5-1-2014

Maturing Temporal Bones as Non-Neural Sites for Transforming the Speech Signal during Language Development

Lisa Hogan

Recommended Citation

Hogan, Lisa. "Maturing Temporal Bones as Non-Neural Sites for Transforming the Speech Signal during Language Development." (2014). https://digitalrepository.unm.edu/anth_etds/32

Lisa A. Hogan

Candidate

Anthropology

Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Hillard Kaplan , Chairperson

Larry Gorbet

JaneBuikstra

Caroline Smith

Sherman Wilcox

Maturing Temporal Bones as Non-Neural Sites for Transforming the Speech Signal during Language Development

by

Lisa A. Hogan

B.A., English, University of Notre Dame, 1980
M.A., English, University of New Mexico, 1985
M.A., Anthropology, University of New Mexico, 1996

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Anthropology

The University of New Mexico
Albuquerque, New Mexico

May 2014

DEDICATION

- To my family and friends—their support over the years has carried me through to finishing this work
- To my committee for their many years of assistance and instruction
- To the speakers of Diné and English from whom I collected data—the most rewarding part of this study was the time we spent together
- To the members of the Students Organized against Dissertation Avoidance Behavior (SO DAB) group—John Y. Anderson, David R. Margolin, Mami McGraw, all of whom are much more to me than professional colleagues

ACKNOWLEDGMENTS

It may take a village to raise a child, but it takes a virtual army of intelligent and educated people to turn out a PhD. I don’t believe I would have ever completed this dissertation if I didn’t feel like it would be a shame to waste the efforts of so many bright people who helped me. Here is an effort to acknowledge and thank the consultants, participants, and members of my dissertation committee who were most closely involved in assisting me with this project:

John Yard Anderson
LeRoy H. Begay, Sr
Ed Bedrick
Becky Zerger Bentley
Jon Bentley
Jane E. Buikstra
Melvatha Chee
Warlance Chee
Ron Christensen
Earl Dressel
Larry Gorbet
Sandra Harada Gabaldon
Lynn A. Hertel
Michael J. Hogan
Janie Hogan
Tess Hogan
Hillard Kaplan
Boris Margolin
David R. Margolin
Miltina Morgan
Mami McGraw
Michelle P. Otterness
Louis Owens
Donald Partridge
Margaret White Pedrotty
Mario Ruiz
Scott Sanders
Caroline L. Smith
Patricia Smith
Hector Torres
Maiya Tyrrell
Shannon Whitehead
Sherman Wilcox

PREFACE

Linguistic anthropology is the study of how verbal signals develop into meaningful messages that, through their use in human communities, forge and sustain cultural practices (Foley 1997). Traditionally linguistic anthropologists have focused their study on forms of linguistic interaction as they are embedded in the context of a particular sociocultural setting with the objective of arriving at cultural understandings—the work done in the ethnography of speaking is an exemplar for the field. Yet what has always informed the study of language by linguistic anthropologists is a perspective that recognizes the roles of not only social factors, but also symbolic, cognitive, and biological factors. As a logical result of the training anthropologists receive in the four fields of the discipline, this perspective has encouraged a wider view of linguistic context than that generally taken into account by investigators in other academic disciplines involved in the study of language. This dissertation has undertaken to study a form of linguistic interaction (that is, the development of vowel systems) as it is embedded in the intersection of the biological context of the developing child and the verbal context of the child's ambient language or languages.

Context, linguistic and extralinguistic, permeates the expression and understanding of language in many ways, but the scientific study of its influences poses a challenge for investigators. As Elinor Ochs and Bambi Schieffelin (1979: 2) observe, one difficulty is that "not all entities in a physical space constitute context. Rather, context consists of environmental features that form part of the language user's universe." In other words, language users select from a range of possible environmental features that to which they will attend. This selection applies to not just environmental context but to linguistic structural levels as well, all the way from the selection from all possible phonetic sounds to make up an inventory that underpins a language's phonology to conscious or unconscious selection from vocabularies and syntactic variants to create a register or style of speech. While some features are "fixed" in the physical environment, participants can exercise the option to ignore one or more of these fixed features. Thus we find the interpretation of fixed features is an area permeated with the influences of culture and society. Further, this "environment" includes other participants in the social setting, as well as their shared and individual histories. We adults then turn to our infant learners and teach them more or less overtly, depending on the community, what does and does not count in their environmental and linguistic worlds.

In his discussion of speech events, Dell Hymes (1974) covers some of the same ground as Ochs and Schiefflin (1979) with his mnemonic "SPEAKING" (setting, participants, ends, act sequence, key, instrumentalities, norms, genre). In Hymes's view, context is a feature of speech events that recurs and that, because of cultural and social conventions, carries a set of constraints and rules. As Hymes points out, speech events vary a great deal from one community to another. One of the things that people know when they know a language is the relationship or configuration of components of the speech events that are recognized within the ambient community. The ethnographic evidence indicates that no known human community specifies all possible components. Which ones are selected for specification and which ones are ignored vary from community to community, again very much like the process that create phonologies from phones. Functionally, the capability to make selections at the contextual level (as well as at more concrete levels of language) forms the basis of variation and flexibility, allowing different communities and different individual actors to produce different combinations of components.

In these respects—selection from a range of possibilities, redundancy, and relationships between different components—speech events resemble other, more formal, more concrete levels of linguistic practice, such as sound systems. This resemblance has been remarked upon by many scholars, among the first of them being Edward Sapir. Presaging work by a generation of cognitive linguists such as Gilles Fauconnier (2002, 1997, 1985) and Ron Langacker (2000, 1991, 1987), Sapir (1921: 17) notes that the “birth of a new concept is invariably foreshadowed by a more or less strained or extended use of old linguistic material…, a thing wrought from linguistic material already in existence in ways mapped out by crushingly despotic precedents.” He goes on to explain:

> The concordance between the initial auditory imagery and the final auditory perceptions is the social seal or warrant of the successful issue of the process. As we have already seen, the typical course of this process may undergo endless modifications or transfers into equivalent systems without thereby losing its essential formal characteristics” (Sapir 1921: 18).

Sapir thus ascribes a role to auditory imagery and perception in shaping the form of other levels of language use. John J. Gumperz summarizes Sapir’s views on linguistic structure thus: “All human beings, informants and linguists alike, tend to prejudge or edit the sounds they hear. Linguistic structure is more than a mere scholarly construct. Structure constrains and potentially predicts the speaker’s perception of verbal stimuli” (Gumperz 1972: 6).

Language context is a construction, the result of negotiation among and within language users (including the results of negotiations passed from generation to generation), a selection made from many possibilities—a filter if you will—brought to bear by participants in a speech community on their interpretation of their experience, linguistic and extralinguistic, resulting in the "reduction of the random by 'restraint'" (Bateson 1972: 132). Linguistic anthropology and this dissertation are concerned with what Goffman (1974: 13) called "the organization of experience," which is not limited to the organization of society. There's cultural filtering and, as this dissertation explores, there's biological filtering. This dissertation will study the influence that immature temporal bones in humans have in filtering speech sounds that the peripheral auditory system (PAS) conducts to the auditory nerve.

MATURING TEMPORAL BONES AS NON-NEURAL SITES FOR TRANSFORMING THE SPEECH SIGNAL DURING LANGUAGE DEVELOPMENT

by

Lisa A. Hogan

B.A., English, University of Notre Dame, 1980
M.A., English, University of New Mexico, 1985
M.A., Anthropology, University of New Mexico, 1996
PhD., Anthropology, University of New Mexico, 2014

ABSTRACT

MATURING TEMPORAL BONES AS NON-NEURAL SITES FOR TRANSFORMING THE SPEECH SIGNAL DURING LANGUAGE DEVELOPMENT

Developmental events in the temporal bones shift the pattern of a given speech sound's acoustic profile through the time children are mapping linguistic sound systems. Before age 5 years, frequency information in vowels is differentially accessible through the years children are acquiring the sound systems of their native language(s).

To model the acoustic effects caused by developing temporal bones, data collected to elicit steady-state vowels from adult native speakers of English and Diné were modified to reflect the form of children's hearing sensitivities at different ages based on patterns established in the psychoacoustic literature. It was assumed, based on the work of psychacousticians (e.g., Werner, Fay & Popper 2012; and Werner & Marean 1996), that the effects caused by immature temporal bones were conductive immaturities, and the age-sensitive filters were constructed based on psychoacoustic research into the hearing of infants and children. Data were partitioned by language, sex, and individual vowels and compared for points of similarity and difference in the way information in vowels is filtered because of the constraints imposed by the immaturity of the temporal bones.

Results show that the early formant pattern becomes successively modified in a constrained pattern reflecting maturational processes. Results also suggest that children may well be switching strategies for processing vowels, using a more adult-like process after 18 months. Future research should explore if early hearing not only affects individual speech sounds but their relationships to one another in the vowel space as well. Additionally, there is an interesting artifact in the observed gradual progression to full adult hearing which may be the effect of the foramen of Huschke contributing to the filters at 1 year and 18 months. Given that immature temporal bones reflect brain expansion and rotational birth in hominids, these results contribute to the discussion of the biological underpinnings of the evolution of language.

TABLE OF CONTENTS

FRONT MATTER **Page**

DEDICATION III

ACKNOWLEDGMENTS IV

PREFACE V

ABSTRACT IX

CHAPTERS **Page**

1 INTRODUCTION **1**

2 THEORIES OF CHILD LANGUAGE ACQUISITION AND DEVELOPMENT **11**

2.1 LANGUAGE ACQUISITION/DEVELOPMENT THEORIES 12

2.2 THEORIES OF THE DEVELOPMENT OF PHONETIC CATEGORIES 20

3 THE HUMAN TEMPORAL BONES AND BRAINS **33**

4 METHODS **61**

4.1 DATA COLLECTION 61

4.2 DATA MEASUREMENTS 66

4.3 DATA FILTERS 67

APPENDIX 4-A – RANDOMIZED ORDERED CARRIER SENTENCES READ BY PARTICIPANTS 73

5 RESULTS AND DISCUSSION 83

5.1 GENERAL EFFECTS CAUSED BY IMMATURE HEARING 83

5.2 AGE EFFECTS ON ENGLISH AND DINÉ VOWELS IN AGGREGATE 91

5.3 DINÉ VOWELS 94

5.4 AGE EFFECTS PARTITIONED BY SEX 98

5.5 CONCLUSIONS ABOUT TRAINING EFFECTS 109

APPENDIX 5-A – COUNTS OF FREQUENCY BINS FOR MALE AND FEMALE VOICES IN DINÉ AND ENGLISH 111

APPENDIX 5-B – AVERAGE FORMANT FREQUENCIES OBTAINED WITH PRAAT'S GET FORMANT COMMAND 115

6 CONCLUSIONS 125

6.1 RETURN TO PREDICTIONS 125

6.2 FUTURE RESEARCH 126

7 REFERENCES...131

1

INTRODUCTION

To study developmental effects on vowels as children grow into their maturity, this dissertation uses an "auditory" (as contrasted with "acoustic") analysis of the speech signal based on a model of psychoacoustic data collected on children's sensitivities to frequencies, which differ from those of adults (Werner et al. 2012; Werner 2007; Werner & Marean 1996). Phoneticians and speech scientists currently understand that in adults acoustic analyses of speech do not match listeners' experience of those sounds for a variety of reasons. Variables that affect all listeners' experience include the direction of the individual listener's attention, the particular linguistic environment, physical properties of airborne sound, and the filtering of sounds through the peripheral auditory system (the outer, middle, and inner ear). Acoustic signals are transformed before they become speech percepts, although frequency information is far from being the only type of information that goes into speech perception (see Hawkins 1999a, b, c for a cogent review).

The mismatch between acoustic signal and speech percept results from the ways the peripheral auditory system responds to air-conducted sound and the ways the central auditory nervous system responds to stimuli. Recent years have seen increased attention paid to the fact that processing of auditory information does not take place solely in the auditory cortex; that is, numerous subcortical sites are performing extraction and analysis of the dimensions of sounds (e.g., Masterton 1992). **This dissertation explores the possibility that the temporal bones function as non-neural "extractors" of dimensions of sound by considering whether or not developmental events in the temporal bone contribute any kind of structure to the acoustic waveform of speech sounds.** For example, Figure 1-1 illustrates the difference between an acoustic analysis (light line) and an auditory analysis (dark line) of a complex wave. (The wave depicted here is composed of a 500-Hz sine wave and a 1,500-Hz sine wave—if all frequencies were present at an equal amplitude, the two lines would be horizontal). In the auditory analysis depicted in Figure 1-1, the peaks for the two component waves appear broader and better separated than in the acoustic analysis. The auditory analysis is based on a model of psychoacoustic data (the bark scale developed by Fletcher & Munson 1933) that shows the adult auditory system is most sensitive to frequencies between 2 and 5 kHz, reflecting the boost given by resonance of the external auditory meatus and the transfer function of the ossicles. This dissertation concerns itself with the fact that, as compared with adults, children have additional, different variables involved in shaping the form in which speech sounds arrive at the inner ear, variables that disappear when they reach skeletal maturity, around age 20 years.

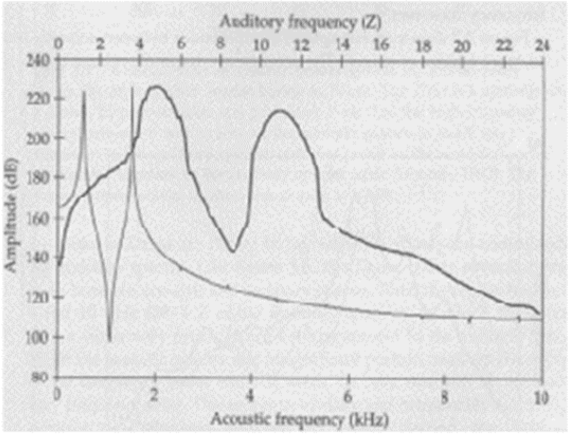


Figure 1-1. Difference between an Acoustic Analysis (light line) and an Auditory Analysis (dark line) of a Complex Wave (from Johnson 1997: 57)

Other experimenters have carried out studies using mathematical modeling to represent filtering of the peripheral auditory system on adult listeners' perception of speech (e.g., de Boer 2000; Kuhl 2000; Lum & Braida 2000; Molis 1999; Slaney 1999; Patterson, Allerhand & Giguère 1995; Ghitza 1993; Holdsworth et al. 1992; Iverson & Johnson 1992; Patterson et al.1992; Forrest et al. 1988; Seneff 1988; Mann & Liberman 1983; Moore & Glasberg 1983; Schroeder, Atal & Hall 1979). The research described in this dissertation continues and extends this prior research into the peripheral auditory system's filtering to include effects on child listeners' perception of speech. Using a series of auditory splines calculated to model child listeners' experience of speech sounds (see Chapter 4), this dissertation explore whether or not human temporal bones transform auditory/linguistic signals in ways important to children learning spoken languages during the years the bones develop and ossify (see Chapter 3 for further details about this developmental process).

To date, no research has used auditory models to consider whether or not the speech signal experienced by children undergoes any transformations of the speech signal that adults no longer experience, even though there has been research conducted in the last 20–35 years that has considered the fact that humans of different ages experience sound differently.[1] For example, Querleu and colleagues (1989, 1985) and Lecanuet and colleagues (1998, 1987) take up the question of what fetuses can experience of sound in utero (fetal hearing is perforce via bone conduction), and it is a common practice for investigators to use low-pass filters to model the speech signal that newborns and infants experience via air conduction (e.g., Mehler et al. 1988; Spence & Freeman 1996). Querleu

---

[1] Though fetal and newborn hearing is discussed in this paragraph, the study of hearing in the aging has a burgeoning literature. For example, Patterson, Nimmo-Smith, Weber & Milroy (1982) studied losses in frequency selectivity that occur with increasing age in adults using a model that featured filtered speech masked by a notched noise.

and colleagues (1980) established that while the high frequencies of voices in the fetus's environment are attenuated, prosody is particularly well preserved. It is established that newborns recognize their mothers' voices (DeCasper & Fifer 1980) which must be based on their experiences in utero beginning around 22–24 gestational weeks of the lower frequencies of their mothers' voices (under 400 Hz), which is the range of the fundamental frequency of voices.[2] The fundamental frequencies of a voice reflect the contribution of the larynx in a voice, influenced by the length, size, and tension of the vocal folds (Laver 1994, 1980) and, as a "speaker-specific acoustic parameter," they are important in humans' ability to identify an individual by his or her voice (Künzel, Masthoff & Köster 1995: 291). It is clearly adaptive behavior for an infant to recognize and interact with the mother's voice from birth.

In addition to the speaker-specific frequency information, fetuses also have available to them information about the pitch-contours, tempo, and rhythms of the language or languages that they are to be born into. DeCasper and Spence (1986) found that fetuses who were read a poem in a loud voice by their mothers between gestational weeks 33 and 37 responded with a decreased heart rate only to the poem they had been hearing for 4 weeks when that poem was read by a third party in a mix of new poems. This experiment shows that a fetus is becoming familiarized with the mother's language, not just her voice. Such prosodic information may be helpful to infants learning to parse the speech stream into a hierarchy of units (e.g., Nespor 1990; Mehler, Sebástian-Gallés & Nespor 2004) or at the very least to tune into speech as speech (e.g., Mehler et al. 1988).

---

[2] Males have an average fundamental frequency of 120 Hz, females, 220 Hz (Fant 1956).

Nonetheless, the possibility that the hearing of human children is affected by auditory effects created by maturational processes has received little attention. No study has yet examined speech sounds using an auditory model that takes into account that children's sensitivity to frequencies differs from that of adults. To this point, most researchers have not thought that any differences between children's and adults' hearing have any functional significance: some compelling behavioral studies have demonstrated that by 6–9 months infants discriminate phonetic contrasts used in their native language(s) and not those used in other languages, where younger infants discriminate all phonetic contrasts (e.g., Best et al.1995; Polka & Werker 1994; Werker & Tees 1984; Aslin et al. 1981; Eimas et al. 1971). This research is analyzed and discussed in some detail in Chapter 2, but it is noted here that these results are based on psychoacoustic data gathered from experiments in which the stimuli are syllables, not just vowels. Nittrouer (1994: 2997) studied developmental changes in auditory capacities and phonemic awareness and reported results that suggested "the weights assigned to various acoustic properties of the speech signal are retuned as children gain experience with their native language." For this dissertation, only steady-state vowels are studied for the frequency information they contain.

One of the goals of this dissertation is to examine auditory "images" that demonstrate how selected speech sounds (steady-state vowels) in two different languages, English and Diné (Navajo), might be experienced by infants and children at different ages, akin to those images produced by researchers working with the development of vision (see Figure 1-2). Vision researchers have advanced the theory that not only do infants see well enough to fulfill their role as infants but also that greater visual acuity might impede them

(Lickliter 1996; Hainline & Abramov 1992; Turkewitz & Kenney 1982). Too much visual acuity too early, they hypothesize, adds visual noise that prevents the infant from attending to the most relevant information; the limitation thus provides structure and order to experience.


(a) (b) (c)

Figure 1-2. This series of photographs (from Hainline 1998: 24) models how the face[3] seen in (a) would look to a 2-month-old infant at (b) a distance of 30 cm—the distance to a face from the crook of the elbow as a child nurses—and at (c) 150 cm, a distance which seems to be close to a threshold for detectability.

This dissertation models the filtering of speech sounds brought about by the noise in the peripheral auditory system of infants and children induced by the immaturity of the temporal bones in order to consider what might be the most relevant information in vowels at different times during the years that the young are building the maps of the vowel systems of their native language(s).

