td>6.91</td>
<td>&lt; 0.005</td>
</tr>
</tbody>
</table>

3.3.1.3 Combined Hemispheres

Table 3.11: Results for combined Wernicke’s area

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std.Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>-55.58</td>
<td>132.91</td>
<td>-2.17</td>
<td>0.07</td>
</tr>
<tr>
<td>trial</td>
<td>0.01</td>
<td>0.31</td>
<td>0.22</td>
<td>0.83</td>
</tr>
<tr>
<td>stimit</td>
<td>-0.67</td>
<td>16.18</td>
<td>-0.22</td>
<td>0.83</td>
</tr>
<tr>
<td>hemi</td>
<td>-176.44</td>
<td>14.5</td>
<td>-63.05</td>
<td>&lt; 0.005</td>
</tr>
<tr>
<td>group:hemi</td>
<td>-39.63</td>
<td>19.45</td>
<td>-10.56</td>
<td>&lt; 0.005</td>
</tr>
</tbody>
</table>

hemi suggests that there is less activity in left Wernicke’s area than in the right. group also suggests that ASD subjects display lower activity in both hemispheres. Moreover, group:hemi suggests that there is even less left-lateralized activity than in TD subjects.

The results in Table 3.11 compare interestingly with those of left AC, which suggest that there is greater left-lateral activity in this region. This left-lateralization, however, may have to do with AC’s closer association with specialized sensory processing mechanisms in left AC.
3.3.2 Broca’s area

3.3.2.1 Left Hemisphere

The results in Table 3.12 demonstrate the opposite effects as the result for the combined Wernicke’s area. Here, hemi indicates increased activity in left
Table 3.13: Results for right Broca’s area

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std.Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>189.69</td>
<td>130.67</td>
<td>1.13</td>
<td>0.3</td>
</tr>
<tr>
<td>trial</td>
<td>-0.45</td>
<td>0.12</td>
<td>-2.92</td>
<td>&lt; 0.005</td>
</tr>
<tr>
<td>stimit</td>
<td>-23.22</td>
<td>3.98</td>
<td>-4.54</td>
<td>&lt; 0.005</td>
</tr>
<tr>
<td>pitchvar</td>
<td>-0.01</td>
<td>0</td>
<td>-2.06</td>
<td>0.04</td>
</tr>
<tr>
<td>trial:stimit</td>
<td>0.16</td>
<td>0.03</td>
<td>3.69</td>
<td>&lt; 0.005</td>
</tr>
<tr>
<td>group:pitchvar</td>
<td>0.01</td>
<td>0</td>
<td>1.78</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Table 3.14: Results for combined Broca’s area

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std.Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>-6.94</td>
<td>132.47</td>
<td>-0.08</td>
<td>0.94</td>
</tr>
<tr>
<td>hemi</td>
<td>642.38</td>
<td>8.99</td>
<td>110.92</td>
<td>&gt; 0.005</td>
</tr>
<tr>
<td>stimit</td>
<td>-3.79</td>
<td>10.03</td>
<td>-0.59</td>
<td>0.56</td>
</tr>
<tr>
<td>trial</td>
<td>-0.09</td>
<td>0.21</td>
<td>-0.66</td>
<td>0.51</td>
</tr>
<tr>
<td>group:hemi</td>
<td>121.42</td>
<td>12.06</td>
<td>15.63</td>
<td>&gt; 0.005</td>
</tr>
</tbody>
</table>

Broca’s area, which contrasts with the decreased activity in left Wernicke’s area. Moreover, group:hemi increases this difference, suggesting that ASD subjects in response to speech may have more left-lateralization in Broca’s area and more right-lateralization in Wernicke’s area. The similar responses of left Broca’s and right Wernicke’s suggest that there is a pattern of interaction between these two regions, which is distinct from the pattern in left Wernicke’s area.
Chapter 4

Discussion

Interpreted together with the caveats mentioned in the preceding section, the results for all ROI's may provide evidence for possibly different sensory processing strategies of ASD and TD individuals. While it is encouraging to find statistically significant results that conform to the literature given such a small sample size, only similar studies over a much larger group of subjects can yield more conclusive results. The following discussion should be held in light of these considerations.

4.1 Significant Group Differences

There appear to be three major group differences in speech perception. First, although there does seem to be a similar attentive process that affects diverse ROI's on the auditory pathway, there may be a slight difference in how this attentive process interacts with the local functionality of Wernicke's area. Second, the results suggest significant hemispheric differences in activity in certain suprathalamic ROI's that are even more marked in ASD subjects. Third, there appear to be significant group differences in left AC responses to stimulus parameters which target specifically temporal mechanisms.

4.1.1 Interaction of Global and Local Attentive Processes

The interaction between stimit and trial appears in IC, thalamus, left PI, right Wernicke's area, and left and right Broca's area. In IC, thalamus, and left PI, the coefficient for trial is negative statistically significant. If it is the case that trial so consistently detracts from the responses, it would suggest that there is some higher-order attentive process whose effects on these sensory processing regions attenuate over time. This could be due to the subjects aclimating themselves to
the environment of the session, to their repeated exposure to similar stimuli, or to simple boredom. In this context, the often-seen positive trial:stimit effect would suggest that recognition of a familiar stimulus might offset the effect of trial as the subject attempts to determine the origin of the stimulus. One drawback to such an explanation is the lack of significant results for these two variables in AC, whose responses are highly subject to attentive processes [68].