Developmental psychobiologists also suggest that the “uneven rate of development and sequential onset of functioning of the sensory systems have consequences for the

---

[3] The author of the article in which this figure appears (Hainline 1998: 24) is careful to state that the figure is published as a tribute to Princess Diana.

development of relationships between them" (Turkewitz & Kenny 1982: 359), meaning that the functioning of later developing senses is assimilated into the existing framework established by earlier developing senses. In human brain development, neuroscientists find the same general pattern for the onset of sensory function (cutaneous, vestibular, olfactory, auditory, and visual) as that identified by Gottlieb (1971) for birds and mammals. They also find that myelination of the central auditory system proceeds on a peripheral-to-central gradient, a pattern shared by other species. Nonetheless, the pattern for the *timing* of sensory maturation differs from species to species (Figure 1-3 shows the human pattern). That is to say, some species are born in a more or less precocial state with "all systems go," while others are born in an altricial or nearly fetal state.[4] Some, like humans, are born with some senses more mature than others and with senses that will mature at different rates. As Michel and Tyler (2005: 157) point out, "de Beer (1958) and Gould (1977) argued that such variability was the foundation for the evolution of species variability. Comparisons of the patterns of development among closely related species revealed that the origin of morphological differences, with important adaptive consequences, derived from differences in the timing of typical sequences of developmental events."

---

[4] Humans are classified as secondarily altricial in comparison with other mammals (e.g., Gibson 1996, 1991). While humans are born with immature locomotive skills, they are neurologically semiprecocial, both in terms of percentage of adult brain size achieved by birth and in terms of the state of neural myelination at birth (see Figure 1-3). By way of reference, newborn rhesus monkeys are highly precocial, skeletally and neurologically. Chimpanzees fall between humans and monkeys skeletally and on the basis of percentage of brain size achieved by birth (neural myelination at birth has not been studied in chimps).

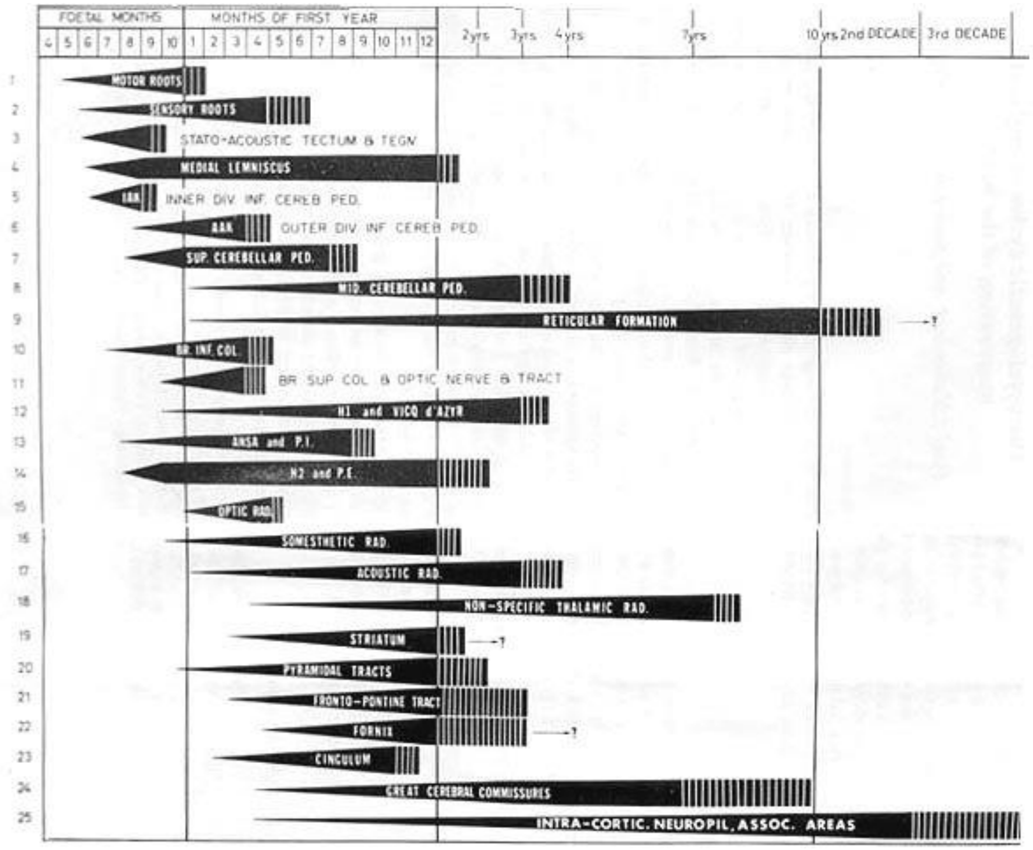


Figure 1-3. This figure (redrawn from Konner 1991: 195) illustrates the patterns of myelination[5] in the human brain. Interrupted bars to the right show the age range for the process's termination. Lines 11 and 15 show that humans attain visual maturity in the first year after birth (some species of mammals are born visually mature). Line 17 shows that the maturation of audition is a more protracted process than the maturation of vision.

The human pattern suggests that postnatal audition is integrated into the framework established by the more rapidly maturing statoacoustic and visual systems. Lines 3 and 5 in Figure 1-3 show that systems subserving detection of postural orientation and vestibular stimulation are fully myelinated at birth. Thus we might find here the biological basis for Len Talmy's (1988, 1983) linguistic analyses examining how spatial orientation and vision interpenetrate the structure of language. Further, recent work with quail, ducklings, and rats has shown that modifications to this species-specific timing—premature visual

---

[5] The formation of myelin (a fatty conductive sheath around neuronal axons) does not indicate functional onset, but rather functional efficiency and specificity (Gibson 1991).

stimulation in these studies—can interfere with normal functioning in earlier maturing systems, which include audition and olfaction in these species (e.g., Foushée & Lickliter 2002; Gottlieb, Tomlinson & Radell 1989; Kenney & Turkewitz 1986). The topic of Chapter 5 (results and discussion) is whether or not the transformations brought about by the developing temporal bones contribute any "training" effects helpful to infants and children during the time they are learning the sound systems of their ambient language(s).

Understanding whether or not the development of the temporal bones affects the way children learn the sound systems of their language can contribute to several areas of scholarship. In general, not enough is known about the way human brains learn language and what can go wrong during this process. For example, dyslexia, a disorder in which otherwise normally intelligent children cannot easily learn how to read and write, is now thought to be not only a visual disorder, but also a deficit in phonological coding and a deficit in phonemic segmentation (e.g., Stein & Walsh 1997; Vellutino 1987). Gerrits and de Bree (2009: 180) found in their data that 3-year children at risk for familial dyslexia and 3-year-olds already diagnosed with specific language impairment had poorer speech perception and production as compared with controls and that "their impaired expressive phonology seemed to be related to a deficit in speech perception." Moreover, children with recurrent otitis media before the age of 3 years experience difficulty learning to read as late as age 9 years (Luotonen et al. 1996).

Studying the normal development of the temporal bone and any auditory effects on the sounds of language may shed light on developmental problems with linguistic coding. Improved understanding of the biology of language might also lead to some kind of device that would make it easier for adults to learn the sound system of a new language. Recent research shows that, given the appropriate kind of listening experience, adult listeners can realize some improvements in learning sound contrasts present in their new language but absent from their first language (Best & McRoberts 2003; Kuhl 2000; Pisoni, Lively & Logan 1994). Golestani and Zatorre (2009) studied 59 English speakers learning to distinguish the Hindi dental–retroflex contrast, as well as a tonal pitch contrast. They concluded that while there was considerable range in the performance of individuals, training resulted in overall group improvement in the ability to identify and to discriminate the phonetic and the tonal contrasts. Interestingly, Bundgaard-Nielsen and colleagues (2012: 643) found "second language learners' vocabulary expansion is associated with improved second language **vowel intelligibility** [emphasis added]" in their study of adult Japanese speakers learning Australian English.

The debate between nativists and constructivists centers on the question of the interaction between brain and culture. This dissertation explores if a third component, the body, ought to be considered in such discussions. Additionally, many anthropologists and linguists assume that any linguistic universals must arise from perceptual mechanisms but the discussion has heretofore concentrated on the role of vision in shaping language; the role of audition, too, ought to be given attention. Last, a better understanding of the biological mechanisms underlying language may contribute to a better understanding of the origins and evolution of language. Many researchers assume that language was a selection pressure on brain size, but as yet there is no well-accepted theory as to how or in what ways (e.g., Hurford, Studdert-Kennedy & Knight 1998; Deacon 1997; Noble & Davidson 1996).

2

THEORIES OF CHILD LANGUAGE ACQUISITION AND DEVELOPMENT

That no known language employs the entire set of possible speech sounds means that all infants extract from their ambient languages a particular inventory of speech sounds. Further, humans cope with a great deal of variability in the individual renditions of speech sounds that they hear from both a given speaker and from across speakers. These two facts prompt several questions. How do language users come to understand a given speech sound as being the same sound when it is produced by different vocal tracts and in different acoustic/articulatory contexts, especially in the view of the fact that what universally characterizes speech processes is variability (Lindblom 2000; Lindblom, Perkell & Klatt 1986)? Most linguists agree that infants learn (and children and adults use) some principled system of categorization for the speech sounds they experience, but what sort of structure does this system have? Does this system look or work the same in all languages at all stages of linguistic development? Does learning influence phonetic development, and does phonetic development influence learning?

Few linguists would maintain that learning and development are completely separable processes (or the one and same process) during the time when children are learning their native languages. Linguists generally agree that learning and development interact when children acquire languages. The principal areas of disagreement concern the nature of the interaction between the two processes. The familiar dichotomy of nativists and empiricists found in the discourse of general scientific investigation becomes cast among linguists in terms of discovering to which extreme of the dichotomy the mechanisms of language learning tend.

2.1 LANGUAGE ACQUISITION/DEVELOPMENT THEORIES

In theories of language learning, a nativist position is one that stresses the role of innate mechanisms. Without the constraints of innate mechanisms, Chomsky (1965) argues, persuasively, that there is little mathematical probability that children would induce the same exact set of rules that their parents use to structure the community's klanguage(s). In this view, children are able to acquire their native language(s) because of specific mechanisms hardwired into the brain and dedicated to language. The environment is mainly seen as a trigger, and experience with a language is necessary for setting "switches" or options within innately determined parameters, but language learning does not affect linguistic structure or representation in the brain in any fundamental way (e.g., Chomsky 2000, 1988; Lightfoot 1991; Meisel 1995; Osheron, Stob & Weinstein 1986; Wexler & Culicover 1980).

Historically, the nativist position in linguistics formed in reaction to behaviorism (a type of empiricism) and its account of child language learning. (The behaviorist account is generally considered to be refuted in Chomsky 1959.) Promulgated by B. F. Skinner (1957), the behaviorist account stressed the role of external stimulus and reinforcement in shaping verbal behavior. In this account, infants bring no innate information to the task of learning languages. Moreover, in the behaviorist account language input itself does not cause language to emerge—language learning is brought about through the frequency, arrangement, and withdrawal of a particular form of reinforcement.

While the empiricist/constructivist position developed in more recent years stresses the role of experience in language learning, it is not a behaviorist argument because it assigns weight to the role of language input in language development. Neither does the constructivist agree with the nativist that what that language input does is set switches within prespecified options. Constructivists theorize that language input interacts with development which in turn advances development that enables or prompts more learning (see Kuhl 2001, 2000 for reviews). Cognitive linguists (a type of constructivist) claim that children are able to learn their native language(s) because of general cognitive mechanisms whose primary task is the organization of perceptual experience, not because of a specialized language acquisition device (e.g., Tomasello 2000; Elman et al. 1996). For instance, Langacker (2001) cites perception, attention, memory, and categorization as examples of cognitive phenomena from which language cannot be dissociated, while Talmy (1983) singles out conceptions of space, force, and time. Langacker (1991, 1987), Talmy (1988), Lakoff (1987), and many others document several ways visual organization penetrates the organization of semantic and syntactic structures.

With their emphasis on the role of social experience and interaction in constructing language, the "social interactionists" (e.g., Tomasello 2001, 2000; Bruner 1983) have also contributed to our understanding of language development, complementing the work of cognitive linguists and anthropologists about the role of language input. Working from a standpoint influenced by the Boasian tradition (Silverstein 1981, 1976; Whorf 1956; Sapir 1949; Boas 1911), anthropologists and others studying language development and social interaction have documented most convincingly that neural mechanisms are only partly responsible for guiding or constraining language development in children (see Ochs & Schieffelin 1995 for a review). With their emphasis on cultural influences and joint attention, social interactionists work within a framework that is more phenomenological in its character (vis-à-vis Husserl or Foucault) than the frameworks within which nativists or cognitive linguists work. Social interactionists claim that when children learn languages, they learn a set of social acts. In these social acts, one interlocutor tries to get another to focus attention on something in the world (e.g., Halina, Rossano & Tomasello 2013; Liebal, Carpenter & Tomasello 2013; Behne et al. 2012; Rossano, Carpenter & Tomasello 2012; Tomasello 1999). Further, some researchers (e.g., King 1994; Turkewitz & Devenny 1993) point out that the interlocutor who actively compels another's attention need not always be the experienced speaker—infants and children, too, act to bring about their own socialization. Through joint attention, proficient language users impart to children a sense of the ways in which previous generations of a social group have construed or ”framed” the perceptual world depending on the immediate communicative goals of the situation (a view which evokes Goffman 1974). To Geertz (1973: 22), who identifies the question of how the great natural variation of language and other cultural forms is “to be squared with the biological unity of the human species” as anthropology's deepest theoretical dilemma, social interactionists would respond that what unifies the human species is cultural transmission, which, while not organic, is a biological mechanism nonetheless.

In general, a nativist takes the view that the complexity of language is not explained by the input children receive (the “poverty of stimulus” argument) and that any adequate model of language requires genetically endowed, innate constraints on the direction and nature of what children generalize from the learning situations into which they are born. Quartz and Sejnowski (1997) observe that for a nativist, since learning does not and cannot influence a genetically endowed development, all of development might as well be an instantaneous process. The nativist position predicts that removing temporal elements from the acquisition profile of a developing system will make no difference. It is the experience itself, not the experience over time, that matters to the language acquisition device (LAD). By contrast, in general, a constructivist holds that children develop the skills to organize their perceptual experience—of which language is but one aspect—by noticing (moreover being assisted in that noticing by experienced speakers) and remembering regularities in the input. It is experience over time with the ambient language(s) that is critical because of the interactions that unfold between learning and experience.

Exploring whether or not the developing temporal bones contribute any training effects to the speech signal will not settle the controversy between nativist and constructivist approaches to language development. But a result of no effect in this investigation would suggest that the current question is framed properly as an interaction of neural and behavioral mechanisms, that is, an interaction of the brain and culture. Should there be an effect caused by the development of the temporal bones, such a result would constitute evidence of a non-neural mechanism at work, one that heretofore has been overlooked, one suggesting that an interaction of brain, body, and culture plays a role in shaping the direction and nature of children's language learning. Such evidence would not necessarily preclude the existence of a genetically endowed language acquisition device as postulated by nativists, but evidence that the speech signal is simplified during early language learning would make it seem more probable that the sound systems of languages could be acquired with procedures for general-purpose learning. The investigation will also provide evidence as to whether or not the development of the temporal bones changes the nature of the language learner's experience with the ambient language(s) over time—this question has not received any attention to date by constructivists.

In recent years, some challenges have arisen to the assumptions about the nature of language and the proper goals of linguistic theory that underlie the nativist and constructivist positions. Both nativism and constructivism share as their point of departure a realist's belief in mental representations, or "symbols," that translate sensory experience into perception (Gordon & Slater 1998; Tomasello 1998; Foley 1997; Elman et al. 1996; Armstrong, Stokoe & Wilcox 1995;). However, some researchers studying language acquisition/development have proposed that the type of mental representation formed by the mind is not symbolic in nature. One form of this theory is called "connectionism" (see Westermann, Ruh & Plunkett 2009 and Plunkett 1998, 1995 for summaries of connectionist research in language learning). In this school of thought, what develops during language learning is a distributed network of simple components (neurons) connected by rules that influence or weight the level of participation of any given element. The ensuing interactions give rise to "emergent" (qualitatively new) global behaviors. Workers in this area are interested in determining how much innate structure is needed to initiate the learning process. Learning, in the words of Seidenberg (1997: 1600), "involves gradual changes to the weights on connections between units that determine patterns of activation in the network." It is possible that the developing temporal bones create a bias toward particular types of information carried in the speech signal, one that, if it exists, could and should be added to connectionist accounts of language development.

Another recent theory is enactionism, which sets aside altogether the notion of cognition as a realist representation of a Kantian pregiven world in its claim that the mind does not form representations at all: "[C]ognition is … the enactment of a world and a mind on the basis of a history of the variety of actions that a being in the world performs," and "knowledge depends on being in a world that is inseparable from our bodies, our language, and our social history—in short from our *embodiment*" (Varela, Thompson & Rosch 1991: 9, 149). A related theory is embodied realism (described in Johnson & Lakoff 2002 and Lakoff & Johnson 1999 and critiqued in Rakova 2002), which does not necessarily embrace the claim that there is no "pregiven" world, but does propose that concepts do not exist independently of a thinking being's body, and that language is fundamentally spatial.

Inquiries into the way embodiment structures language have focused on the role of the visual system in creating universal "image-schemas[1]" which structure perceptions and linguistic expressions and which people use to reason (e.g., Fauconnier 1997; Langacker 1991, 1987; Lakoff 1990, 1987; Talmy 1988; Lakoff & Johnson 1980). Yet the sense of hearing and its role in structuring experience with schemas has by and large escaped notice. Given that no known hearing community has ever adopted a signed language as its primary

---

[1] As Deane [1991: 364] points out, image schemas are "embodied schemas" that function as "cognitive models of the body and its interaction with the environment."

language, linguists' explorations into the visual system's role in language can only be complemented by some consideration of the auditory system's contributions.

Psycholinguistic research over the last 40 years into the capabilities of infants and children to perceive and produce speech has figured prominently in the theoretical debates sketched above (see reviews in Fernald 2000; Jusczyk, Houston & Goodman 1998; Aslin, Pisoni & Jusczyk 1983). While researchers studying the development of phonetic categories generally accept that linguistic experience alters phonetic perception (Kuhl et al. 1992), they do disagree, predictably given the current theoretical milieu, as to the nature of the outcome of that interaction and the mechanisms that underlie it. Research over the last 20 years has shown that even the youngest infants use signal-complementary and top-down processes to listen to speech; that is, they use sources of information in addition to the signal itself, probably to identify words rather than linguistic phonetic units, and that early language learning has effects on language performance later in life (e.g., Marchman & Fernald 2008). Indeed, infants have more access to phonotactics[2] (e.g., Jusczyk et al. 1993) and prosodic structure (e.g., Myers et al. 1996; Jusczyk, Cutler & Redanz 1993) and attend more to regularities in the speech they hear than previously recognized (e.g., Brent & Cartwright 1996; Miller & Eimas 1996; Saffran, Aslin & Newport 1996; Fisher & Tokura 1995; Werker & Lalonde 1988). Speech perception is a complex process, ultimately relying on the ability of humans to integrate cues and information from more

---

[2] Phonotactic information is information about which sequences of sound are possible in a given language. For example, in English, words never end in "spr," although they can begin that way, and the "eng" nasal only occurs after some of the short vowels, not all of them.

than one sensory modality and many sources. Theories of speech perception also need to account for systematic sources of contextual variation in the signal such as coarticulation effects (the physical signal reflects the fact that the articulation of one sound affects the articulation of others near it), speaker effects (different vocal tracts produce physically different sounds), and style effects (speakers use different speaking styles in different communicative situations).

Faced with the complexity of the language learning task, some researchers have proposed innate mechanisms that function independent of experience with the ambient language(s) and that constrain the learning process, something like the highly specialized auditory system for echolocation in bats (Gould, Cooley & Barnick 1981; Gould 1975) or song in crickets (Bentley & Hoy 1996), currently thought to have genetic bases. Others researchers emphasize that infants and children are superior learners and explore the extent to which general cognitive mechanisms could be constraining the process of language development.