Interestingly, there does not appear to be a significant group difference in this process. The only significant group difference in one of these variables is in the results for group:stimit in left Wernicke’s area, shown in Table 3.9. This suggests that stimit is correlated with a decrease in activity for TD subjects, perhaps a function of the ROI’s familiarity with the stimulus. This would conform to the role of Wernicke’s area in lexical access, although the results are not conclusive as to whether this is happening in the ROI. If this were the case, the stimuli would register as discrete, meaningful sensory units which would be held in some sort of local memory in the ROI. If this ROI responded with more activity to a novel stimulus, then TD subjects would have diminishing responses after first hearing a stimulus. If the data demonstrates that ASD subjects do not demonstrate such a familiarization, this could be related to the results of Gomot et al. [24], in which ASD subjects demonstrated superior performance in an auditory novelty detection task. Accordingly, ASD subject’s increased sensitivity to novel stimuli could stem from a diminished ability in left Wernicke’s area to aclimate to previous stimuli in comparison to TD subjects. This could then relate to the above-mentioned difficulties contextualizing the meanings of words in a conversation. The results of Harpaz et al. [30] suggest that Wernicke’s area is used to resolve lexical ambiguities. In Natural Language Processing, an effective assumption for computationally resolving lexical ambiguity of a text is “one sense per discourse” [22] - the assumption that a word’s meaning depends on the meaning of the meanings of the words in the same discourse. If this assumption applies to the brain’s linguistic processing, then resolving lexical ambiguity depends heavily on relating it to other words which are in working memory. Therefore, if the vowel stimuli are being processed as lexical items, it is very possible that these ASD responses in Wernicke’s area to previously introduced stimuli, which are similar to responses to novel stimuli, may reflect differences in how ASD subjects relate new lexical items to previously introduced lexical items.

4.1.2 Hemispheric Differences

The effect of ASD appears most strikingly in the hemispheric distribution of activity in AC, Wernicke’s areas, and Broca’s areas. In AC and Broca’s area, the activity
is left-lateralized for TD subjects, but this effect is dampened in ASD subjects. In Wernicke's area, the activity is right-lateralized for TD subjects, consistent with the results of Anderson et al. [4]. However, it is not, as one might expect, the right hemisphere which is dampened for ASD subjects. Rather, it is the left hemisphere, as the large negative coefficient for group:hemi indicates. This suggests that ASD subjects are not undergoing some sort of hemispheric reversal of activity, nor are they more evenly distributing this activity between hemispheres. Rather, the results suggest that there is some left-lateralized process in ASD subjects which is dampening their responses in this region.

The lack of any significant group differences in IC and the thalamus suggests that this process does not extend to the midbrain. Indeed, the apparent responsiveness of these regions to the am and fs variables and to the attential processes indicated by trial and stimit suggests that their functionality does not significantly differ between subject groups.

Although work has been cited herein as evidence of group differences in response patterns of the insula, Broca's area, and Wernicke's area, no such group differences appear in the present results. The results herein suggest similar response patterns in left PL in both subject groups, response patterns that conform to the literature on the ROI's role in speech processing for TD subjects. The responses in Wernicke's area and Broca's area are considerably more difficult to interpret, but the lack of group differences in these responses calls into question the results of Just et al. [36] that ASD speech processing is characterized by less activity in Broca's area and more activity in Wernicke's area.

This is not to say that there are no group differences in activation patterns in these ROI's, nor does it invalidate the results of Just et al., Anderson et al. [4], and Lange et al. [42], all of whom noticed different activation patterns in these regions for ASD subjects. It may nevertheless be meaningful that the only group difference in how an ROI responds to a stimulus is in left AC. This suggest that this region may be fundamental to ASD-specific differences in speech processing. Bodaert et al. [8] and Ceponiene et al. [11] demonstrated fundamental differences in how left AC responds to speech stimuli.

### 4.1.3 Differences in Response Stimulus Parameters

The hemispheric differences in left AC should also take responses to stimulus parameters. The significance of group:autocorr in the results suggests that it is a
particular property of the recorded speech signal which accounts for these functional differences. Tables 2.7 and 2.8 show a striking difference in this metric between recorded and synthesized stimuli. This is a metric which is sensitive to frequency components above 800 Hz. Three non-deterministic features of speech may be apparent above these frequencies - (1) rapid transients, which appear in both consonants and certain types of glottal events such as breathiness, (2) noisy sound sources in speech arising in the vocal tract, most of which are above 1500 Hz, and (3) subtle stochastic processes in the glottis would more likely disrupt the periodicity of higher harmonics [71].

Given what is known about the auditory pathway, a purely tonotopic, spectral representation of these would not be adequate to characterize them. Transients, such as the release of air in stop consonants followed by a brief pulse of wideband noise, need a precise temporal representation to distinguish them from other click-like stimuli, or from non-click stimuli with rapid onsets. A purely spectral representation of a noise band, at the resolution permitted by the cochlear filters, would not convey the stimulus’ stochastic quality. Finally, a purely spectral representation of the higher harmonics would not allow one to perceive irregularities in periodicity in the harmonics, to which speech perception has been shown to be very sensitive [64]. Therefore, it is very likely that autocorr targets features of the stimulus which require high temporal sensitivity which may not be afforded by a purely spectral representation in the auditory pathway. An ROI’s response patterns’ sensitivity to autocorr would strongly suggest an exclusively temporal mechanism. This is reinforced by the results’ suggestion that left AC is most responsive to this metric, and not pitchvar and fs, which are derived from spectral representations. If this is true, then the observed group difference in responsiveness to autocorr would suggest that there are significant group differences in the functionality of this mechanism.

The results also suggest that left AC in TD subjects is functionally sensitive to both low and high values of autocorrelation. In ASD subjects, however, left AC seems to demonstrate similar response patterns when autocorrelation is above a certain threshold. If there is some temporal-sensitive mechanism in left AC whose response properties autocorrelation captures, then these results would suggest that this mechanism is deficient in ASD subjects, and only a signal with sufficient autocorrelation, and therefore sufficient periodicity at high frequencies, could activate it in left AC.
Determining the candidates for this mechanism are beyond the scope of this study. The lack of any group interaction with $am$ in left AC discourages the choice of exclusively AM-sensitive cortical neurons. Moreover, the sensitivity to $fs$ and $am$ in the thalamus discourages the choice of AM and FM sensitive neurons in IC. Four previously observed possibilities remain. One is the temporally precise neurons in Heschl’s gyrus or other regions of AC which respond to pitch [55, 28]. An argument for this mechanism would be counterbalanced, however, by the lack of significant results for the variable $pitchvar$, which suggests that pitch modulation alone cannot account for these differences. Another possibility is the temporally precise FM-AM sensitive neurons postulated by Luo et al. [44], suggested to be independent of another group neurons exclusively sensitive to AM. The noted ASD differences in myelination could affect the temporal sensitivity of these putative neurons, and their ability to encode both FM and AM through phase modulation. If these neurons are solely responsible for FM sensitivity in the cortex, however, this candidate would be discouraged by the findings of Ceponiene et al. [11], who showed that ASD subjects do have basic FM processing capabilities. However, these basic FM processing capabilities may depend on a separate mechanism altogether - perhaps even a spectral mechanism. Another possibility is a deficiency in the inhibitory mechanism suggested by Elhahili et al. [18]. Here, the AM-sensitive gate which allowed the cortical neurons to follow the signal - perhaps equivalent to Luo et al.’s putative second group of AM sensitive neurons - would not function in a way which would allow this to function. Finally, there is the possibility that the mechanism to detect autocorrelation depends in some way on remote connectivity with other ROI’s such as right AC. Given the specializations of left and right AC, if right AC were one such region, the mechanism would most probably be involved in integrating a temporally precise representation of the signal with a spectrally precise representation. This final candidate would conform to the results of Just et al. [36] and Roberts et al. [60], who related ASD speech processing to difficulties in synchronization among various brain regions. Effective synchronization over remote regions would be essential for such an intensive task.

4.2 A “Deficient Input” Model for ASD Speech Processing

If ASD-specific differences in speech processing are at least partially localized in left AC, a scenario could arise where left AC provides significantly different input to other ROI’s whose functionality is comparable to TD subjects. In ASD subjects, other regions would receive deficient input from left AC when the autocorrelation of the signal was not sufficiently high. Anderson et al. [4] suggest that the left PI showed reduced activity in response to speech in ASD subjects, but Lange et
al. [42] support the belief that this is not because left P1 is functionally deficient. In ASD subjects, the region tended to respond to song in a comparable way to its response to speech in TD subjects. Due to their stationary pitch, sung vowels have a higher degree of autocorrelation than spoken vowels. Therefore, song stimuli might pass the threshold of autocorrelation in left AC, postulated above, whereas normal speech, with its more varied pitch fluctuation, might not as easily pass. However, the stimuli in these studies are significantly longer, posing a caveat to comparison.

The possibility of deficient input from left AC may also explain the present study’s results for Broca’s and Wernicke’s area. The results for these areas do not suggest a hemispheric reversal in functionality. Rather, they suggest a dampened leftward response in Broca’s area and Wernicke’s area. This explanation would only be plausible if Wernicke’s area were highly dependent on input from AC. The results suggest that this region is sensitive to stimulus-specific parameters, supporting the idea of such a dependence. Since Wernicke’s area is connected to AC, and since Broca’s area has strong connectivity to Wernicke’s area, deficient input from AC in Wernicke’s area would perhaps result in deficient input to Broca’s area.