Theories of phonetic development constitute but a subset of the rich literature in the area of speech perception theory. While focusing on those theories that have developed explanations for how children learn phonetic systems, the next section in this chapter also touches on several recent theories of speech perception that are not focused on how perception develops in children in order to present something of the full breadth of thinking in the field.[3]

2.2 THEORIES OF THE DEVELOPMENT OF PHONETIC CATEGORIES

Eimas (1975) provides a statement of what researchers initially proposed as a explanation for phonetic development, that is, that the human genetic endowment includes innate feature detectors for all phonetic contrasts possible in human languages. The theory was that language input has the effect of throwing certain switches so that some phonetic contrasts are "selected" and some contrasts, the ones not used in the ambient language(s), are lost. As mentioned in Chapter 1, several behavioral studies have demonstrated that by 6–10 months, infants discriminate phonetic contrasts used in their native language(s) and not those used in other languages, where younger infants discriminate all phonetic contrasts (e.g., Best et al. 1995; Polka & Werker 1994; Werker & Tees 1984; Aslin et al. 1981). The results of these experiments were initially interpreted as evidence that infants were losing phonetic contrasts, implying that they start with a larger innate inventory from which they select the contrasts relevant in their native language(s).

In subsequent years, these results were reinterpreted by investigators working in other theoretical traditions. For example, Kuhl (2001, 1994, 1991) and Garcia-Sierra et al. (2011) now explains the phenomenon by theorizing that 6-month-old infants are building or constructing a language-specific map of the sound systems of their ambient language(s).

---

[3] Hawkins (1999c) and Massaro (1994) provide excellent reviews of the theories of speech perception, the organization of which the discussion below follows closely.

Kuhl (2000: 11852) points out that “to refute the selectionist position, studies must demonstrate that infants listening to ambient language are engaged in some other kind of process, a process that is not fundamentally subtractive in nature.” In other words, to refute the selectionist position inherent in this nativist explanation of phonetic development, researchers generally seek to demonstrate that infants are building something (whether that be representations or distributed networks or image-schemas) rather than selecting from among innately specified options.

Hawkins (1999b, c) classifies theories of speech perception according to the role they envision for “abstract referents.” That is, some theories “conceptualize the process of speech perception as one of filtering out an invariant ‘signal’ from irrelevant ‘noise’” (Hawkins 1999b: 199). Such theories posit the existence of an invariant core or reference point. This invariant core unites a group of phonetic signals that users of a given language report experiencing as the same sound across speakers and contexts (although there is disagreement about whether or not this invariant core is innately specified). Other theories suggest that there is “invariance in the percept but not the object” (Hawkins 1999c: 233), arguing that the reason phoneticians have not yet learned to identify a core constancy under the signal surface is that no such invariance exists.

Invariant signal. In her comparison of speech perception theories, Hawkins (1999b) designates two theories as classical: the motor theory of speech perception (e.g., Liberman 1996; Liberman & Mattingly 1985) and the quantal theory of speech (Stevens 2000, 1989, 1972). In motor theory, speech perception is viewed as “the product of a specialized module that recovers talkers’ intended articulatory gestures” (Kluender 1998:

26). Babies are born with these genetically endowed modules and a listener's extracting the abstract articulatory gesture that the speaker intended to make is an automatic process. Contrasting with motor theory and its search for invariant articulatory properties is the quantal theory of speech, which is a search for invariant acoustic or auditory properties that can be linked to linguistic units; in this way perception remains stable across speakers and phonetic contexts. Quantal theorists do not subscribe to the notion of the specialized module that extracts articulatory gestures coded in the speech signal as postulated in motor theory—rather they argue that suitable candidates for speech sounds are those for which the properties of the signal are relatively resistant to variations in the way people articulate them, and that languages seek out such regions of stability. General auditory mechanisms perceive acoustic patterns that are there because of consistencies in articulatory gestures. Though quantal theorists do not explicitly address the issue of child language development, infants and children with normal hearing presumably learn the phonetic categories inherent in the ambient language(s), which have exploited several naturally occurring regions of stability.

Both these theories emphasize that listeners recover the speaker's articulations, directly in motor theory or indirectly in quantal theory. Motor theory would not predict that the developing temporal bones would simplify the speech signal because the basis of speech is not the sounds of speech; the basis of speech is the intended articulatory gesture. If such auditory simplification exists, it could not be of any consequence; the basic units important in speech perception (i.e., the recovered intended motoric representations) are entirely separate from the signal. Neither would quantal theory predict acoustic regularities that did not stem from articulatory gestures.

Invariant percept. The invariant percept theories, more recently developed than the classical theories, deemphasize or reject the notion that there is anything invariant about the speech signal itself. In general, these theories do not posit the existence of innate mechanisms. As in general linguistic theory, however, researchers working in these different theoretical paradigms disagree about what form of mental representation people learn, if any. The discussion below sketches several of these theories—native language magnet theory, hyper/hypo theory (H&H theory), auditory enhancement theories, continuous information theories, and direct realism.

Native language magnet theory. Native language magnet theory has an empiricist/constructivist orientation. In native language magnet theory, Kuhl (2001, 1994, 1991) and colleagues propose that what people learn is a map of phonetic categories. This map is initially based on auditory boundaries that cause listeners to hear sounds categorically and that occur naturally in humans and other vertebrates such as chinchillas (Kuhl & Miller 1978, 1975), macaques (Kuhl 1991; Kuhl & Padden 1983), and Japanese quail (Kluender et al. 1998). During development, a phonetic “prototype” (in the sense developed by Rosch & Mervis 1975) perceptually pulls other members of its category toward it, meaning that a listener does not perceive differences between a prototype and other instances that acoustically are not prototypes. The signal is variable but the percept is not because of the warping caused by prototype magnets. The formation of phonetic prototypes is based on best instances of a sound category. Evidence suggesting that some process of categorization is at least available for listeners to use includes studies that show people can identify one speech sound as being more prototypical than another (Iverson & Kuhl 2000; Grieser & Kuhl 1989), although these results do not necessarily mean that all listeners in all language communities always categorize individual speech sounds to process speech.

Researchers working within this theoretical stance have shown a great deal of interest in the development of language in infants and children. For instance, several studies have explored the role of "baby talk" in providing the best instances of a sound category so that children can develop their conceptions of prototypical speech sounds (Fernald 2000 and Kuhl et al. 1997 include up-to-date sketches of work in this area). The conclusion to be drawn from the work of several ethnographers, however, is that cultures vary in their use of infant- and child-directed speech (e.g., Heath 1983 on southern rural African-Americans; Ochs 1988 on western Samoans; Pye 1992 on the K'iche' Maya community; Schieffelin 1990 on the Kaluli of Papua New Guinea; Smith-Hefner 1988 on the Javanese). Additionally, infant-directed speech has not been proved to provide best instances of categories. Kuhl and colleagues (1997) analyzed acoustic data from U.S. English, Russian, and Swedish speech directed at infants ranging in age from 2 to 5 months. They found that mothers addressing their infants produced words with instances of /i/, /a/, and /u/ that were more distinct from each other than instances of the same vowels in the same words uttered by the same women in speech directed at adults. Nevertheless, they did not find a greater frequency of best instances or less variability in the tokens present in infant-directed speech as compared with adult-directed speech.

Arguing that prototypes should be directly observable at the level of the acoustic signal and that the existence of the perceptual magnet effect depends on there being some sort of heightened or enhanced shape for the linguistic input, Davis and Lindblom (2001) use an acoustic analysis of their data and conclude that while baby talk is possibly less variable than adult-directed speech, baby talk is indeed variable and that prototypical tokens of the vowels they studied (the ones commonly described as "long e"—/i/—and "short i"—/ɪ/—in English) formed a very small subset of the data. In another, earlier study, Lindblom and colleagues (1992: 365) speculate that the acoustic variability they found in baby talk, adult-directed spontaneous speech, citation speech, and clear speech "might be considerably reduced if the measurements were translated into more perceptually relevant dimensions." In other words, they believe that, while an acoustic analysis of the data might reduce nonsystematic (random) variability, infant-directed speech cannot be the only mechanism responsible for providing the best instances around which phonetic categories are presumably built in native language magnet theory.

H&H theory. On the basis on such studies as those described above and in order to explain the lack of phonetic invariance, Lindblom and colleagues (e.g., Lindblom 2000; Lindblom et al. 1992; Lindblom 1990) advanced the hyper/hypo theory, called H&H theory, which rejects the notion of a prototype at the center of a category:

> Speakers can, and typically do, tune their performance according to communicative and situational demands, controlling the interplay between production-oriented factors on the one hand, and output-oriented constraints on the other. For the ideal speaker, H&H claims that such adaptations reflect his tacit awareness of the listener's access to sources of information independent of the signal and his judgment of the short-term demands for

> explicit signal information. Hence speakers are expected to vary their output along a continuum of hyper- and hypospeech. (Lindblom 1990: 403)

Like native language magnet theory, H&H theory believes that language learners create a mental representation in their interaction with speakers. In native language magnet theory, however, a learner builds a collection of discrete phonetic categories; in H&H theory a learner maps a constellation of distinctive (not invariant) reference points.

Adaptive Dispersion Theory. Adaptive dispersion theory grows out of H&H theory in that it posits that distinctive sounds of a language tend to be positioned in phonetic space so as to maximize perceptual contrast (Flemming 1996; Johnson, Flemming & Wright 1993; Lindblom 1990; Lindblom and Engstrand 1989) One working hypothesis of dispersion theory is that “speakers develop a ‘feel’ for the ‘survival value’[4] of phonetic forms through a process not unlike natural selection” (Lindblom 1990: 405). But these phonetic forms have no structure individually. They are something like electrons and quarks in the quantum field theory of physics, that is, fundamental structureless points that interact according to rules that hold them together in a configuration or shape. Adaptive dispersion theory attributes the emergence of discrete entities in the speech signal to the ability of language learners to compare the speech and meanings of those speaking around them. Learners then use their experience of speech sounds across many words and different speakers to build in their memory a map of the locations in which these recurring sounds

---

[4] Ladefoged (1956) defines a sound’s phonetic value as the linguistic and sociolinguistic information present in the sound taken together, that is, information that is not the result of idiosyncratic features such as those due to the shape of an individual’s vocal tract. Lindblom’s “survival value” seems to refer to the linguistic component alone of a speech sound.

pool relative to one another. These pools of sound interact with each other according to rules, and speakers also learn the principled ways in which the map can be shifted or rotated as a whole. Map shifts are adaptive in that they signal important social information, such as regional dialects and social distance between interlocutors. What children are mapping is an entire system that emerges from phonetic content. The elements of that system are learned in relation to one another, something like a musical scale or a map; with experience, a listener's map can shift, expand, or contract given the various functional aspects of the speaker's social or communicative situation.

Several researchers have noted that approximately 15% of the world's known languages do not conform to the predictions of dispersion theory, including McDonough (1992, 1994) who observes that Diné is one of these exceptions. Indeed, Maddieson (2003: 2331) proposes that "the overall maximal separation of vowels in the acoustic space is therefore not the major organizing principle of these [four-] vowel systems." He suggests that instead such vowel systems' configuration "is anchored by the polarity of /i/ vs /a/."

Auditory enhancement theories. As a rule, auditory enhancement theories argue it is "little more than sensitivity to statistical regularities of language input together with organizational processes that serve to enhance distinctiveness of regions in that input" that allows people to learn the sound system of a language (Kluender et al. 1998: 3580). These groups of experimenters (Diehl, Kluender & Walsh 1990 provide a survey) see perceptual needs as determining articulatory patterns, but they do not assume acoustic invariance. The speech signal is redundant, yet the redundancy is structured. Acoustic properties combine to form intermediate perceptual properties, which combine to form distinctive features that define phonemes—in this way, these researchers also assume that people form mental representations, albeit in a different way and form than those in native language magnet theory or in H&H theory. They have concentrated on studying the perception of certain speech sounds that they believe cannot be explained with a motoric or gestural theory (Hawkins 1999c), such as vowels (e.g., Diehl 2000). On the whole, auditory enhancement theorists have not as yet considered language development in great detail. The “organizational processes” to which Kluender and colleagues (1998) refer could presumably incorporate a role for social experience and interaction in focusing the learner’s attention on linguistic elements.

Continuous information theories. As a group, continuous information theorists do not believe people form symbolic mental representations (see Massaro 1994 for a summary of this work). As Hawkins (1999c: 266) puts it, “they build on a basic assumption that all information is potentially salient and allows information about a particular feature or phone to vary in clarity” in a network of activated connections, and they focus on how decisions are made in word recognition rather than on the nature of the input.

The point of departure for this group of theorists is that the goal of speech perception is to identify words, not individual speech sounds. Speech is far more complex than an assemblage of sound blocks. For example, Peterson, Wang & Sivertsen (1958) segmented a stream of speech from one language and then recombined those bits into new words in another language, with disastrous results in terms of producing recognizable words**.** Klatt (e.g., 1979) has suggested that entire words are stored as spectral patterns rather than individual segments; he envisioned a process for word recognition in which the acoustic signal is compared against those stored patterns and a best match determined. Continuous information theories generally incorporate this best match process. Of the models produced by continuous information theories, the connectionist TRACE model (e.g., McClelland & Elman 1986) incorporates an interest in language learning, although at this time the researchers in this area do not investigate child language learning as a different phenomenon from adult learning.

One of the interesting findings of such research is that these networks can recognize words and arrive at linguistic categories by integrating disparate sources of information over time without early segmentation of the signal into individual speech sounds. But what these networks model is how word recognition could happen, perhaps for people who do not have access to segmentation of speech sounds, for example, the deaf or the hard of hearing or, as recent research shows, apparently dyslexics (Gerrits & deBree 2009; Munson, Edwards & Beckman 2005; Elliott & Hammer 1993).

Direct realism. As in the classical theories of speech perception, direct realists see articulatory (vocal tract) gestures as the basic commutable units of speech (e.g., Fowler 1994, 1986). As in enactionism, direct realists insist we know the world "'directly' as an immediate experience of the actual object in the environment" (Hawkins 1999c: 233). Direct realists do not believe that listeners construct a mental representation of that immediate experience.

In a direct realist account of speech perception, learning to understand speech is learning about the behavior of vocal tracts. What separates direct realists from motor theorists is the belief that speech perception is another aspect of perception in general, not the result of an innate specialized mechanism and that what listeners recover from speech is the actual articulatory gesture, not an abstract "intended" gesture. "Moreover," as Hawkins (1999c: 235) states, "unlike the motor theory, direct realism allows that invariance could be in the acoustic signal as well as in the percept, but is no more interested in acoustic invariance than in acoustic variability because the acoustic signal is not what is directly perceived." During the time infants and children are learning their native language(s), they are becoming attuned (in the sense promulgated by Gibson 1971, 1950) to the complex coordination of simple gestures, which leads to higher-order invariants that ignore the details of the acoustic/gestural signal. Despite the variable input, certain abstract patterns are preserved. As Hawkins (1999c: 235) points out, "In a sense, this describes a process of learning to be efficient at picking out the critical parts of the information flow."

Finding training effects in the way youngsters experience their vowel systems over the years of childhood will not prove that people form mental representations to some theorists because the project will only describe changes in the form of input available to people at different ages, not if people actually use any such training effects. Native language magnet theory, H&H theory, adaptive dispersion theory, and auditory enhancement theories (which assume people do form mental representations) would predict that if training effects exist that they would have a role in shaping the prototypes, maps of sound constellations, or distinctive regions posited in their respective theories. Continuous information theories could possibly find a role for effects caused by the development of the temporal bone in the form of weights in their networks of activated connections. Direct realists would not predict any **functional** role for training effects caused by effects on hearing.

The goal of this dissertation is to study any patterns of differences in the ways infants, children, and adults experience language in order to identify possible training effects. The project thus proceeds on the assumption that people do form mental representations of the sound systems of their native language(s) in childhood that may shape their perceptions as adults. Though the project proposes to compare auditory analyses (i.e., perceptually relevant analyses based on the way infants, children, and adults hear) of vowels to see if the temporal bones are making different types of information more prominent than others during the years of language acquisition/development, the project also assumes the activity of many other organizing processes during speech perception (e.g., the importance of formant transitions discovered by Cooper et al. 1952 or the role of dynamic spectral variation in vowels first studied by Liberman et al. 1967). These organizing processes range from language-specific knowledge about prosody and rhythm that helps identify words in the speech stream to culture-specific devices for directing the attention of young learners to what constitutes communication and what is important in the event, as well as multimodal sources of information such as knowledge about vocal tract behavior and visual organization. It is entirely possible that any training effects identified in this research project are not absolutely necessary to speech perception in modern-day humans.

Notwithstanding the complexity of speech perception in modern-day languages, however, any effects on speech sounds caused by the temporal bones remain interesting in that they would have evolved in connection with behavioral and biological adaptations to certain strategies thought to be important in human evolution. In an evolutionary sense, the morphology of human temporal bones and their pattern of development (described in the next section) result from bipedal hominids committing to a certain nutrient-rich diet that required more effort and training to procure, growing big-brained infants whom they provisioned for increasingly extended periods of time, and experiencing benefits from creative and flexible behaviors. So a better understanding of the influence of the temporal bones on the perception of sound (which would include species-typical vocalizations) may provide clues about the precursors of language, supplementing other work that has been done on the evolution of speech production. See, for example, Lieberman et al. 1992 on the reconstruction of hominid vocal tracts from fossil cranium; Kay, Cartmill & Balow 1998 on the reconstruction of hominid tongues from the hypoglossal canal located in the skull's occipital bone of the skull; Degusta, Gilbert & Turner 1999 on a refutation of Kay et al. 1998; and Enard et al. 2002 on the role of the FOXP2 gene in the evolution of human linguistic abilities.

3

THE HUMAN TEMPORAL BONES AND BRAINS

Arguing that thinking and behavior are artifacts of an entire body (i.e., that the brain has a body), a number of cognitive scientists have emphasized the interaction of the nervous system, the body, and the environment (e.g., Barrett 2011 is a thorough review; see also Chiel & Beer 1991; Clark 2008; Rowlands 2010):

> …[A]daptive behavior also depends on interactions among the nervous system, body and environment: sensory preprocessing and motor post-processing filter inputs to and outputs from the nervous system; co-evolution and co-development of nervous system and periphery create matching and complementarity between them; body structure creates constraints and opportunities for neural control; and continuous feedback between nervous system, body and environment are essential for normal behavior. Chiel & Beer (1991: 553)

This point of view is a challenge to interpretations that assume complex, flexible behaviors like language are "mere by-products of internal cognitive activity" (Farina 2012). This dissertation undertakes to study "sensory preprocessing" that filters sounds delivered through the temporal bones to the nervous system at different ages.

Temporal bones are skull bones that house the bony portion of the external auditory meatus (the ear canal), the middle ear, and the inner ear (see Figures 3-1 through 3-5). The primary function of the mammalian external ear, which includes the pinna and the external auditory meatus, is to collect acoustic energy and conduct it to the tympanic membrane at the entrance of the middle ear (e.g., Perkins & Kent 1986; Yost 1994). As Pickles (1988) points out, the external ear has two kinds of effect on incoming sound: (1) directional effects important in sound localization, and (2) resonance effects on the sound pressure at the tympanic membrane.