The major counterweight to this explanation is the lack of a group difference for the left P1. Indeed, the region seems to respond to speech-like stimuli for both subject groups. This, however, may be due the duration of the stimuli. Both Anderson et al. and Lange et al. [42] used long speech, while Lange et al. also used song stimuli, whereas this study used short stimuli of 200 ms. Because this time is comparable to the 100 ms time constant for responses in the central nervous system, it is less likely that the insula commenced any sort of looped response to a continuous stimulus. However, it may have been long enough to encode sufficient periodicity to contrast group activation patterns in left AC.

### 4.3 Conclusion

To the degree that any conclusions can be drawn from so few subjects, these conclusions suggest that there are significant differences in left AC which correspond specifically with its ability to process temporal features of the stimulus which cannot be accurately represented tonotopically. These differences for left AC may be related to the group differences in suprathalamic differences in hemispheric activity. More tentatively, the differences for left AC may also be related to functional differences in how Wernicke’s area responds to novel and familiar stimuli, although
the lack of significant results for stimulus-specific parameters discourages concluding any sort of relationship between the two.

If these results are valid, the study will have related a quantifiable property of the signal to ASD subjects' responses to speech. In so doing, it will have narrowed down the candidate mechanisms in the brain that could contribute these differences. These results encourage further work to test the functionality of specific groups of cortical neurons in ASD subjects. It encourages a comparison MEG responses of ASD and TD subjects to speech stimuli parametrized according to AM, FM, pitch variation, and autocorrelation, to get a sense of how these stimuli relate to predicted electromagnetic responses to speech stimuli. Finally, it encourages more neuroimaging work to be done not only on the whole brain, but on different subsections of the ROI's. After all this is done, however, it may prove impossible to localize autism to any one pattern of morphological and functional differences. Temple Grandin was adamant to point out in conversation that autism, although definable by a behavioral phenotype, had a diverse range of manifestations and most likely a diverse range of causes [25].
Chapter 5

Appendix: The Source-Filter Model in an Acoustic Medium

1.1 Propagation of Sound in an Acoustic Medium

Sound is the perception of a physical event by the energy which this event has caused to propagate through an acoustic medium. A sound source is this physical event. Although energy propagates through the medium via molecular collisions, the collective behavior of the molecules allows the propagation of energy to be modeled with continuous functions.

1.1.1 Molecular Collisions Propagating Energy.

When it is not propagating energy, the molecules of an acoustic medium are at equilibrium - which can be Brownian motion in a gas, or some other configuration in a solid or liquid. A sound source causes a net displacement of molecules in the medium which propagates through the medium through directed molecular collisions. The velocity of propagation is written as $c$, and is a constant depending on the compressibility of the medium. When these collisions meet a solid surface, their directionality reverses in a process known as reverberation.

1.1.2 The 1-Dimensional Wave Equation to Model Propagation of Energy.

1.1.2.1 The Acoustic Particle.

Energy in an acoustic medium is often modeled in terms of an acoustic particle, a theoretical volume of the medium which is large enough to model energy in terms
of continuous gradients, but small enough to have near-linear gradients in energy across the volume. The number of molecules in the particle is constant, but its volume is variable. In a plane wave propagating in one dimension, the acoustic particle is a cubic volume. Energy is applied to a single surface of this volume. A particle is mass-like in that it can accelerate when force is applied. It is spring-like in that it compresses when force is applied and decompresses when force is released. Energy in the medium can be measured in terms of particle velocity \( v \), which is the velocity of the accelerated particle, pressure \( p \), which is the force applied to one side of the particle divided by the area of the surface, and volume velocity \( u \), which measures the particle velocity divided by the same surface area. Figures 1.1 and 1.2 are schematics showing the behavior of an acoustic particle in response to a sinusoidal pressure wave.

### 1.1.2.2 The 1-Dimensional Wave Equations.

The acoustic medium is compressible, and is therefore subject to Boyle’s Law of adiabatic pressure, which states that pressure and volume are inversely proportional in a closed system. Since the acoustic medium demonstrates the inerter (inertive properties) of a mass, Newton’s Law applies. Because the medium is a closed system, conservation of mass also applies. The wave equation is derived from these three laws. It defines acceptable 1-dimensional functions for \( p, v, \) and \( u \), where \( t \) and \( x \) are time and space, respectively:

\[
\frac{\partial^2 u}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} \tag{1.1}
\]

\[
\frac{\partial^2 v}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 v}{\partial t^2} \tag{1.2}
\]

\[
\frac{\partial^2 p}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} \tag{1.3}
\]