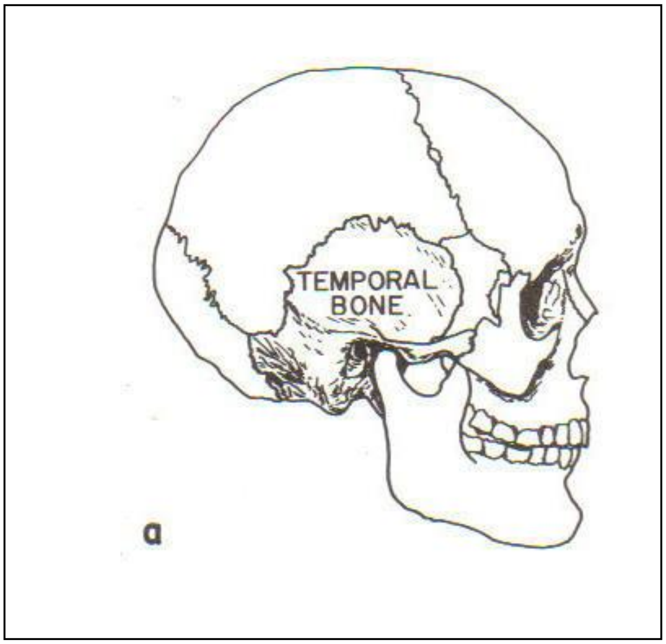


Figure 3-1. Located at the side and base of the skull, the temporal bone is a dense, complex bone (from Durrant & Lovrinic 1977: 106).

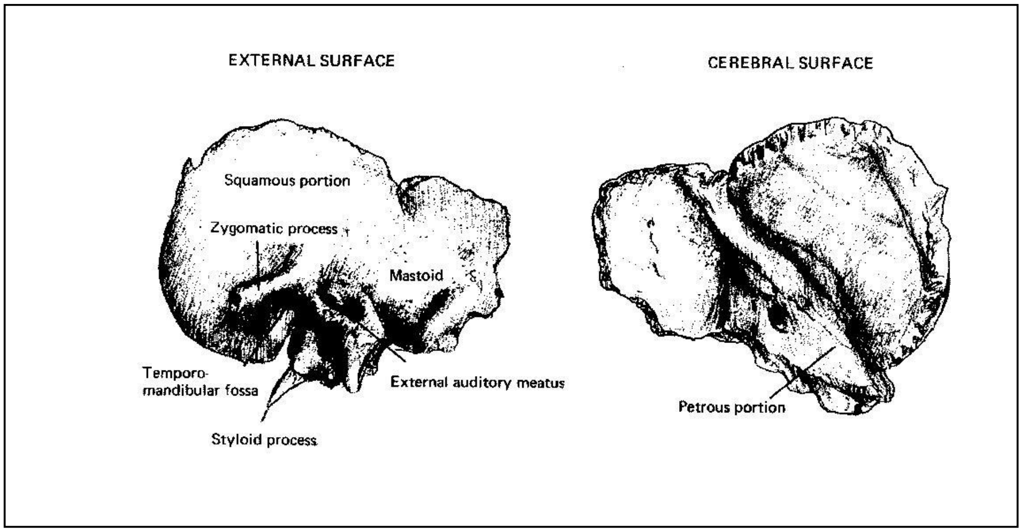


Figure 3-2. The temporal bone has three main parts: the squamous, the mastoid, and the petrous (from Bass 1995: 48), each part having its own trajectory for postnatal development. The tympanic also has its own independent course of postnatal development, discussed in more detail below. This is a left temporal; the external view is on the left.

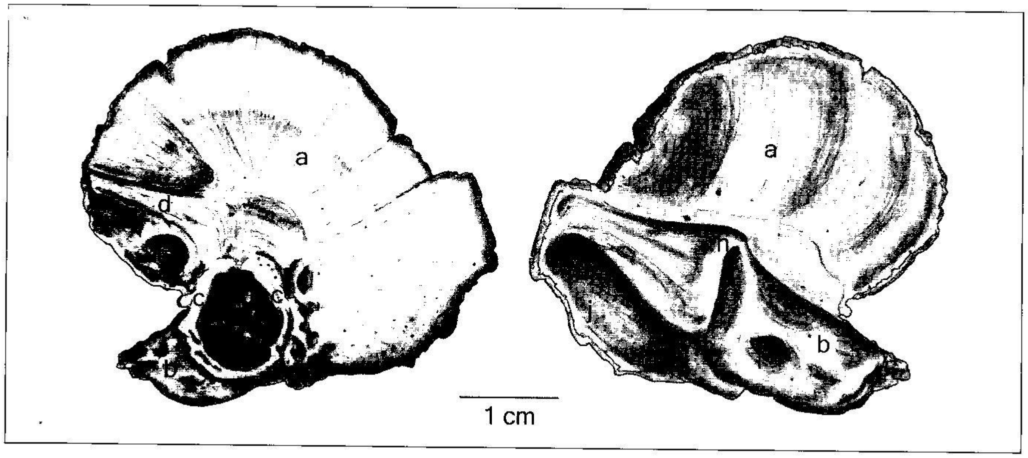


Figure 3-3. Left Infant Temporal Bone (from Baker, Dupras & Tocheri 2005: 37). The view on the left is the external view. "A" is the squama; "B is the petrous portion.

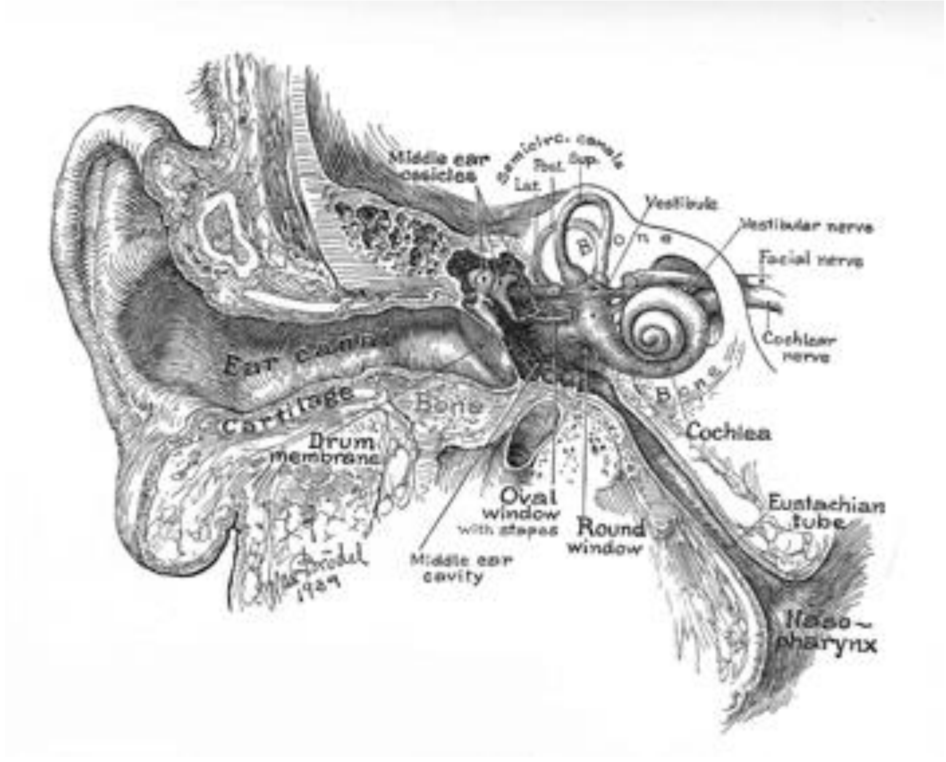


Figure 3-4. The outer, middle, and inner ear constitute the peripheral auditory system (Brödel 1939)—the labeled "bone" is the temporal bone.

Figure 3-5. This figure (from Proctor 1989: 20) is a frontal section of an adult left temporal bone in which one is looking from the back of the head in the direction of the face. The overlapping bones that form the roof of the middle ear cavity are the squamous and petrous parts of the temporal bone. Petrosal bone, which is very dense and hard, overlaps the squama and, surrounding the organs of hearing, forms the floor of the middle ear cavity as well. The petrosal encasement of the middle ear is considered a diagnostic characteristic of the Primate order (e.g., Szalay & Delson 1979). Note the location of the cavity that houses the jugular (labeled 6) in the lower right corner of the figure, separated from the middle ear by petrosal bone in adults.

Humans are born with immature skull bones (Scheuer and Black 2000; Williams et al. 1982) which allow the skull to be compressed slightly as the infant rotates and passes through the birth canal (Leutenegger 1982; Lindgren 1977) and which permit the considerable postnatal growth of the brain that will occur (e.g., Bogin 1999; Sinclair 1973; Tanner 1989), as Figure 3-6 illustrates.

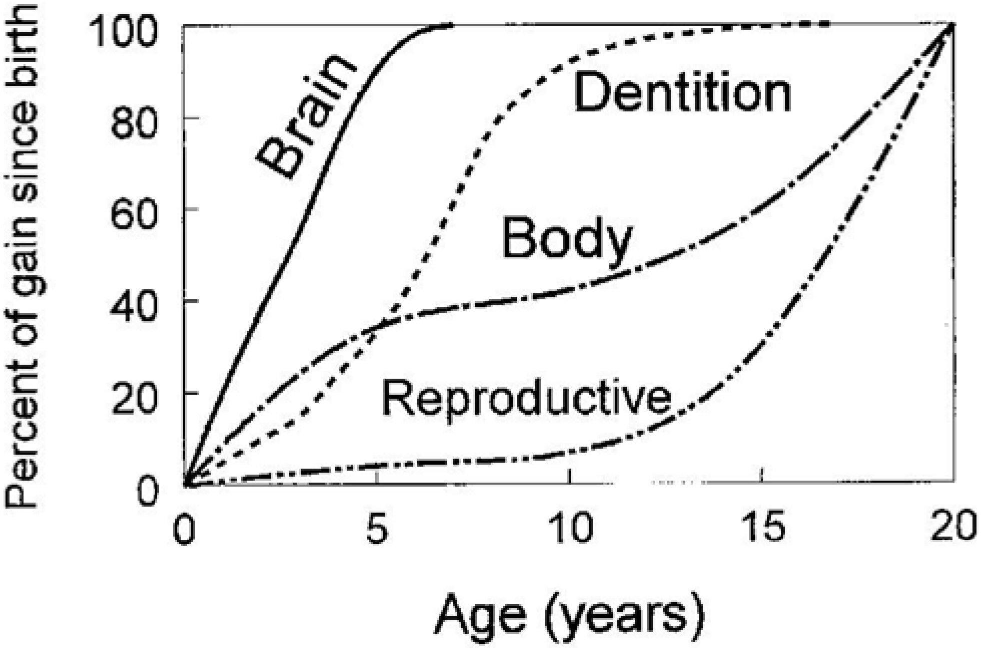


Figure 3-6. Human Growth Curves for Body Tissues Showing the Human Brain Grows More Rapidly in Childhood Than Any Other Tissue (from Bogin 1999: 73)

Falk and colleagues (2012) examined an endocast of Taung child (a 3–4-year-old specimen of *Australopithecus africanus*) based on a reconstruction composed by Falk and Clarke (2007) and compared the metopic suture (see Figure 3-7) to the relevant frontal bone morphology of fetal to adult chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus)*, and modern humans (*Homo sapiens*). An unfused metopic suture is part of the immature skull's accommodation to a trip through the bipedal pelvis, though other selection pressures such as post-natal growth and reorganization of the frontal cortex are likely mutually non-exclusive aspects of perinatal ontogeny (Falk et al. 2012).

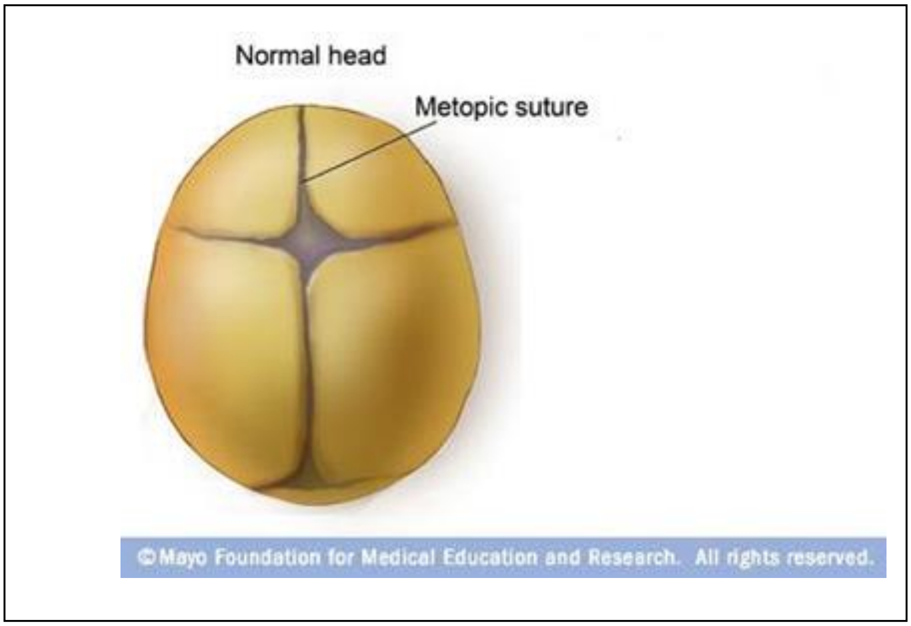


Figure 3-7. Illustration of Metopic Suture in Modern Human Infants (cropped from a MayoClinic.com illustration)

The results from Falk and colleagues (2012) indicate that, unlike the metopic sutures in monkeys and *Pan* that are fused at the time of birth, that Taung child had an unfused metopic suture (MS), "thus taken as evidence that a human-like pattern of late MS fusion was already present in mid-to-late Pliocene gracile hominins" and that "the late fusion of MS may have become adaptive relatively early during hominin evolution" (Falk et al. 2012: 8469). Taung child is dated at 2.5 MYA; *A. africanus* was an obligate bipedalist.

In keeping with Michel and Tyler's (2005: 157) observation quoted in Chapter 1 of this dissertation that "comparisons of the patterns of development among closely related species revealed that the origin of morphological differences, with important adaptive consequences, derived from differences in the timing of typical sequences of developmental events," Leigh (2004: 139) concludes from his comparison of brain growth in several species of primate that "primates are characterized by significant variation in patterns of brain growth. In addition, the degree to which brain growth is allocated to either the pre- or the postnatal period varies substantially." Leigh (2012, 2004) goes on to point out that both chimpanzees and humans experience about the same length of duration for post-natal brain growth (until age 5–6 years; see Figure 3-8). However, humans experience a much greater **volume** of postnatal brain growth in the neocortex. This growth principally takes place in two areas of the neocortex, one being in the parietal association areas which are involved in sensory integration, and the other being in the prefrontal lobes, important in long-term planning and creativity (Byrne 1995; Deacon 1997; Gibson 1996). As the neocortex expanded with larger brain sizes in different primate species, the internal ear remained connected to the external ear by the formation of a tunnel. Bramblett (2004) states that in Platyrrhines "this tube is enclosed by a cartilaginous structure that terminates in an ectotympanic ring that anchors the eardrum. In Catarrhines an ossified ectotympanic tube [the external auditory meatus] connects the eardrum to the outside."

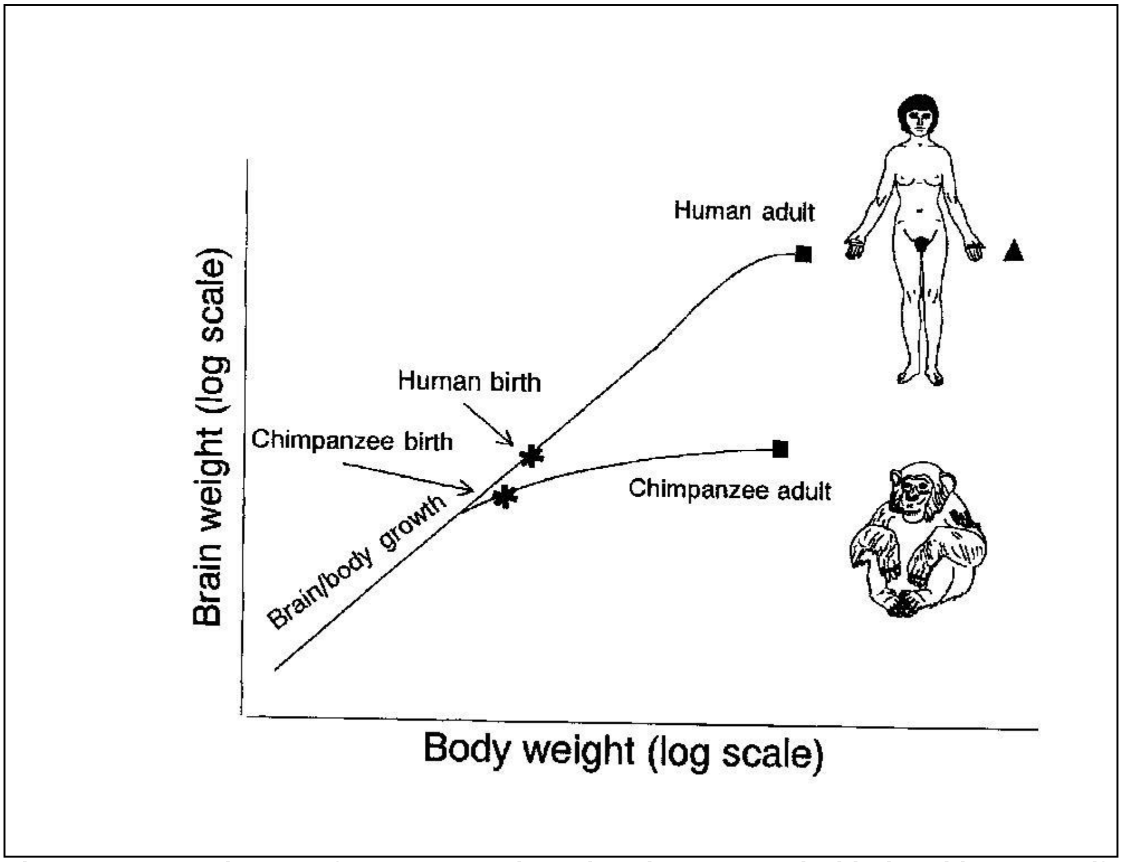


Figure 3-8. Growth Curve for Human Brain and Body Compared with the Chimpanzee (from Bogin 1999: 184)

In Leigh's (2012) judgment, brain growth in the earliest hominins, including Taung child, resembles the general **pattern** of chimpanzee brain growth. Bastir and Rosas (2004) comment that modern humans (*H. sapiens*) do follow a different pattern from chimpanzees, in that they have higher prenatal rates of skull growth and lower peri- and postnatal maturation rates in the skull as compared with chimps. Schultz (1941, 1940) provides another perspective, observing that greater than 40% of brain growth has occurred in all nonhuman apes at birth, where 23% has occurred in *Homo* at birth. Schultz further found that orangutans have completed 90% of their brain growth at age 2 years and 97% at age 6 years, where humans reach 90% of brain growth at age 6 years and 95% at age 10 years, and he concludes that the pattern of brain growth in orangutans and chimpanzees is more similar to each than either of these two apes is to humans. Penin, Berg & Baylac (2002) conclude that, although chimpanzees and humans reach the same overall size of brain and face as adults, human growth accelerates through the eruption of the first molar but then decelerates as compared to that of chimpanzee growth. Faulk and colleagues (2012) note that the metopic suture in non-human great apes closes shortly after birth, but considerably later in humans, persisting into adulthood in about 3–4% of modern humans. Schultz (1941, 1940) observes that nearly all sutures are closed at or shortly after birth in nonhuman apes.