Sinusoids can be solutions to the system described by Equations 1.3, 1.1, and 1.2. A sinusoid is a periodic function which is either a cosine, sine, or a complex exponential. Complex exponentials can be used because of the equivalence \( e^{ja} = \sin(a) + j \cos(a) \). A sine describing a one-dimensional \( p, v, \) or \( u \) over space and time (a space-time sinusoid) is written as \( \sin(2\pi ft \pm \frac{2\pi f}{c} x + \theta) \), where \( f \) is frequency and \( \theta \) is the phase. \( f \) is the number of cycles per second, measured in Hertz (Hz), and can also be defined in terms of the time to complete a single cycle \( T \), as \( \frac{1}{f} \). Since a sinusoid repeats at integer multiples of \( 2\pi \), it will equal the same value when \( \frac{f}{T} \) and \( \frac{\theta}{\pi} \) are integers. A positive \( \frac{2\pi f}{c} x \) signifies that the wave is traveling forwards, while a negative \( \frac{2\pi f}{c} x \) signifies that it is traveling backwards.
Figure 1.1: A schematic of an acoustic particle when a sinusoidal pressure wave \( P(\theta) = P_m \sin(\theta) \) is applied, from \( \theta = 0 \) to \( \theta = \pi \). Dark spheres in the particle are molecules. The solid arrow is the pressure wave. The dotted arrow indicates velocity. The wave schematic indicates the phase in the cycle of a wave. \( P_M \), \( U_M \), and \( V_M \) are the maximum pressure, volume velocity, and particle velocity, respectively. Note the change in volume of the particle.
Figure 1.2: A schematic of an acoustic particle when a sinusoidal pressure wave \( P(\theta) = P_M \sin(\theta) \) is applied, continuing from Figure 1.1, from \( \theta = \pi \) to \( \theta = 2\pi \). Dark spheres in the particle are molecules. The solid arrow is the pressure wave. The dotted arrow indicates velocity. The wave schematic indicates the phase in the cycle of a wave. \( P_M, U_M \) and \( V_M \) are the maximum pressure, volume velocity, and particle velocity, respectively. Negative values indicate negative pressure, volume velocity, and particle velocity. Note the change in volume of the particle.
The phase of a sinusoid is a value between 0 and $2\pi$. It is an offset in the cycle relative to a wave whose phase is 0. Therefore, $\sin(2\pi ft + \frac{\pi}{2})$ will lead $\sin(2\pi ft)$ by one-quarter of a cycle.

1.1.2.3 Solutions for the 1-Dimensional Wave Equations.

D’Alembert’s general solution for the 1-dimensional wave equation is:

$$\sin(2\pi ft + \frac{2\pi f}{c}x + \theta) + \sin(2\pi ft - \frac{2\pi f}{c}x + \theta)$$

(1.4)

Equation 1.4 demonstrates several necessary properties of a function for $u$, $v$, and $p$. First, the function must be twice-differentiable over both space and time. Spacetime sinusoids satisfy this property. Secondly, the fact that the general solution contains two waves travelling in opposite directions shows that the sinusoids can propagate in either direction. Thirdly, the fact that the second partial is a linear operator allows the solution to be either a space-time sinusoid or a sum of space-time sinusoids. This is known as the superposition principle. Because the solutions can be sums and because the sinusoids can be imaginary, it is also possible to use complex exponentials as solutions.

The amplitude of a sinusoidal function for $p$, $v$, or $u$ describes the maximum value that this sinusoidal function can take. For a real-valued sinusoid, the amplitude is also real-valued $A$. For a complex exponential, it is also possible to use a complex amplitude which encodes both a real-valued amplitude and phase, such as $A = Ae^{j\theta}$.

Converting a real-valued signal into a complex signal can be done by phase-shifting the real-valued signal by $\frac{\pi}{2}$, and adding the phase-shifted signal as an imaginary term. This can be done via the Hilbert transform [35] for a real-valued signal $x(t)$:

$$\hat{x}(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau$$

(1.5)

where $P$ is an operator for the Cauchy principal value $Pf(x) = \lim_{\alpha \to \infty} \int_{-\alpha}^{\alpha} f(x) dx$. The complex signal is the sum $x(t) + j\hat{x}(t)$.

1.1.3 Time and Frequency Domain Representations

In a sum of sinusoids, each sinusoid has its own frequency, phase, and amplitude. Therefore, there are two ways of representing an acoustic waveform. In one, the
Figure 1.3: A spectrum of a vowel which is low-pass filtered at 2600 Hz. The spectrum represents the magnitudes of the Fourier coefficients. Here, a logarithmic measurement (dB) is used. Frequencies are in Hz x 1000.

waveform is tracked over time and space. This is called a time-domain representation. In this representation, the superposition principle depicts the signal as a sum of space-time sinusoids, each with its own frequency \( f_n \), real-valued amplitude \( A_n \), direction \( d_n \in \{1, -1\} \), and phase \( \theta_n \):

\[
x(t) = \sum_{n=1}^{N} A_n \sin(2\pi f_n t + d_n \frac{2\pi f_n}{c} + \theta_n)
\]

(1.6)

It is possible to describe the signal in Equation 1.6 as a sum of complex amplitudes, each associated with a given frequency. This is called a frequency-domain representation, or a spectrum. A spectrum is derived by a Fourier Transform. In its most general form, the Fourier Transform takes a complex-valued signal \( x(t) \), which can be derived from a real-valued signal for a frequency \( f \) using methods such as those described in Equation 1.5:

\[
F(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt = \int_{-\infty}^{\infty} x(t)(\sin(2\pi ft) - j \cos(2\pi ft)) dt
\]

(1.7)

Alternately, it is possible that \( x(t) \) is a real-valued function, although the complex-valued function yields more information about phase. This integral can be interpreted as an infinite sum of sinusoids, each with infinitessimally small amplitudes, over infinitessimally small frequency intervals. The complex exponential is useful because it encodes both phase and real-valued amplitude of a wave.
1.2 Classifying Sound Sources

A sound source is any physical process that begins this propagation through a displacement of adjacent molecules. The three major types of sounds are clicks, periodic sounds, and noise, each with its own type of originating physical process and mathematical representation. These representations describe a sound's periodicity, transience, and stationarity. Because functions to describe pressure, volume velocity, and particle velocity for a sound are respective solution to the wave Equations 1.3, 1.1, and 1.2, they can be described as a sinusoid or sum of sinusoids.