It seems reasonable to expect that the general patterns described above for the ossification and maturation of the skull across different species of primate would hold for the temporal bones as well, but the ontogeny of the temporal bone in nonhuman primates and its perinatal form have not been well studied to date. One study providing some data in this area is provided by Torre, Giacobini & Ardito (1978) who compared development of premature newborns of *Pongo* and *Homo* (both at 7 months' gestational age). They found that the petro-mastoid region has "attained a more advanced degree of ossification in *Pongo;* semicircular canals are more evident" (Torre, Giacobini & Ardito 1978: 145). Inspection of the radiographs in Plate 3 of the article reveals a greater preponderance of dense bone in the temporal region stretching back to the much more developed occipital in the *Pongo* fetus as compared with the temporal region of the *Homo* fetus. Torre, Giabcobini & Ardito (1978: 147) conclude that while the *Pongo* and *Homo* fetuses have similar degrees of ossification, there are "differences in the maturity of some skeletal districts" which may be discussed "on the basis of different specializations reached by hominid and pongid evolutionary lines." They note that orangutan newborns are known to be able to cling to their mother's body from the very first days of postnatal life and state that the ability to cling even during the mother's brachiation implies a "high maturity of the system controlling muscular tone and synergy" and a good fixity of the head, which is allowed by highly developed nuchal and spino-appendicular muscles that insert on the occipital. Torre, Giabcobini & Ardito (1978) further speculate that the marked development of the semicircular canals in *Pongo* indexes advanced development of proprioceptive systems. These are interesting observations because they suggest that features of the temporal bone's ontogeny have been subject to selection pressures from behavioral and biological responses to the environment as would be predicted by, for example, Chiel and Beer (1991).

It is well established that modern human temporal bones do not approximate their mature form until age 5–7 years (see Dahm, Shepherd & Clark 1993 for an extensive study of the postnatal growth of the temporal bone). Nonetheless, there is no difference in the size of the inner ear between adults and newborns (Hoyte 1997), which fits with the neurological evidence that postural orientation and vestibular function are mature at birth (e.g., Konner 1991; see Figure 1-3 in this dissertation) and with the psychoacoustic and anatomic evidence that, while there are neuronal and conductive immaturities in the auditory system, the cochlea itself seems to be mature at birth (Bredberg 1968; Pujol & Lavigne-Rebillard 1985; Werner, Fay & Popper 2012; Werner & Marean 1996;). Furthermore, as reported by Dahm, Shepherd & Clark (1993), middle ears are adult-sized at birth; what increases is the size of the pneumatized mastoid. The explanation for this pattern of development probably lies in the pattern and types of ossification for the temporal: the petrous portion of the temporal bone forms in cartilage prior to ossification; the squamous and tympanic portions develop intramembranously (Baker, Dupras & Tocheri 2005).

Despite the maturity in size of the middle ear, the conductive apparatus of the middle ear is not mature. The neural immaturities that exist at birth[1] do not negate the effect of conductive immaturities (Werner & Marean 1996; Werner, Fay & Popper 2012). In terms of absolute sensitivity, which is the most common way to assess an auditory system's functional status, "there is little evidence to suggest that improvement in responses arising peripheral to the auditory brainstem is not completely accounted for by maturation of the conductive apparatus" in humans during the postnatal period (Werner & Marean 1996: 80). The neural immaturities that underlie wider auditory filters (causing poorer frequency tuning) in 3-month-olds as compared with older children and adults affect frequencies higher than 4 kHz (i.e., mostly higher than the frequencies important in speech); infants have adultlike auditory filters at 1 kHz.

During the time that the central auditory system matures, so does the conductive apparatus (the outer and middle ear) of the peripheral auditory system. In the middle ear, between birth and ages 5–7 years, sutures in the roof of the middle ear cavity ossify, the surrounding bone grows denser, and the malleus (the ossicle in contact with the tympanic

---

[1] Auditory filter width, which is important in frequency resolution, appears to be affected by neural immaturities until postnatal age 6 months. At 3 months, auditory filter width is adultlike at frequencies below 4 kHz (Spetner & Olsho 1990). The auditory filter width is the range of frequencies over which the auditory system integrates energy; only noise that falls within the filter width will mask the signal. The functional consequence is, as Werner and Marean (1996: 91) note, "the narrower the auditory filter width, the lower the signal-to-noise ratio will be at threshold." Neural immaturities clearly affect temporal processing, which is not mature until age 6 years (when the skull itself has nearly reached its mature size), and possibly intensity resolution (Werner & Marean 1996 summarize this research).

membrane) moves into its adult position. The most spectacular growth in the temporal bone, though, happens laterally to the tympanic membrane in the external auditory meatus, growing from about 8 mm in length at birth to 25 mm in adults. The pinna grows larger; the external auditory meatus lengthens, ossifies, and increases in diameter; and the tympanic membrane, which is supported by the tympanic ring, moves from a 15°–25° angle in the horizontal plane to an 80° angle (Anson & Donaldson 1981; Ballachanda 1995; Proctor 1989; Saunders, Kaltenbach & Relkin 1983; Williams et al. 1982). The inner and middle ears maintain in essence their same position in the skull from birth and the skull grows laterally to support the growing brain.

At around 1 year of age the human tympanic ring begins to extend posterolaterally to become cylindrical, growing into a fibrocartilaginous tympanic plate that will ossify around age 5 years (Anson & Donaldson 1981). This growth lengthens the external auditory meatus, changing the resonant properties of the canal. The foramen of Huschke forms in all children when the two tuberosities on the tympanic ring grow posterolaterally and then touch, leaving an opening in the tympanic plate (see Figure 3-6). The foramen of Huschke develops in the floor of the external auditory meatus inferior to the tympanic membrane by age 1 year, filling in with bone sometime between ages 3 and 5 years in most individuals (see Figures 3-6 through 3-9).

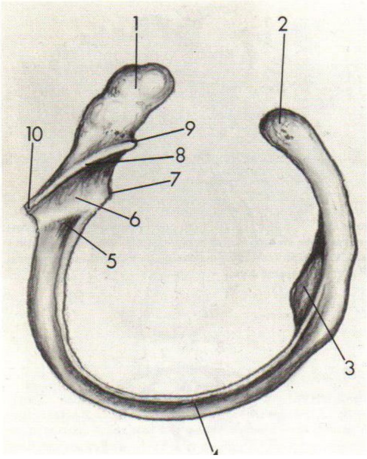


Figure 3-6. At birth the tympanic portion is an incomplete ring (Proctor 1989: 6). The tuberosities that grow laterally to form the external auditory meatus are labeled 3 and 5. The anterior and posterior tympanic tubercles grow towards each other across the ring and fuse together, forming the foramen of Huschke, posterior and medial to the external auditory meatus.

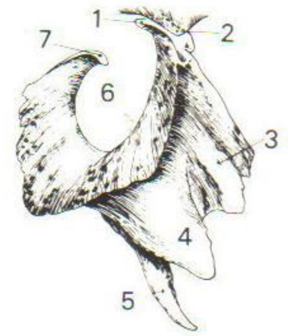


Figure 3-7. Adult Tympanic Ring. Labels 1 and 7 indicate the anterior and posterior tympanic crest that articulate with the squama to form the external auditory meatus (labeled 6).

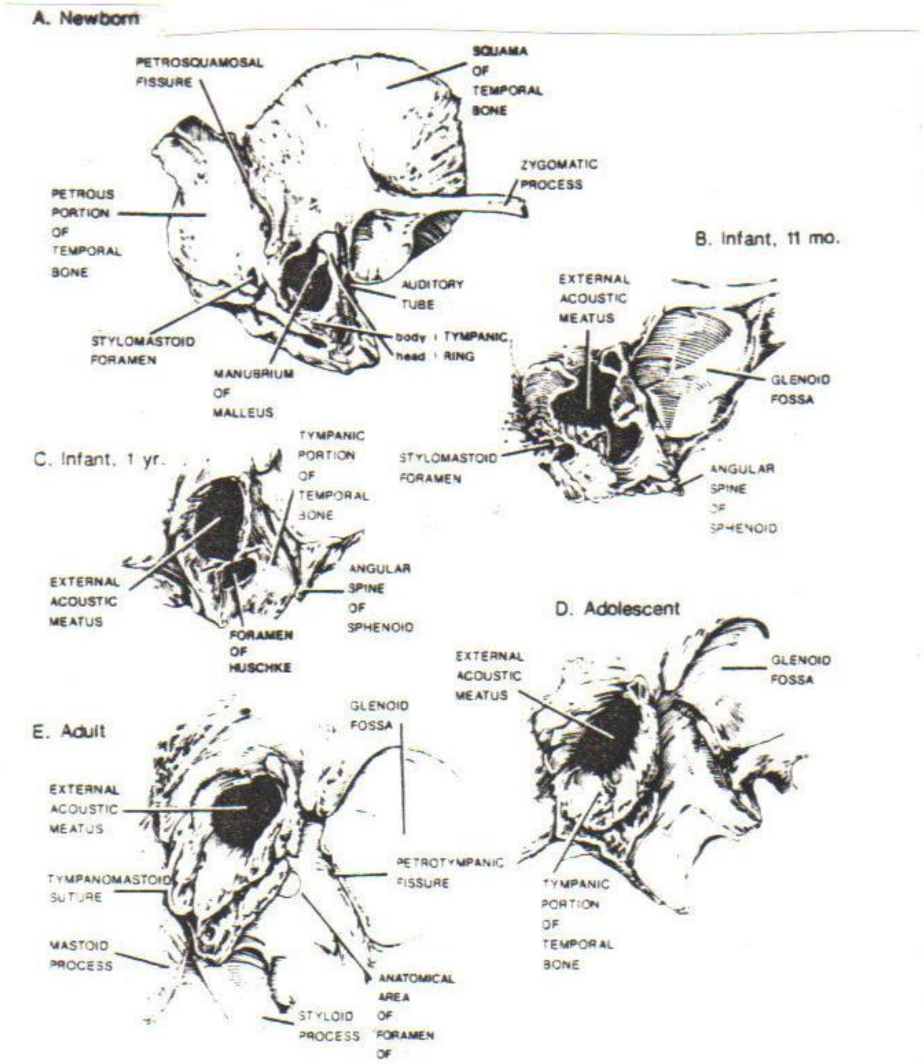


Figure 3-8. The tympanic portion of the temporal bone develops between birth and adulthood (illustration from Herzog & Fiese 1989: 268). The foramen of Huschke begins to form around age 10–11 months and ossifies around age 5 years. The length of the external auditory meatus increases until adolescence.

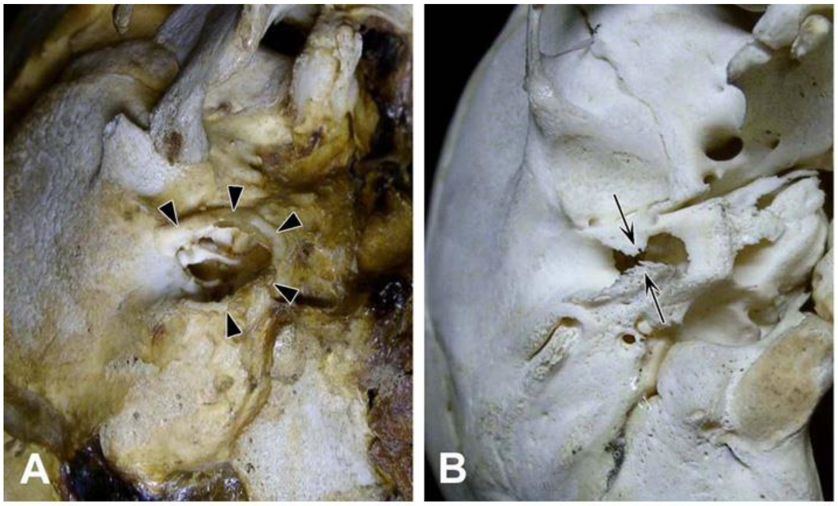


Figure 3-9. In these two photos from Hashimoto, Ojiri & Kawai (2011), A shows a newborn's tympanic bone (note how shallow the external auditory meatus is), and B shows the tuberosities on the tympanic ring beginning to grow towards each other in a 10-month-old.

Before the tympanic plate ossifies to fill in the foramen of Huschke, the foramen is covered with a fibrous membrane. The foramen of Huschke persists in some percentage of adults, depending on population[2] occurring bilaterally and unilaterally. Humphrey and Scheurer (2006) identified three types of persistent foramen of Huschke: single (the foramen comprises a single perforated area larger than 1mm in at least one dimension); cribriform (the foramen is divided by one or more bone bridges, occasionally forming a web-like structure and resulting in two or more separate areas of perforated bone); and trace (a small pinhole-sized foramen).

---

[2] Hashimoto, Ujiri & Kawai (2011) counted foramen of Huschke in 1994 temporal bones from 997 Japanese skulls at a range of ages. One of their findings was that the foramen of Huscke persists in 12% of males and 20% of females. Lacout et al. (2005) found an incidence of 4.6% in 130 ears of a French population. Wang et al. (1991) studied 377 Chinese skulls and found an incidence of 7%. Humphrey & Scheurer (2006) studied samples from Spitalfields (Christ Church) and Abingdon (St. Bride's Church) and found that foramina of Huschke persist past age 5 much more frequently than commonly thought and that males tend to have a foramen of Huschke unilaterally if they have one persist into adulthood. They found persistent foramen of Huschke in adults in 15.8% in Spitalfields and 23.3% in Abingdon.

The mechanical properties of the cartilage covering the foramen of Huschke have not been studied, but Heffez, Anderson & Mafee (1989) report that on CT scans of closed and open mouths a bulge and depression can be seen in external auditory meatuses where the axial view revealed a patent foramen, suggesting some pliability. Because the covering membrane is not rigid, the configuration of the foramen forming in the external auditory meatus floor near the tympanic membrane suggests the possibility that the foramen might be functioning as a notch filter. Often used in air conditioning ducts to remove an annoying resonant frequency, a notch filter screens out frequencies within a certain range, passing all frequencies except those in a stop band around a center frequency. As the foramen of Huschke fills in, a different band of frequencies would be filtered out, varying with its size.

Dempster and Mackenzie (1990) studied 250 children from ages 3 to 12 years and established an oft-cited figure of 3089 Hz as a mean resonance for the external auditory meatus of 4-year-olds (3009 Hz for 15 children who were under the age of 4 years), which gradually decreases until age 7 when the mean resonance declines to an adultlike 2700 Hz; they did not note any effects that would suggest the foramen of Huschke affects resonant frequencies in any way. Nonetheless, their study was critiqued by Bentler (1991) who pointed out that, because of the way the microphone must be placed in children’s ears, what Dempster and Mackenzie measured was actually external ear resonance effects and not external auditory meatus resonance effects. Bentler implicates canal “geometry” as another variable affecting resonant frequency, and Dempster and Mackenzie in their reply to her critique acknowledged that there was considerable variation in resonant frequencies at younger ages.

In order to consider whether the foramen of Huschke might be functioning as a notch filter, this investigator measured pure tones' input and output in three tubes modeling the external auditory meatus[3]: an open-ended PVC tube 14.3 cm in length[4] with an inside diameter of 2 cm, that same tube with the output end covered in a tightly stretched rubber glove to model the tympanic membrane, and that same tube with an output end with dual openings, one at the end of the tube and the other in the floor of the tube wall to model the foramen of Huschke, both covered with a tightly stretched rubber glove.

The tube was mounted on top of a box in the configuration depicted in Figure 3-10 to allow placing the loudspeaker and microphone within the tube walls. The oscillator was an HP audio oscillator Model 200AP, set to generate tones at 10dB, and the oscilloscope for measuring the input and output signal was a Tektronic 2211.

---

[3] It is common for the external auditory meatus to be modeled as a rigid tube. The external ear produces these effects with structures that are usually viewed as static and passive (Rosowski 1994). Traditionally, when directional effects are ignored, the outer ear has been modeled by simple horns and uniform tubes (Rosowski 1996; Shaw & Stinson 1983). For example, modeling ear canals as hollow cylinders with rigid walls open at one end and closed at the other, Harrison (1998, 2001) studied the role that statistical fluctuations play in thermal excitation of resonant ear canals in establishing auditory thresholds for primates. Some researchers have proposed more complex models to accommodate viscous and thermal losses in tubes and horns which affect frequencies below 1 kHz in the ears of very small animals like guinea pigs and gerbils (Rosowski 1994; Zuercher, Carlson & Killion 1988).

[4] The 14.3 cm length instead of the standard 15 cm for simplified models was the result of a lab mishap.

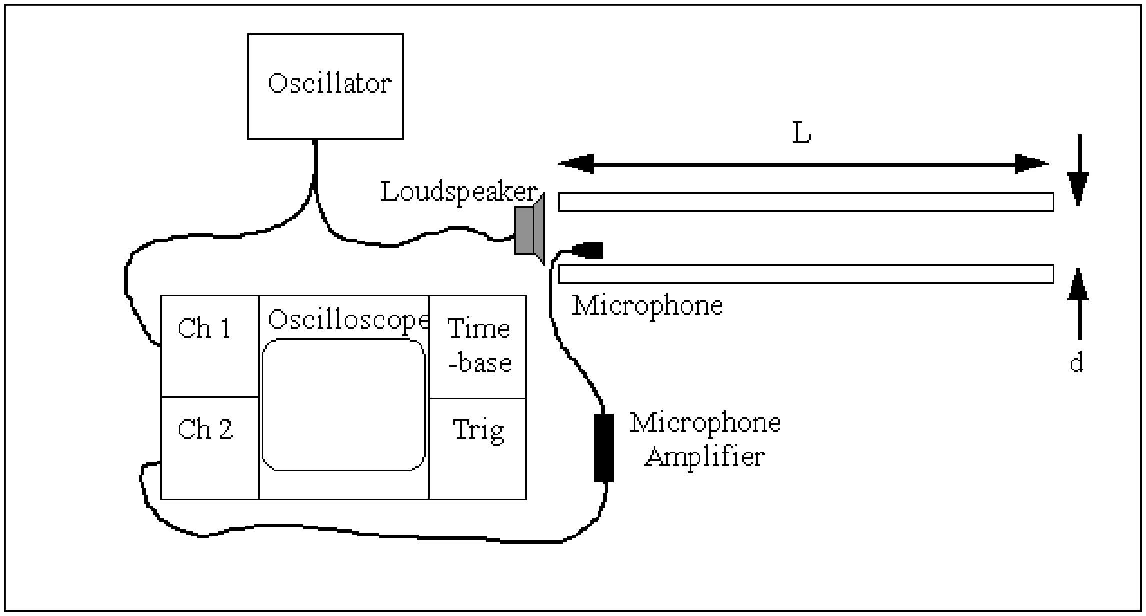


Figure 3-10. Configuration for Tone Generator and Microphone for Measuring Resonance in an Open-Ended Tube and Models of the External Auditory Meatus with and without a Foramen of Huschke

Acoustic resonance is the tendency for an acoustic system to absorb more energy when it is driven at a frequency that matches one or more of its own natural frequencies of vibration. The basic formula for determining resonant frequencies in a tube is F = (n)(v)/2L where n is a positive integer (1, 2, 3) representing the resonance node, L is the length of the tube, and v is the speed of sound in air (~ 343 m/sec, but varies with the temperature of the room). For a tube of 14.3 cm length, the equation would predict a third resonant frequency of approximately 3500 Hz, where the measured results in Figure 3-11 show a resonant frequency of 3250 Hz. Given that (1) the model is fairly rough, (2) the speed of sound in air was estimated, and (3) the loudspeaker and microphone partially blocked one end (explaining the slight resonance at 2900 Hz), these results are in line with the prediction. (The results in Figures 3-11 through 3-13 show the difference between the input and output signals in millivolts on the Y-axis.)

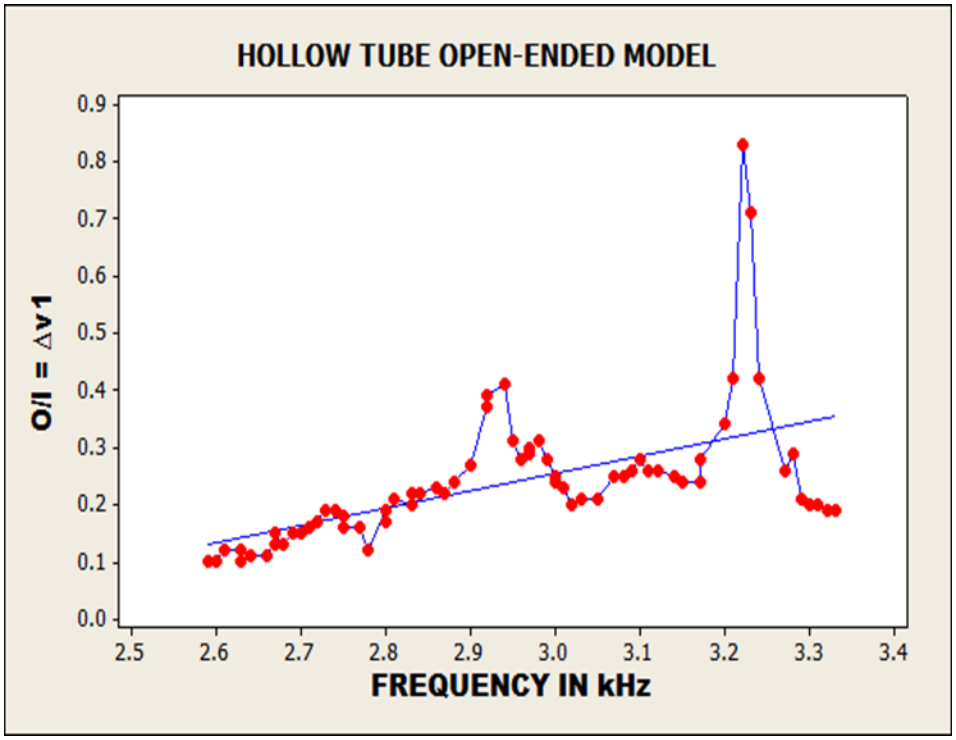


Figure 3-11. Resonance in an Open-Ended Tube

For close-ended tubes the formula is adjusted to account for the fact that the tube resonates when the length of the tube is ¼ the wavelength of the tube. In this model, the end is closed with a rubber glove, not a rigid wall, to model the tympanic membrane. The results are given in Figure 3-12.

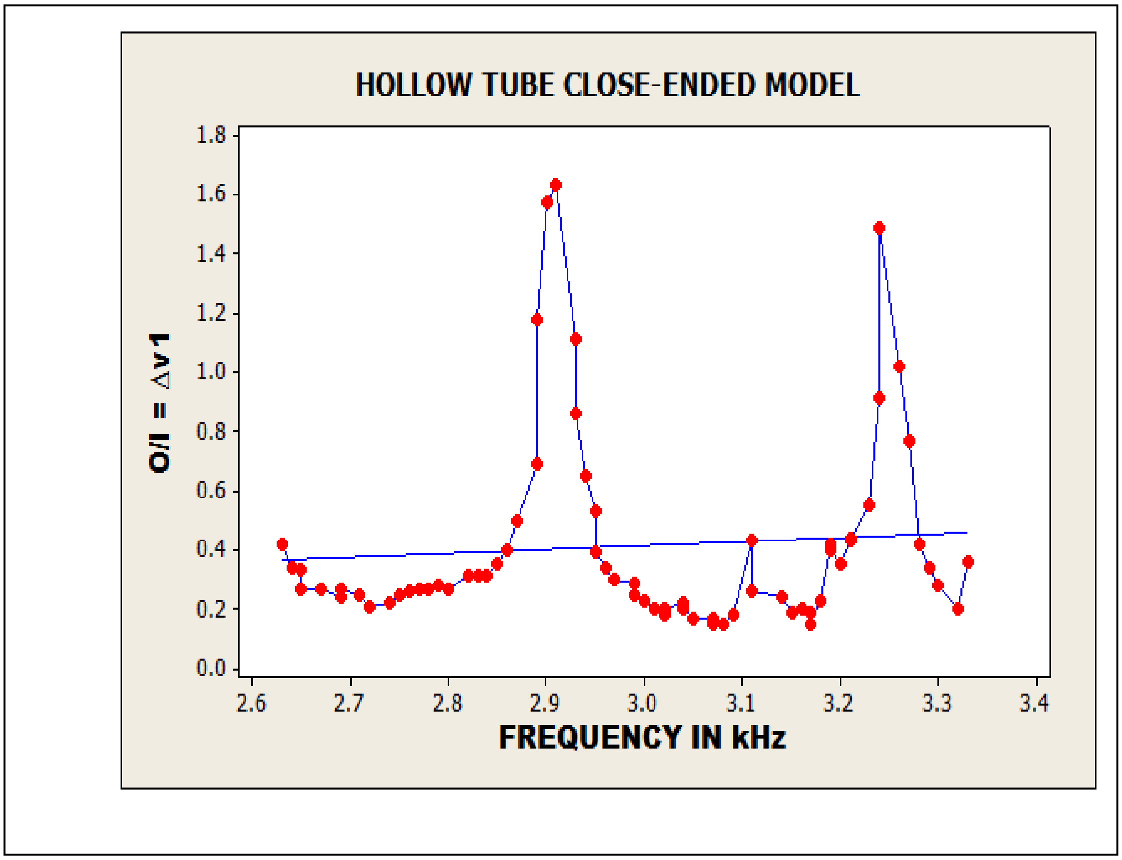


Figure 3-12. Resonance Results from Closing One End with a Rubber Glove to Model the Tympanic Membrane

Figure 3-13 presents the results from adding an aperture in the floor of the tube near the closed end to model the foramen of Huschke. This aperture was also covered with a tightly stretched rubber glove to model the fibrocartilaginous membrane. The result is to attenuate the resonance at around 3250 Hz.

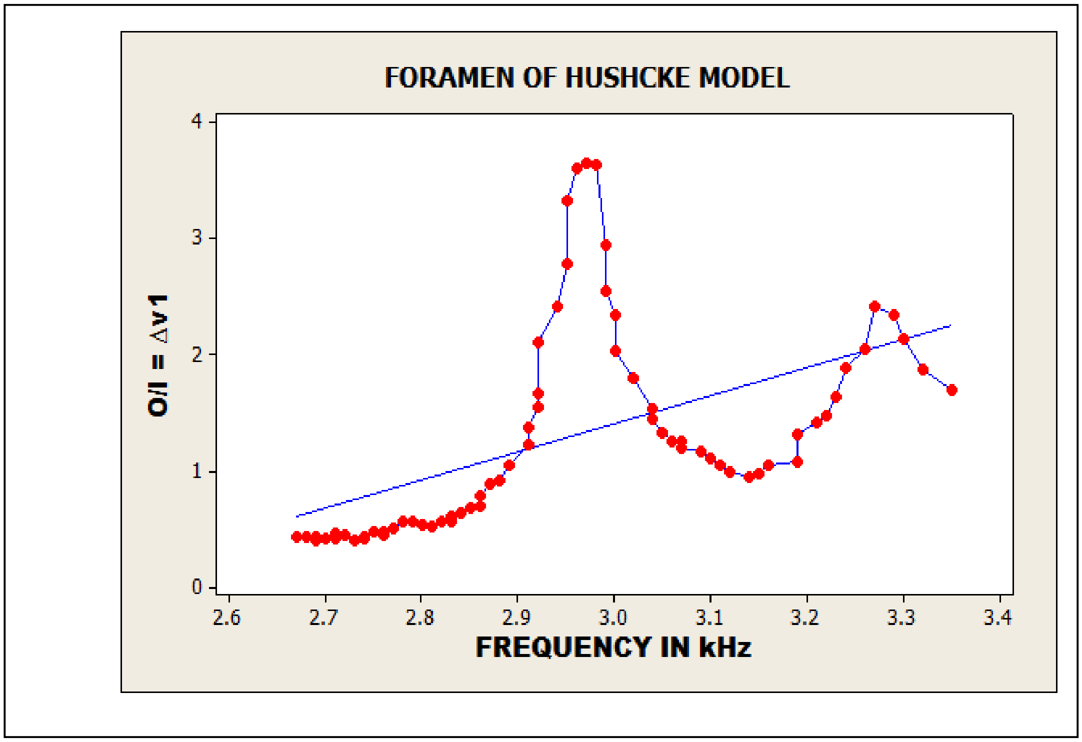


Figure 3-13. Resonance Results from Adding a Model of the Foramen of Huschke to a Closed-End Tube

The models depicted in Figures 3-11, 3-12, and 3-13 are crude, but they do suggest that the foramen of Huschke conditions sounds, including speech sounds, in a principled way that would vary throughout the development of the external auditory meatus and in the frequencies important to speech. But the foramen of Huschke is only one source of immaturity in the temporal bone during childhood. For example, one further consequence of the immaturity of the bone encircling the cavities of the ear is that infants and children likely have less stable cavities that are less isolated from other skull cavities as compared to adults. The immature temporal bones probably permit disturbances of auditory structures that have nothing to do with airborne speech sounds, for example, infants probably experience vibrations from sucking and from the throbbing of the jugular (Figure 3-4 above illustrates the position of the jugular relative to the cavities of the ear).

Further support that the auditory system can be subject to vibrations that have nothing to do with airborne sound even in adults comes from Minor et al. (1998), who identified superior canal dehiscence syndrome (SCDS). SCDS is a condition in which a small hole in the superior semicircular canal leads to debilitating problems with hearing and balance and complaints that patients can hear the movements of their eyeball, the crunching sound of their own footsteps, their heart beating, the echo of their own speaking voice, or disturbingly loud reverberations when brushing their hair or shaving (see also Janky et al. 2012). These auditory effects come about because the semicircular canals in the inner ear are abnormally activated through vibration traveling through the small hole in the semicircular canal. Additionally, noise may also arise from thermal excitation of air in the smaller volume of children’s ear canals (Harrison 1998, 2001).

One more consideration is that right and left temporal bones ossify asymmetrically (Ossenberg 1981), perhaps because of the greater innervation of the dominant hemisphere of a human’s brain, which may slow the rate of ossification on that dominant side (Torgersen 1951). This asymmetry may cause the right and left sides of the brain to receive different forms of the auditory signal, including speech sounds. It is well accepted that timing and intensity differences between sounds reaching the two ears are critical to the ability to localize the source of a sound (Pickles 1988), and indeed the neural mechanisms subserving sound localization are one of the last to mature, which makes sense since the head and therefore the distance between ears continually changes in childhood; perhaps differences in auditory images of speech sounds arriving at the auditory nerve from each of the ears are important to speech perception.

In utero, infants experience sounds through bone conduction, that is, direct stimulation of the cochlea through the bones of their skulls. Specifically, Guild (1936) emphasizes that in bone conduction, sound waves enter the intralabyrinthine fluids of the inner ear primarily through mechanical vibration transmitted through the osseous trabecula connecting the posterior wall of the external auditory meatus to the horizontal semicircular canal. In practice, this means that in utero, infants hear frequencies below 400 kHz. After birth, there is no question that brains of hearing infants and children are processing auditory information, including speech, through air-conducted transmissions that arrive at the auditory nerve from the cochlea. It seems possible that in general the patterning of sound waves in the peripheral auditory system changes as the temporal bone ossifies, and this dissertation is a step towards establishing that age-related changes in the peripheral auditory system "condition" the speech signal in ways that have perceptual consequences (many researchers have proposed various elements of frequency information that have perceptual consequences, e.g., Assmann, Nearey & Hogan 1982; Benguerel & McFadden 1989; Bladon 1983; Hillenbrand et al. 1995; Kewley-Port & Atal 1989; Nearey 1989; Strange 1989; Syrdal & Gopal 1986; Traunmüller 1984; Verbrugge et al. 1976; Zahorian & Jagharghi 1993).

As mentioned in Chapter 1, many investigators have used computer programs or other mathematical methods to model adult listeners' experience of speech, but these models have not accommodated the fact that auditory function in children develops over time in early childhood (e.g., Nozza, Rossman & Bond 1990; Olsho 1985; Rosowski 1996, 1994; Ruben 1992, 1995; Schneider, Trehub & Thorpe 1991; Sinnott & Aslin 1985; Werner & Gray 1998; Werner & Marean 1996; Werner & Rubel 1992). For example, Slaney (1999) used MATLAB m-functions to create six programs that primarily model cochlear physiology. Slaney (1999) also supplies a test sentence in the .wav format that can be imported into MATLAB using the wavread function. Based on psychoacoustic research, Patterson (1976) models filtering based on critical bands (also called auditory filter widths). Patterson, Allerhand, and Giguère's (1995) Auditory Image Model (AIM) is written in C. Like Slaney's MATLAB toolbox, the AIM programs read .wav files and carry out functional and physiological models of the cochlea's basilar membrane and neural firing patterns. Another interesting feature of the AIM programs is that they can be linked together with the output from one model serving as the input to the next stage in the process of audition.

To study the age-related changes caused by the development of temporal bones, I constructed a model based on psychoacoustic data from studies of children's hearing as reported in Schneider and Trehub (1992), described in Chapter 4 – Methodology. More recent research into the hearing and speech perception of infants and children (e.g., Fellman & Huotilainen 2006; Tharp & Ashmead 2001; Tsao, Liu, & Kuhl 2004) has continued to support the shape of the curves as reported by Schneider and Trehub (1992). This choice of auditory model relies on a few assumptions:

- *That the conductive immaturities, including immaturities of the middle ear (as summarized earlier in this chapter), can be treated in the main as temporal bone immaturities.*

One source of conductive immaturity located outside the temporal bone is the concha. A hollow of the pinna positioned inferior and dorsal to the opening of the ear canal, the concha contributes to the spectral transfer function. At birth in humans, the shape of the external ear (including the concha) is mature, but the size increases until about 9 years of age (Anson, Davies & Duckert 1991). The concha's contribution has been studied in adults (Shaw 1974) and in infants at ages 1, 3, 6, 12, and 24 months (Keefe et al. 1994). For example, there is a gain of approximately 10 dB at 5.5 kHz for1-month-olds and at 4.5 kHz for 24-month-olds (Keefe et al. 1994), which is higher than the frequencies most important in speech sounds. In adults, the pressure gain enhances the transmission of frequencies in the range of 2–7 kHz relatively uniformly.

Another source of conductive immaturity arguably located outside the temporal bone is the portion of the ear canal that is an extension of the external ear, about 8 mm in both infants and adults (Hollinshead 1962). At birth, the bony part of the ear canal is but a slender ring of little length laterally; the postnatal growth of the temporal bone's tympanic ring and squama produces the bony two-thirds of the external auditory meatus (Anson, Davies & Duckert 1991; Hollinshead 1954; Lee 2003). The diameter of the cartilaginous portion increases during development, but for the most part, changes in the resonance properties of the external auditory meatus develop because the temporal bone develops.

Because these two sources of conductive immaturity are relatively small in the frequencies important to speech, I have not adjusted Schneider and Trehub's (1992) curves for sound sensitivities in infants and children of different ages in an attempt to estimate only the effect of the temporal bones. I have also assumed that the foramen of Huschke is accounted for in the general picture of infant and child hearing sensitivities as reported in Schneider and Trehub (1992).

- ♦ *That the response of the cochlea to frequency information in speech is mature at birth and the frequency response of the central auditory nervous system is mature at 3–6 months.*

In terms of absolute sensitivity, which is the most common way to assess an auditory system's functional status, "there is little evidence to suggest that improvement in responses arising peripheral to the auditory brainstem is not completely accounted for by maturation of the conductive apparatus" in humans during the postnatal period (Werner & Marean 1996: 80). The neural immaturities that underlie wider auditory filters (causing poorer frequency tuning) in 3-month-olds as compared with older children and adults affect frequencies higher than 4 kHz; infants have adultlike auditory filters at 1 kHz.

- *That the external ear can be modeled as a rigid tube.*

The external auditory meatus will be conceptualized as a rigid tube uniform in shape in this study, which is a commonly accepted model as discussed above.

Page Left Intentionally Blank

4

METHODS

To test the hypothesis that during childhood the developing temporal bones contribute training effects to speech sounds, the auditory model outlined in Chapter 3 was used to filter phonetic data collected from adult speakers (this auditory model is described below). "Training effects" in this dissertation means filtering effects that simplify or otherwise modify the speech signal and possibly influence perceptual organization in adults.

4.1 DATA COLLECTION

The phonetic data collected controlled for dialectal variation. The adults participating in this collection of data were native speakers of English and bilingual speakers of Diné (the language of the Navajo people). The English speakers were close in age, within 8 years of one another. Both English speakers, male and female, lived in the Bay area of California well into their 20s. The Diné speakers were brother and sister, both in their 20s, who were monolingual speakers of Diné until they went to school where the language of instruction was English. Ms. Melvatha Chee, a linguist and a native speaker of Diné, acted as a consultant to the researcher translating the Navajo speakers' stories and assisting with compiling the list of words that contained the target vowels in parts of the word that were not grammatical morphemes (for those unfamiliar with Diné, the language exhibits a productive and extensive inflectional morphology).

The materials that the adult speakers recorded were designed to elicit the steady-state vowels of English and Diné.[1] Vowels are more acoustically distinct than consonants; they are longer in duration and louder in intensity[2] (e.g., Denes & Pinson 1993). Gervain and Mehler (2010: 205) report research from Nespor and colleagues in 2003 and Toro and colleagues in 2008 that suggests "consonants and vowels have different linguistic functions."[3] Furthermore, frequency information is more important as an acoustic cue in vowel perception than in consonant perception. For instance, several recent studies have investigated the contributions of consonants and vowels to intelligibility of words and sentences, and they have found that vowels are more important to intelligibility than consonants. Kewley-Port, Burkle, and Lee (2007) report that "vowels carry more information about sentence intelligibility" than consonants, and Fogerty and Humes (2012: 1490) in reviewing several such studies conclude "[t]hus, it appears that vowels highlight an important locus of speech information." Therefore, for this initial foray into the question of how children are experiencing the speech sounds of their ambient language(s) at different ages, vowels are a good place to look for training effects. Using only steady-state vowels was a useful way to limit the size of the study's wordlists while controlling for questions raised by issues like timing and formant transitions in English diphthongs and Diné tones.

---

[1] Steady-state vowels are s subset of the vowel inventories of both English and Diné.

[2] Caramazza et al. (2000) promote the view that vowels and consonants are just labels used to distinguish intensity differences between peak (vowel) and non-peak (consonant) parts of a continuous stream of sound and conclude that that vowels and consonants are processed by distinct neural mechanisms, which supports the view of their independent status in language production as well as in perception.

[3] Where consonants are thought to encode the lexicon, vowels signal morphological form and syntactic functions.

The languages included in the study are unrelated and have different vowel systems. English is a West Germanic language within the Indo-European family; Diné is an Athabaskan language traditionally spoken in Arizona and New Mexico. The vowels of American English are dispersed throughout the vowel space (e.g., Dillon 2003) traditionally thought to be defined by physiological constraints of the vocal tract's constrictions (e.g., Pickett 1999). Diné, by contrast, has fewer oral vowels than English, and the vowel system shows a different pattern of dispersion. Diné has four vowels, none of which is a high back vowel (Maddieson 1984; McDonough, Ladefoged & George 1992). Figure 4-1 below presents illustrations of the dispersion of the two vowel systems.

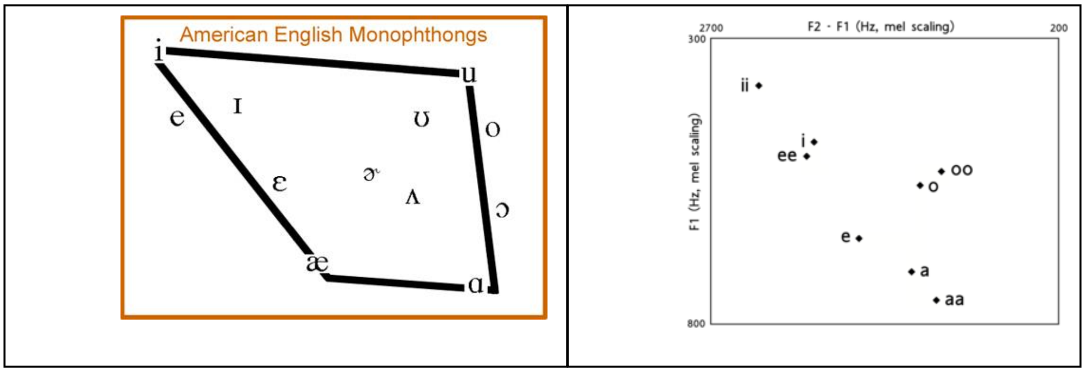


Figure 4-1. The graph on the left diagrams the monothongs of English (figure from http://laits.utexas.edu/texas_english/interactive/monophthongchart.jpg) and the graph on the right diagrams the steady-state vowels in Diné (from McDonough 2003).