1.2.1 Clicks and Click-like Sounds

The simplest sound is the result of a single displacement in the medium, a click. A click is a sum of sinusoids at the same phase, and an ideal click is an infinite sum of sinusoids across all frequencies. Its time-domain representation is:

\[ x(t) = \int_{-\infty}^{\infty} A e^{j2\pi ft} df \]  \hspace{1cm} (1.8)

where \( A \) is real-valued and nonzero. Since all sinusoids in the integral cancel one another out at every point except where they all equal 1 or \( j \), the end result is a infinitessimally short spike with infinite amplitude, the Dirac delta function:

\[ \delta(t) = \begin{cases} +\infty & t = 0 \\ 0 & t \neq 0 \end{cases} \]  \hspace{1cm} (1.9)

Real-world clicks do not feature an infinite range of frequencies, but rather a sum of sinusoidal components over a wide range of frequencies, as shown in Figure 1.4. Clicks are purely transient stimuli, happening very quickly over time. Physical events that cause click-like sounds are events which cause a single displacement of air molecules, such as a hand clapping or an abrupt movement of a flat surface.

1.2.2 Periodic Sounds

Periodic sounds are caused by a sound source’s repeated displacement of molecules in the medium at regular time intervals. This repetitive property is called periodicity, and a periodic sound which maintains the same pattern of vibration is called stationary. A vibrating surface, such as that of a tuning fork, is an example of this type of sound, since its excursions create positive values in the waveform and its incursions create negative values. The degree to which a sound repeats can be measured as periodicity.
1.2.2.1 Pure Tones and Tone Complexes

If the positive values are equal to the negative values, the sound can be modelled as a single sinusoid, or pure tone. A related type of sound is the tone complex. In this, there are multiple sinusoids in the signal. These can be created by multiple sources, or a single source, and are modelled as a sum of sinusoids.

1.2.2.2 AM and FM Tones

Amplitude-modulated (AM) and frequency-modulated (FM) tones introduce periodicity into $f$ and $A$. If $f_{AM}$ is the AM frequency,

$$A[1 + \sin(2\pi f_{At})] \sin(2\pi ft)$$

(1.10)
Figure 1.5: A spectrum and waveform of a noisy stimulus, in this case a pronunciation of the fricative consonant /sh/. Note how the maximum amplitude is at 2500 Hz, and attenuates at 6000 Hz.

describes an AM tone. If $f_{FM}$ is the FM frequency, then

$$A \sin(2\pi[f + \sin(2\pi f_{FM}t)]t)$$  \hspace{1cm} (1.11)

is an FM tone. In describing AM and FM tones, $f$ in Equations 1.10 and 1.11 is the carrier frequency, and $f_{AM}$ and $f_{FM}$ are the modulator frequency.

**Harmonic Tone Complexes.** A particularly important type of tone complex is the harmonic tone complex. This occurs when there is distortion, in which positive values in the waveform are greater than the negative, or vice-versa. This creates a waveform of lopsided “pulses”, which occur at a frequency known as the fundamental frequency. Pitch is perception of this fundamental frequency. The Fourier transform of this signal features frequency components at the fundamental frequency $f_0$ and at integer multiples of this frequency. Components at these integer
Figure 1.6: Spectrum and waveform for a pulse train which produces harmonics. This was taken from a recording of a “buzzing” sound made with the lips. The frequency of excursions of the lips was 126 Hz. This fundamental frequency is perceived as pitch. In the spectrum, harmonics appear at 272 Hz, 383 Hz, and 523 Hz - all approximately integer multiples of the fundamental frequency. This is similar to the signal produced by the glottis.

Multiples are called harmonics:

\[ x(t) = \sum_{n=1}^{n=\infty} A_n e^{j(2\pi nf_0)t} \]  

where \( A_n = A e^{j\theta_n} \). Because these stimuli repeat every period, they are periodic. Figure 1.6 shows the waveform and spectrum of a periodic tone.