The words included in the test materials were nouns that infants and children encounter in all languages. Clark (1979, 2001), for example, summarizes cross-linguistic evidence about the content of children's first 50 words: words for food, body parts, clothing, vehicles, toys, household items, and people. While the content of infants' and children's first 50 words in Navajo communities has not been specifically studied, extrapolating from this cross-linguistic research seems a reasonable way to decide on the content of the test materials to be constructed for this pilot study. Words were also pulled from spontaneous narratives produced by two native speakers[4] of English and Diné of a sequence of pictures from a children's book (*Silly Ruby*). The stories from these two speakers were recorded and transcribed, then analyzed for basic vocabulary in English and Diné that the researcher thought anyone narrating the pictures likely would produce (e.g., words for sheep, tree, apple, dibé, t'iis, bilasáana—see Figure 4-2). These words were included in the carrier sentences in the hopes they would help prime the words' appearance in speakers' stories.

(a) (b)

Figure 4-2. (a) The sheep is lying asleep under a tree when an apple hits it in the head, and (b) The sheep kicks the tree that dropped the apple. The sequence of eight pictures shown to speakers for narration came from the children's book *Silly Ruby* (Friend 2000).

The test materials were designed to elicit the steady-state vowels of the two languages in the different contexts of several words. The materials elicited the words containing the vowels in two different styles. In the first style there was a list read twice in two different random orders with the words containing the target vowels embedded in a carrier frame. In this study, the frame "Simon says (word)" was used for English speakers and in Diné the frame was "Díish (word) daolyé," which roughly translates into

---

[4] The recordings from these two speakers were not included in the data analyzed—they were used only to compile the word lists.

English as “What is it that you call (word).” Repetition 2 of these sentences in both languages has been included as Appendix 4-A to this chapter. The Diné list has been amended to provide a gloss and to identify the target vowel. The words included in the statistical analysis for this dissertation have been highlighted.

Speakers were instructed that they would be reading the two repetitions of the sentences at a comfortable rate and at a comfortable level of vocal effort. In recognition that the bilingual Diné speakers being asked to produce Diné in a context (data collection) heavily cued for English, speakers of both languages were encouraged to take a few minutes to practice reading at least a few of the sentences aloud before recording while the researcher left the room. Speakers were urged to moisten their vocal folds with a drink of water before beginning the recordings and instructed that it was fine to stop and take a sip of water at any time.

In the second style of elicitation, speakers were asked to narrate the sequence of eight pictures from *Silly Ruby* as described above[5]. Using a little-known book in the spontaneous narrative task rather than a classic, such as Mercer Mayer’s *Frog, Where Are You?* which is often used in language acquisition research (MacWhinney 2000), ensured that all speakers were equally unfamiliar with the sequence of pictures that they narrated.

The recordings were made on a Sony WM D6C with a Shure headset unidirectional mike. At the outset of recording participants were asked to state their

---

[5] The speakers’ stories were transcribed and the recordings digitized, but the vowels collected in these stories were not included in the statistical analysis for this dissertation in the interests of simplifying the data set.

name, the date, and the place of the recording so that the investigator could set the input level dial at a setting appropriate to the speaker's voice.

4.2 DATA MEASUREMENTS

The Computerized Speech Laboratory (CSL) software from Kay Elemetrics was used to digitize the recordings at 10K which is adequate for study of the frequencies important to vowels (under 5K). During digitization notes were made about data that needed to be discarded because of pauses or other disfluencies.

The average amplitudes of the voices as recorded was 73 dB (30 dB is whispering, 50 dB is ordinary conversation at a distance of 3 meters, 70 dB is ordinary conversation at a distance of approximately 30 centimeters, approximately the distance for a infant being held in the crook of an adult's elbow). Included in this dissertation's analysis is a selection of the data collected. In English /i/, /e/, /æ/, /a/, /^/, /o/, /ʊ/, and /u/ were analyzed, and in Diné, /i:/, /i/, /e:/, /e/, /a:/, /a/, /o:/, and /o/. Each of these vowels was included in two repetitions of three words. For example English /i/ was collected in two repetitions each of the words "sees," "sheep," and "keep."

Using the freeware Praat (Boersma & Weenink 2014), FFT analyses were performed at the midpoint of the vowels, both with the command Get Formant (using the default range 0–5500 kHz) and a script to save a spectrum over a 40 ms window over the midpoint calculated from markers set by the researcher at the vowel's beginning and end. The Get Formant command yields average formants for the first, second, third, and fourth formants (F1, F2, F3, and F4), while the spectrum analysis yields the real and imaginary values for power (amplitude) at 1,029 regularly spaced frequencies between 400 Hz and

5000 Hz for a two-sided frequency domain. These values were used to calculate a one-sided power spectral density using this formula (Boersma 2001):

$$PSD(f) \equiv 2|X(f)|^2 / (t_2 - t_1)$$

where $|X(f)|^2$ is calculated from squaring the real component and adding that to the square of the imaginary component and $t_2 - t_1$ is the 40 ms window.

The power spectral density is a measure of the average power in a sound during a certain time range and in a certain frequency range, and it is expressed in the unit Pa^2/Hz. Since the auditory model was constructed from data expressed in decibels it was necessary to convert the FFT values to decibels with this formula:

$$PSD_{dB}(f) = 10 \log_{10} \{ PSD(f) / P_{ref}^2 \}$$

where PSD stands for power spectral density, "f" is frequency, and $P_{ref} = 2 \text{ X } 10^{-5}$ Pa. Excel was used to calculate the PSD from the real and imaginary values reported by Praat for each frequency bin sampled and then converting the PSD values to decibels. Once the amplitudes were in decibels, "residual amplitudes" were calculated for the frequency bins that passed through the filter modeling children's sensitivity to frequencies. In other words, the amplitude in decibels for that bin was subtracted from the threshold value (also in decibels). The rationale for calculating residual amplitudes is that they emphasize the strength at which the signal exceeded the cutoff thresholds.

4.3 DATA FILTERS

As mentioned in Chapter 3, this dissertation uses a model based on psychoacoustic data collected in experiments into the speech perception of infants and children conducted by several groups of researchers working in different labs and using different techniques to collect data as appropriate to the age of the infants and children being studied. Schneider and Trehub (1992) and Werner and Marean (1996: 67) summarize these studies as shown in Figure 4-3.

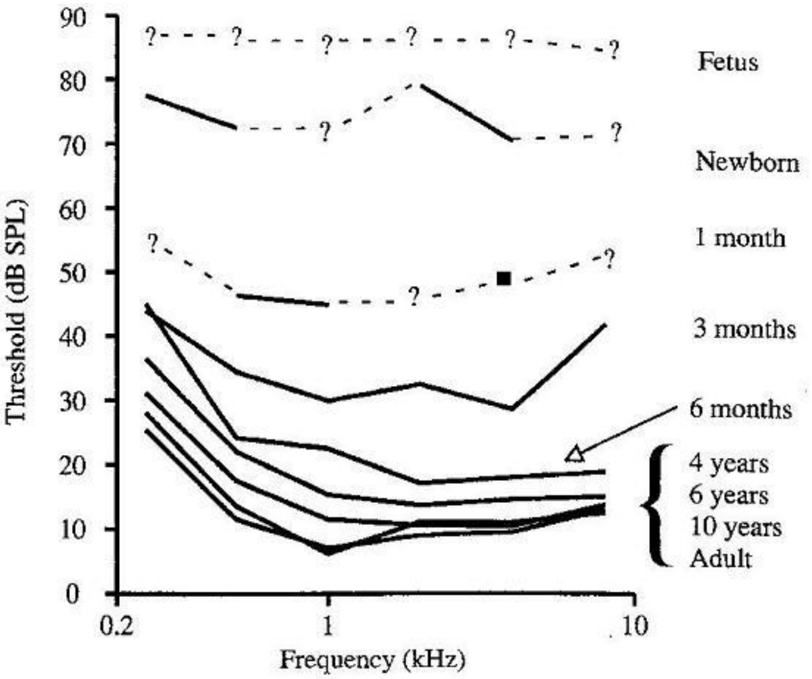


Figure 4-3. Audibility Curves Plotted from the Psychoacoustic Literature (Werner and Marean 1996: 67). Question Marks and Dashed Lines indicate Missing Data.

As can be seen in Figure 4-3, infant thresholds can be 60–70 dB higher than adults above 500 Hz (based on Weir 1979). At the same time, infants also show better sensitivity to frequencies below 500 Hz than higher frequencies, with progressive, fastest improvement in higher frequencies occurring until age 6 years. Low frequency thresholds continue to drop until age 10 years. The sounds of human speech fall in the range of 50 Hz to 10,000 Hz, with the range of 500–4000 Hz generally considered the most important for vowel perception (Denes & Pinson 1993).

The auditory model constructed for this dissertation relies on the patterns of sensitivities to frequencies at different ages summarized in Figure 4-3. The graph in Figure 4-3 was enlarged and point values were determined along the lines from the use of a more finely graded scale. The slopes of the four main regions of the curves[6] for children aged 6 months, 1 year, 18 months, 3 years, 5 years, and 20 years (adult) were thus estimated in order to create a series of auditory splines, as plotted in Figure 4-4.

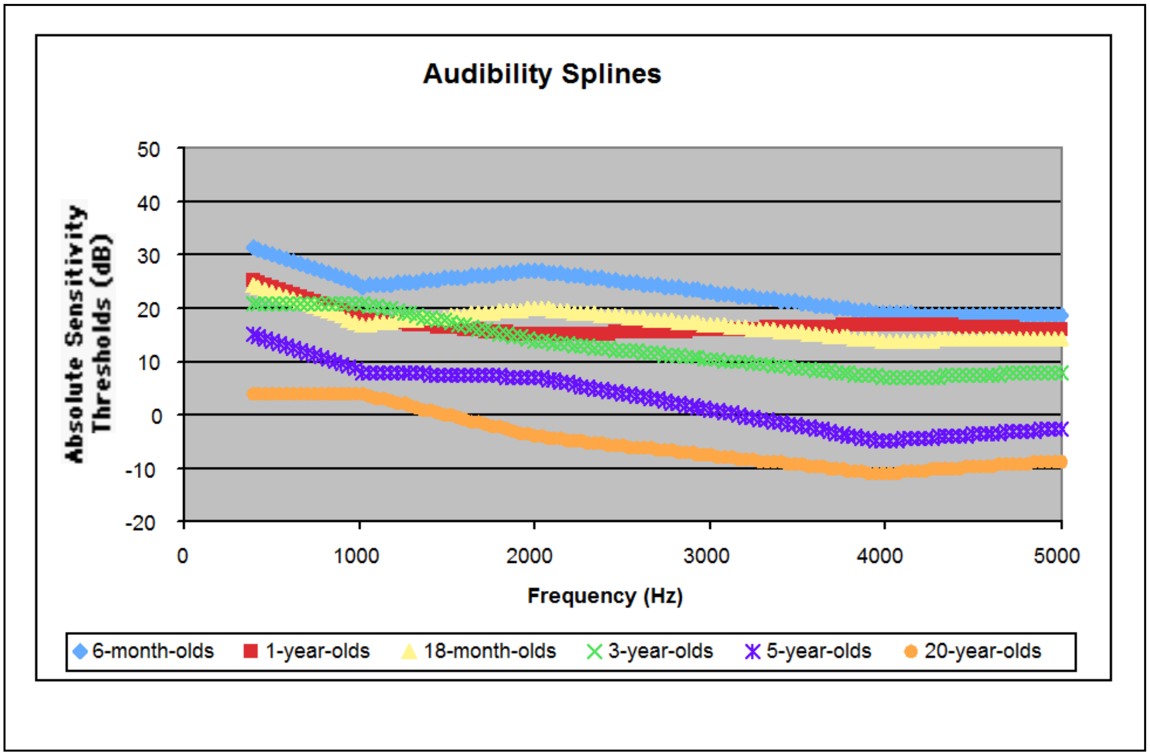


Figure 4-4. Splines Calculated from Schneider and Trehub (1992) and Werner and Marean (1996) Showing That Children and Adults (20-year-olds) Differ in Their Absolute Sensitivities to Frequencies

Once the splines were drawn it was possible to interpolate values for frequencies that fell between the frequencies that were actually tested in these studies. A series of cutoff points in decibels for each age at each frequency to be studied was calculated in an

---

[6] Data from more than one study for the reported ages were available for the regions between 400 Hz and 1000 Hz; 1000 Hz and 2000 Hz; 2000 Hz and 4000 Hz; and 4000 Hz and 10000 Hz.

Excel worksheet. Then each frequency sampled in each vowel in the two repetitions of the three words was evaluated against the cutoff points (Figure 4-5 is a graphic depiction of these evaluations for one vowel at each of the ages—each bar of the spectrum represents the amplitude for that frequency bin sampled). Those frequency bins with amplitudes that did not exceed the threshold of the cutoff points were set aside. A "residual" value was also calculated for the remaining amplitudes as a measure of how strongly the amplitude exceeded the threshold.

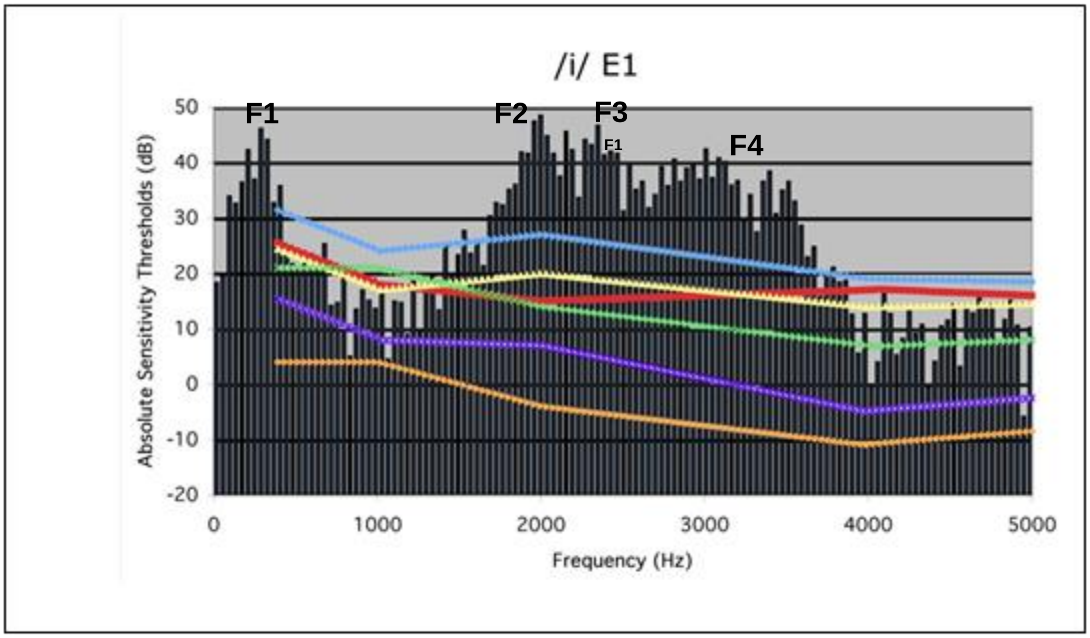


Figure 4-5. Spectrogram of English /i/ Overlain with Auditory Splines. The blue line depicts auditory sensitivities at 6 months, the red line 1 year, the yellow line 18 months, the green line 3 years, the purple line 5 years, and the orange line 20 years. Frequency bins that fall below the splines were set aside in the statistical analysis (this figure was created by Shannon Whitehead.)

Visual inspection of Figure 4-5 shows that at 6 months (the pattern traced by the blue line), the area of greatest sensitivity to the formants[7] of /i/ comes in the frequencies higher than F1, F2, and F3. In other words, the lowest point of the blue line shows the point of greatest sensitivity—the 6-month-olds in the psychoacoustic studies summarized by Schneider and Trehub heard frequencies at 4000 Hz and higher delivered at only 18 dB where they did not heard frequencies at 2000 Hz until they reached 28 or so decibels in intensity. At age 18 months (the pattern traced by the yellow line), children appear to lose sensitivity in the F2/F3 (2000 Hz–2500 Hz) region that they had at 1 year.

Using a combination of statistical software (Palaeontological Statistics [PAST] v2.17 and Minitab v16), the results reported in Chapter 5 explore the effects of such sensitivity patterns in the spectra for all vowels analyzed.

---

[7] Vowel formants are resonances of the vocal tract. Changing the configuration of the vocal tract changes the vowel that's produced and therefore the formant pattern changes. Denes and Pinson (1993: 142) note that "usually the first three or four formants are adequate for satisfactory perception."