1.2.3 Noise

Noise is a sum of sinusoids over a contiguous frequency interval. Each sinusoid has a random phase, \( 0 \leq \theta_r(f) \leq 2\pi \):

\[ x(t) = \int_{f_1}^{f_2} A_{r} e^{j(2\pi f)t} df = \int_{f_1}^{f_2} A_{f} e^{j\theta_r(f)} e^{j(2\pi f)t} df = \int_{f_1}^{f_2} A_{f} e^{j(2\pi f)t+\theta_r(f)} df. \]  

Figure 1.5 shows the spectrum for a noisy stimulus, with an amplitude peak beginning at around 1500 Hz.
1.3 The Source-Filter Model

In acoustics, the source-filter model is often used to describe the interaction of sound with its environment. In this model, there is a sound source which is coupled to a filter, a physical structure which alters the signal. The filter can amplify and attenuate different frequency components of the signal, as well as alter their phase. A source’s interaction with a filter is called coupling, and is modeled using convolution. It is important to note that the categories of source and filter are arbitrary. For example, in one model the glottis can be modeled as a source, and the vocal tract can be modeled as a filter. However, if one wanted to model how a room altered a human voice, the glottis-vocal tract coupling would be the source, and the room would be the filter. A resonance is a frequency at which a filter amplifies the source. Many purely ideal filters are modeled solely in terms of complex values, and are characterized as purely reactive. Real-world filters, however, feature loss of energy through random activity in the molecular collisions. This value is termed resistance.

Because a source has a defined amplitude, it can be defined in terms of a Fourier Transform. A filter, on the other hand, may be a purely passive structure, and will respond differently to stimuli of different amplitudes. Therefore, it is characterized by a transfer function, which specifies how a filter alters both the amplitude and phase of a stimulus.

1.3.1 Time Domain Models of Convolution

In the time domain, a source \( s(t) \) is convolved with a filter whose transfer function is \( Y(f) \) as:

\[
   s(t) * y(t) = \int_{-\infty}^{\infty} s(\tau) y(t-\tau) d\tau = \int_{-\infty}^{\infty} s(\tau) \int_{-\infty}^{\infty} Y(f) e^{j2\pi f (t-\tau)} df d\tau \quad (1.14)
\]

In order to observe the effects of a filter on a source, it is important that the source contains frequency components at or near the resonance. The three types of sounds described in Section 1.2 can all be used to demonstrate the resonances of a filter. Because an ideal click covers all frequencies, it is often used to characterize a filter. The response of a filter to a click stimulus is called the impulse response. Noises and harmonic sources are also ideal to characterize the filter. A harmonic source coupled with a filter will have harmonics amplified according to their proximity to the resonance. For example, a glottal source with fundamental frequency \( f_0 \) can be modeled as \( s(t) = A \sum_{n=1}^{\infty} e^{j2\pi n f_0 t} \). A filter, with no resistance and a resonance
of $f_R$, $y(t)$, has a Fourier Transform of $Y(f)$:

$$s(t) * y(t) = \int_{-\infty}^{\infty} s(\tau) y(t-\tau) d\tau = \int_{-\infty}^{\infty} s(\tau) \int_{-\infty}^{\infty} Y(f) e^{j2\pi f (t-\tau)} df d\tau. \quad (1.15)$$

This ensures that the magnitude of the transfer function for the resonator is maximum at $f = f_R$ when $\tau - t = 0$. Since this happens at the beginning of every cycle of the resonant frequency, it ensures that the harmonics of $s(t)$ will be amplified according to their proximity to the resonance.

### 1.3.2 Frequency Domain Models of Convolution

A simpler model for the coupling of a source and filter is convolution in the frequency domain. The Fourier Transform for the source $S(f)$ can be convolved with the filter $F(f)$ by a product. In the frequency domain, the convolution is the product of these two factors $C(f) = F(f)S(f)$. The general pattern of amplitude accounted for by the filter is called the envelope of the spectrum, and can be determined by the impulse response, while the harmonics are often termed the fine structure.

### 1.4 Digital Signal Processing Basics

Although most of the ideal models for acoustics are described in continuous terms, their implementation in a digital environment requires that they be discretized. The field of digital signal processing covers this task.

#### 1.4.1 Representing a Time-Domain and Frequency-Domain Signals

##### 1.4.1.1 Time-Domain Representations

In digital signal processing, a sound is often partitioned into a series of samples, and is thus represented as a series of real-valued scalars. The frequency at which these samples occur is called the sample frequency, and is usually either 8000 Hz or 16000 Hz. In a common format for sound files, the .wav format, the amplitude of the sound is normalized to the sound system, and metrics for this are often relative to the sound system.
1.4.1.2 Frequency-Domain Representations

Integrals, such as those in Equations 1.7, 1.8, and 1.13, are implemented as sums of a discrete number of sinusoids. In these sums, all frequencies are positive. Often, a Fourier Transform is not used, due to its computational complexity, but rather some variation thereof, such as a Fast Fourier Transform (FFT). Moreover, when one wants to take multiple spectra in a signal, one uses a windowing function, which partitions the signal into a series of intervals on which the Fourier Transform can be performed. A window tapers off the signal on either side of the interval according to some mathematical function. Windows are specified by their type, their time, and their overlap. The time, or width, of a window function, is a time interval specifying the taper-off points of the function. In discrete samples, it is written as $N$. One of the most commonly used windowing functions is the Hamming window, defined for the $n$-th sample of the file as:

$$w(t) = 0.54 - 0.46 \cos \left( \frac{2\pi t}{N - 1} \right)$$

(1.16)

The overlap is expressed in percentages, and describes the overlap in the audio file of one time interval with another.

1.4.2 Implementations of the Source-Filter Model

Digital signal processing has the challenging task of modeling continuous functions in a discrete framework.

1.4.2.1 Implementing the Source

Most digital signal processing implementations of the source are straightforward, involving manipulation of the individual samples of the waveform.

Clicks. Although it is impossible to implement an ideal click described in Equations 1.8 and 1.9, it is possible to achieve click-like effects by setting the amplitude of a single sample at a significantly higher amplitude than the preceding or following samples.

Periodic Sources. Generating a single sinusoid or a sum of sinusoids is a straightforward process, since it is easy to find discretized implementations of these functions. A strategy to generate a series of pulses, similar to those produced by the glottis, is the sawtooth wave described in Flanagan [20]. The Klatt synthesizer uses polynomial approximation of this [38].
Noise. If there are no requirements for the bandwidth of the noise, it is possible to use a random number generator to generate the amplitudes of the individual samples. Alternately, it is possible to sum a series of sinusoids. Each sinusoid represents a frequency in the contiguous interval. Each sinusoid has a randomly generated phase.

1.4.2.2 Implementing the Filter

Filters are often described not in terms of their physical parameters, but rather in terms of their resonances. As seen in Equation 1.14, time-domain convolution ultimately depends on a frequency-domain transfer function for the filter. Therefore, it is expedient to describe a digital signal processing system in terms of the transfer functions of the filter. A digital signal processing system will often use digital resonators, which amplify the source at a single resonance. For example, frequency-domain transfer function for a theoretical resonator with only one resonant frequency is:

\[ R(f) = \frac{A}{1 - Be^{-j2\pi fT} - Ce^{-4\pi fT}} \]  

where \( A, B, \) and \( C \) are amplitudes and \( T \) is the period of the resonant frequency \([38]\). Another example would be:

\[ R(f) = \frac{1}{(j2\pi f - j2\pi f_R)(j2\pi f + j2\pi f_R)} = \frac{1}{pi^2(f_R^2 - f^2)} \]  

where \( f_R \) is the resonant frequency \([71]\). It is important to note that this has no resistance \( R \), which is defined by a real value. Therefore, the magnitude of this ideal resonator would be infinity at resonance - the transfer function for a resonator with a resistance of \( R \) would have a magnitude of \( \frac{1}{R} \) at resonance. Multiple resonant frequencies in a resonator can be modelled by taking the product of these transfer functions, as in \( T(2\pi f) = \prod_{i=1}^{N} R_i(2\pi f) \). Therefore, the transfer function of an individual tube can be modelled by taking the product of these transfer functions with frequencies at the resonances of the tube, as in \( F_T(2\pi f) = \prod_{i=1}^{N} T_i(2\pi f) \). The impulse response of a non-Helmholz concatenation of tubes can be modelled as a product of the transfer functions for each tube: \( F_R(2\pi f) = \prod_{i=1}^{N} T_i(2\pi f) \). If the concatenation forms a Helmholz resonator \( R_H \) this function becomes: \( F_R(2\pi f) = R_H(2\pi f) \prod_{i=1}^{N} T_i(2\pi f) \)

1.4.3 Deconvolution

Deconvolution is a method of extracting information about the source and filter from a signal. Most algorithms which perform this in the frequency domain focus
Figure 1.7: The magnitude of a transfer function for the resonator described in Equation 1.18. Frequency is in arbitrary units. From Stevens [71].

\[ S(2\pi f) = \frac{C(2\pi f)}{F(2\pi f)} = \frac{F(2\pi f)S(2\pi f)}{F(2\pi f)} \]  

(1.19)

Clearly, extracting the transfer function of the filter is the crucial part of this operation.
Bibliography


[45] I. Noens, I. van Berkelae-Onnes, R. Verpoorten, and G. van Dujin. The com-
for: an instrument for the indication of augmentative communication in peo-
ple with autism and intellectual disability. *Journal of Intellectual Disability

[46] M. T. Le Normand, S. Boushaba, and A. Larechet-Dujour. Prosodic distur-


[49] J.M. Paccia and F. Curcio. Language processing and forms of immediate
echolalia in autistic children. *Journal of Speech and Hearing Research*,


[51] E. Paulesu, C. D. Frith, and R. S. J. Frackowiak. The neural correlates pf the

[52] V.B. Penhune, R.J. Zatorre, J.D. MacDonald, and A.C. Evans. Interhemi-
spheric anatomical differences in human primary auditory cortex: probabilis-
tic mapping and volume measurement from magnetic resonance scans. *Cere-
bral Cortex*, 6:661–672, 1996.

[53] H. Platel, J. C. Baron, R. Wise, J. Lambert, R. S. J. Frackowiak, B. Lecheva-
lier, and F. Eustache. The structural components of music perception. a func-

-port/teach/641/hearing.for.linguists.html.

activates pitch processing centre in heschl’s gyrus. *Neuroimage*, 49(2):1641–
1649, 2010.


[57] A. Rees and A.R. Moller. Responses of neurons in the inferior colliculus of

82