Page Left Intentionally Blank

APPENDIX 4-A

RANDOMIZED ORDERED CARRIER SENTENCES READ BY PARTICIPANTS

(Highlighted words are words used in Chapter 5's statistical analysis; bold characters are the target vowels)

Page Left Intentionally Blank

English Carrier-Sentence List
Repetition 2

1. Simon says f**a**ll. Target vowel: /a/
2. Simon says m**a**d. Target vowel: /æ/
3. Simon says sh**ee**p. Target vowel: /i/
4. Simon says sh**a**kes. Target vowel: /e/
5. Simon says gr**a**ss. Target vowel: /æ/
6. Simon says t**a**king. Target vowel: /e/
7. Simon says g**o**es. Target vowel: /o/
8. Simon says fr**e**ckles. Target vowel: /ɛ/
9. Simon says g**oo**d. Target vowel: /ʊ/
10. Simon says n**o**se. Target vowel: /o/
11. Simon says c**ur**ly. Target vowel: /ɝ/
12. Simon says pops. Target vowel: /a/
13. Simon says h**ea**d. Target vowel: /ɛ/
14. Simon says st**u**ff. Target vowel: /ʌ/
15. Simon says f**e**ll. Target vowel: /ɛ/
16. Simon says h**ur**t. Target vowel: /ɝ/
17. Simon says k**i**cks. Target vowel: /I/
18. Simon says **a**pples. Target vowel: /æ/
19. Simon says s**ee**s. Target vowel: /i/
20. Simon says **u**p. Target vowel: /ʌ/
21. Simon says l**e**gs. Target vowel: /e/
22. Simon says b**a**ck. Target vowel: /æ/

23. Simon says s**aw**. Target vowel: /Ɔ/

24. Simon says f**oo**led. Target vowel: /µ/

25. Simon says thr**ee**. Target vowel: /i/

26. Simon says sl**ee**ping. Target vowel: /i/

27. Simon says h**i**t. Target vowel: /I/

28. Simon says s**o**. Target vowel: /o/

29. Simon says tw**o**. Target vowel: /µ/

30. Simon says s**ai**d. Target vowel: /ɛ/

31. Simon says l**i**ttle. Target vowel: /I/

32. Simon says g**ir**l. Target vowel: /ɝ/

33. Simon says t**oo**k. Target vowel: /ʊ/

34. Simon says c**o**vered. Target vowel: /ʌ/

35. Simon says s**ear**ching. Target vowel: /ɝ/

36. Simon says tr**ee**. Target vowel: /i/

37. Simon says s**a**ys. Target vowel: /ɛ/

38. Simon says m**oo**d. Target vowel: /µ/

39. Simon says w**a**kes. Target vowel: /e/

40. Simon says p**i**cks. Target vowel: /I/

41. Simon says l**oo**ks. Target vowel: /ʊ/

42. Simon says p**o**kes. Target vowel: /o/

43. Simon says th**a**nk. Target vowel: /ãe/

44. Simon says f**oo**ling. Target vowel: /µ/

45. Simon says l**o**ts. Target vowel: /a/

46. Simon says **an**gry.	Target vowel: /ãe/
47. Simon says t**a**kes.	Target vowel: /e/
48. Simon says h**i**tting.	Target vowel: /I/
49. Simon says asl**ee**p.	Target vowel: /i/
50. Simon says f**a**lling.	Target vowel: /a/
51. Simon says thr**o**ws.	Target vowel: /o/
52. Simon says s**ear**ch.	Target vowel: /ɝ/
53. Simon says n**a**p.	Target vowel: /æ/
54. Simon says b**oo**k.	Target vowel: /ʊ/
55. Simon says s**o**me.	Target vowel: /ʌ/
56. Simon says sh**oo**k.	Target vowel: /ʊ/
57. Simon says c**o**mes.	Target vowel: /ʌ/
58. Simon says s**i**tting.	Target vowel: /I/
59. Simon says h**i**ll.	Target vowel: /I/
60. Simon says w**o**ke.	Target vowel: /o/
61. Simon says k**ee**p.	Target vowel: /i/
62. Simon says f**oo**d.	Target vowel: /µ/
63. Simon says fr**ui**t.	Target vowel: /µ/
64. Simon says thr**e**w.	Target vowel: /µ/
65. Simon says l**oo**king.	Target vowel: /ʊ/
66. Simon says l**a**mb.	Target vowel: /æ/
67. Simon says r**u**de.	Target vowel: /µ/
68. Simon says k**i**cking.	Target vowel: /I/

Diné Carrier-Sentence List
Repetition 2

1	Díish nábid**ii**łne’ daolyé.	Gloss: ‘it [the apple] hit the sheep’ Target vowel: /i:/
2	Díish yinaan**ee**ztą́ą́ daolyé.	Gloss: ‘it [the sheep] investigated again’ Target vowel: /e:/
3	Díish ashdl**a**’ daolyé.	Gloss: ‘five’ Target vowel: /a/
4	Díish ł**i**zhin daolyé.	Gloss: ‘black’ Target vowel: /i/
5	Díish néid**ii**nil daolyé.	Gloss: ‘she pick several [apples] up’ Target vowel: /i:/
6	Díish bin**ii**’jį' daolyé.	Gloss: ‘her face’ Target vowel: /i:/
7	Díish yidl**o**hgo daolyé.	Gloss: ‘she is smiling’ Target vowel: /o/
8	Díish héid**ee**z’99’o daolyé.	Gloss: ‘she is looking for it’ Target vowel: /e:/
9	Díish **a**t’ééd daolyé.	Gloss: ‘girl’ Target vowel: /a/
10	Díish han**oo**lne’ii daolyé.	Gloss: ‘it poked its head out’ Target vowel: /o:/
11	Díish n**aa**násdz9ǫ daolyé.	Gloss: ‘while it is standing there again’ Target vowel: /a:/
12	Díish náád**ee**st’99’ daolyé.	Gloss: ‘it is looking again’ Target vowel: /e:/
13	Díish ałk'ínaniid**ee**’ daolyé.	Gloss: ‘they fell on top of one another’ Target vowel: /e:/
14	Díish b**i**lasáana daolyé.	Gloss: ‘apple’ Target vowel: /i/
15	Díish ł**i**tso daolyé.	Gloss: ‘yellow’ Target vowel: /i/

16	Díish bin**e**’d66’o daolyé.	Gloss: ‘from behind’ Target vowel: /e/
17	Díish ałh**o**sh daolyé.	Gloss: ‘it is asleep’ Target vowel: /o/
18	Díish naalts**oo**s daolyé.	Gloss: ‘paper’ Target vowel: /o:/
19	Díish báh**oo**chįįdgo daolyé.	Gloss: ‘it got mad’ Target vowel: /o:/
20	Díish d**i**nilchíí’ daolyé.	Gloss: ‘pink’ Target vowel: /i/
21	Díish ch’il lát**a**h hozhóón daolyé.	Gloss: ‘flower’ Target vowel: /a/
22	Díish bijáádii daolyé.	Gloss: ‘legs’ Target vowel: /a:/
23	Díish d**oo**tł’izhgo daolyé.	Gloss: ‘blue’ Target vowel: /oo/
24	Díish n**aa**neiztał daolyé.	Gloss: ‘the sheep kicked [the tree] again’ Target vowel: /a:/
25	Díish ts**o**sts’id daolyé.	Gloss: ‘seventeen’ Target vowel: /o/
26	Díish bitsiits’**ii**n daolyé.	Gloss: ‘its [the sheep’s] head’ Target vowel: /i:/
27	Díish yilch**áá**zh daolyé.	Gloss: ‘fluffy’ Target vowel: /a:/
28	Díish dibé daolyé.	Gloss: ‘sheep’ Target vowel: /i/
29	Díish kón**e**’ daolyé.	Gloss: ‘across’? Target vowel: /e/
30	Díish tátł’idgo d**oo**tł’izh daolyé.	Gloss: ‘blue watermoss’? Target vowel: /oo/
31	Díish tł’**o**h daolyé.	Gloss: ‘grass’ Target vowel: /o/

32	Díish áłts**óó**zí daolyé.	Gloss: ‘thin’ Target vowel: /o:/
33	Díish bin**aa**’ daolyé.	Gloss: ‘its eyes’ Target vowel: /a:/
34	Díish h**o**zh= daolyé.	Gloss: ‘good’ Target vowel: /o/
35	Díish bigh**aa**’ daolyé.	Gloss: ‘its wool’ Target vowel: /a:/
36	Díish n**oo**d==z daolyé.	Gloss: ‘striped’ Target vowel: /o:/
37	Díish nináániid**ee**’ daolyé.	Gloss: ‘they dropped’ Target vowel: /e:/
38	Díish báh**a**ch9’ daolyé.	Gloss: ‘angry’ Target vowel: /a/
39	Díish h**a**lchíí’ daolyé.	Gloss: ‘the area is red’ Target vowel: /a/
40	Díish y**i**shch’il daolyé.	Gloss: ‘curly’ Target vowel: /i/
41	Díish ł**i**tsxo daolyé.	Gloss: ‘orange’ Target vowel: /i/
42	Díish neid**ii**łts’in daolyé.	Gloss: ‘picks up several sticks’? Target vowel: /i:/
43	Díish t**áá**’ daolyé.	Gloss: ‘one’ Target vowel: /a:/
44	Díish bik’ín**aa**lts’id daolyé.	Gloss: ‘fell on top of the sheep’ Target vowel: /a:/
45	Díish ninádin**oo**dah daolyé.	Gloss: ‘they will drop again’ Target vowel: /o:/
46	Díish ts**ee**bíí daolyé.	Gloss: ‘eight’ Target vowel: /e:/
47	Díish nin**ee**z daolyé.	Gloss: ‘tall’ Target vowel: /e:/

48	Díish h**o**t’éé lá daolyé.	Gloss: ‘the end’ Target vowel: /o/
49	Díish t’áá ákód**i** daolyé.	Gloss: ‘that is all’ Target vowel: /i/
50	Díish ts**i**n daolyé.	Gloss: ‘stick’ Target vowel: /i/
51	Díish nináád**oo**lts’ił daolyé.	Gloss: ‘it will drop again’ Target vowel: /o:/
52	Díish ałk’**ii**stł’in daolyé.	Gloss: ‘they are stacked on one another’ Target vowel: /i:/
53	Díish dah n**oo**chał daolyé.	Gloss: ‘it is fluffy’ Target vowel: /o:/
54	Díish nihin**e**ł9 daolyé.	Gloss: ‘the sheep is looking at us’ Target vowel: /e/
55	Díish gón**e**’é daolyé.	Gloss: ‘inside’ Target vowel: /e/
56	Díish tsíd**ee**sk**ee**z’go daolyé.	Gloss: ‘when it [the sheep] began to think about’ Target vowel: /e:/
57	Díish t’**ii**s daolyé.	Gloss: ‘tree’ Target vowel: /i:/
58	Díish b**i**kéé’ii daolyé.	Gloss: ‘its feet’ Target vowel: /i/
59	Díish d**ee**z’99’ daolyé.	Gloss: ‘it [the sheep] is looking’ Target vowel: /e:/
60	Díish d**a**h násk’id daolyé.	Gloss: ‘they are piled up’ Target vowel: /a/
61	Díish n**aa**ki daolyé.	Gloss: ‘two’ Target vowel: /a:/
62	Díish yizt**a**ł daolyé.	Gloss: ‘[the sheep] kicked [the tree]’ Target vowel: /a/
63	Díish yin**e**ł’98go daolyé.	Gloss: ‘when [the sheep] is looking at it’ Target vowel: /e/

64	Díish yich’**i**do daolyé.	Gloss: ‘she is scratching [her head]’ Target vowel: /i/
65	Díish n**ee**znáago daolyé.	Gloss: ‘when there are ten’ Target vowel: /e:/
66	Díish ł**i**chíí’ daolyé.	Gloss: ‘red’ Target vowel: /i/

5

RESULTS AND DISCUSSION

The overarching goal of this dissertation is to explore the idea that the development of the temporal bones exerts “training effects” on children’s perception of vowels. If training effects exist, then the formation of the perceptual strategies of adult hearing would thus be constrained in the developing child by the acoustic information made available by the temporal bones. This chapter presents a series of exploratory analyses conducted on the English and Diné data described in Chapter 4.

5.1 GENERAL EFFECTS CAUSED BY IMMATURE HEARING

The first analysis was carried out in PAST to establish that the auditory modeling makes the vowels different from the unfiltered input (see Figures 5-1 and 5-2). Dataset 1 includes one instance of /u/ collected from the male English speaker (E1) in the word “food.” Because back vowels are inherently less energetic than front vowels (e.g., Moser, Michel & Fotheringham 1961; Weber & Smits 2003), the filtering at the youngest ages—6 months, 1 year, 18 months—drastically simplifies high back vowels, particularly in male voices (see the discussion on male-female differences in Section 5.4). This simplification is a reduction in the number of frequency bins that arrives at the auditory nerve of young children. The scales are different from each other in Figures 5-1 and 5-2 because the unfiltered /u/ graphed in Figure 5-1 includes frequencies from 0 to 5000, while the filtered /u/ graphed in Figure 5-2 shows output constrained to the 400 Hz to 1120 Hz range. The filtered /u/ graph plots only those frequencies that were of sufficient intensity to exceed the sensitivity thresholds for 6 month-olds (as described in Chapter 4) for each frequency bin sampled in the vowel.

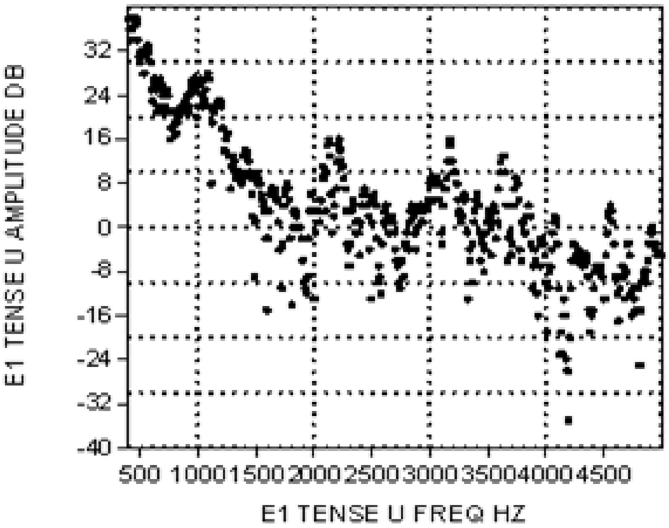


Figure 5-1. This graph plots the unfiltered data (frequency in Hz on the x-axis and amplitudes in dB on the y-axis) for one instance of speaker E1's back vowel /u/.

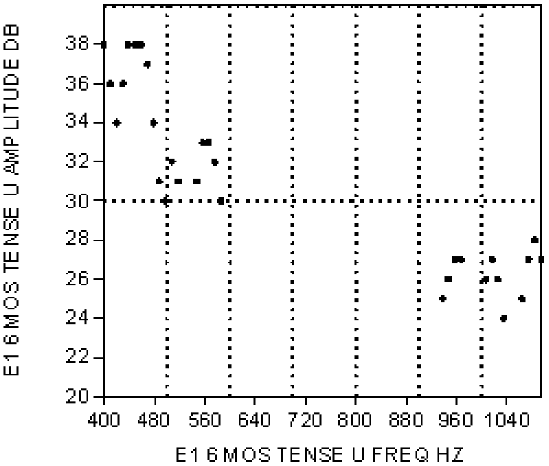


Figure 5-2. This graph plots the effect of the 6-month filter of one instance of speaker E1's back vowel /u/.

Figure 5-3 displays PAST plots of Dataset 2, comprising all instances of vowels collected from the male Diné speaker (D1) filtered according to the 1-year and 3-year models. Inspection of the two plots in Figure 5-3 suggests a slightly different shape for the vowels taken in aggregate at the two ages. At age 1 year frequencies above 4000 Hz are for the most part missing from all vowels. The dense bow along the bottom of the 1-year-old plot reflects the fact that children at age 1 year have hearing that is more sensitive to frequencies between 500 Hz and 2000 Hz relative to the frequencies above 2000 Hz as compared to the results filtered according to hearing sensitivities at age 3 years.

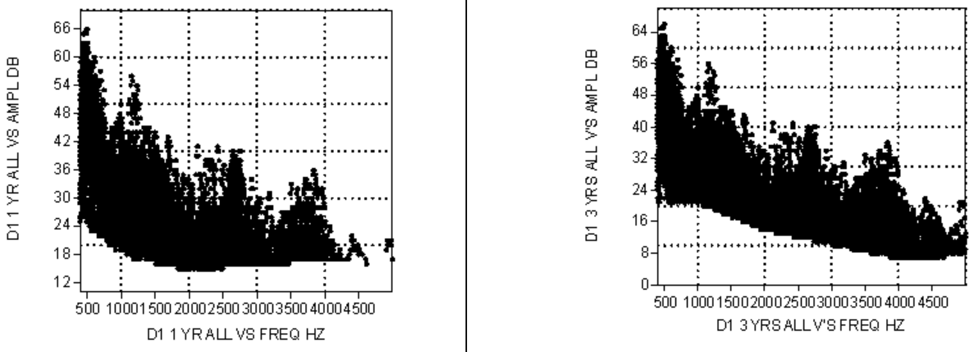


Figure 5-3. Data for all Diné vowels in aggregate using the 1-year (left) and 3-year (right) filters are plotted for speaker D1.

To explore how significant such general differences in aggregated vowel shapes are for a language, the analysis turns to a statistical technique for looking at the differences in multidimensional data. This dissertation's analysis uses a cubic generalized linear model (GLM) throughout. Whereas a linear GLM (also called a first order model) can show a steady rate of increase or decrease in the data, and a quadratic model, or second order model, can account for curvature (roughly in the shape of a U or an inverted U) in the data.

However, a cubic, or third order, model was necessary for the analyses of this dissertation's data to describe the peak-and-valley pattern in the data collected (Minitab 2010). Figure 5-4 uses the Akaike Information Criterion (AIC) to simplify the exploration of the third-order multidimensionality of the aggregated vowels in the Diné data (see Akaike 1974). The AIC constrains the maximal amount of information in multidimensional space. Using the AIC prevents "Overdimensionality" or overspecification of the model referring to the fact that any collection of numbers (including values for frequencies and amplitudes) has patterns within it, but some of the patterns will be "trivial," or uninformative. The point of the AIC stepping procedure is to reach a point where there is no longer any improvement in fit from adding dimensions. When the AIC's value no longer reduces, the result is a graph in which uninformative patterns in the data are minimized.

In this dissertation's analysis, the AIC stepping procedure results in a graph of the most informative parts of the data.[1] Excluding the trivial information focuses the plots in Figure 5-4 on formant peaks and reduces the diffuseness of the peaks seen in Figure 5-3's graphs of the same data. The lower the AIC, the more robust the model, and an AIC of 5.8 is low. (An AIC value depends on the number of data points. Dataset 2 encompassed 21,454 data points, and a high AIC would be 12 or higher.) AIC interpretation is necessary when

---

[1] In multidimensional space there exist near-infinite sets of singular points of relationship (singularities) of all points to each other and these relationships approach totality as more of the singularities are included. Obviously complete specification of all singular relationships in multivariate space is uninformative. The other end of the spectrum is complete atomization of all relationships, which is equally uninformative. Increasing AIC eventually leads to complete specification of the data (the opposite of atomization). To use an analogy, consider that there exists one optimal trail through a forest and the use of the AIC can be thought of as a way to establish that trail, with the realization kept in mind that the number of trails that approximate perfection is nearly infinite (for example, one might vary from the absolute optimal trail in a dense forest by going around one tree or not).

examining large datasets because it allows one to extract the maximal regression fit in as few steps as possible.

The r^2s of .555 and of .628 indicate good explanatory fits, with the model being a slightly better fit for the 3-year data, as discussed in the caption. A sinusoidal regression line (the red line) using the cubic model at a 95% confidence level shows the locations of the tightest relationships between energy (amplitude) and the frequency bin as sampled and represents the ebb and flow of energetic peaks inherent in the data. The 95% confidence level indicates that 95 out of a hundred times the sinusoids are an exact representation of the formant peaks. The data are three-dimensional and the plots are in two dimensions, so the peaks in the sinusoids do not exactly match the peaks in the plots of the data points. Though slightly different from each other, the sinusoids have a similar shape at 1 year and 3 years in these two graphs of Diné vowels in the aggregate

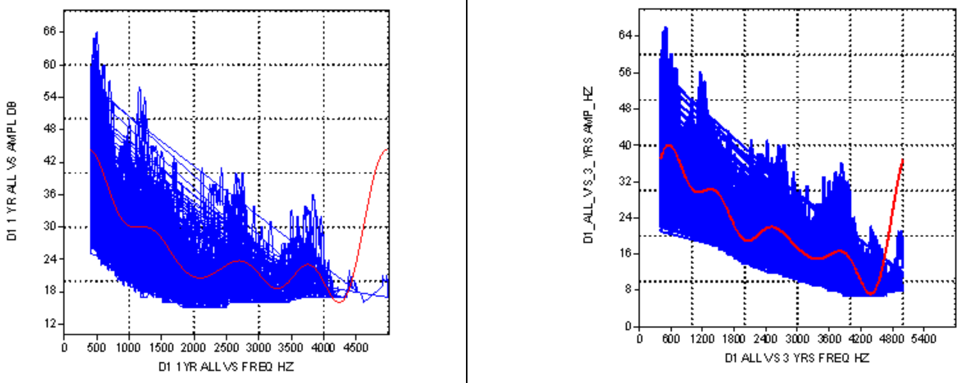


Figure 5-4.The graphs display the filtered data for all vowels collected from D1 at 1 and 3 years with sinusoidal regression lines (red) modified by the AIC procedure, 5 steps. AIC = 5.843 in both procedures. The r^2 for the 1-year data is .555 and for the 3-year data is .628. The significantly better fit at 3yrs indicated by the higher r^2 demonstrates increasing discrimination of formant peaks with age.

Based on this extension of least-squares regression into multidimensional space, sinusoidal modeling creates representations separating informative sinusoids from non-informative noise (Hammer, Harper & Ryan 2001). The formant peaks in the sinusoids are the informative three-dimensional patterns emergent from the data.Visual inspection of the sinusoidal cubic regression lines in Figure 5-4 suggests that the most difference in the data for the two different age-specific filters occurs in the region between 500 Hz and 1200 Hz.

An informal inspection of the filtered data shows a general pattern of increasing sensitivity as the temporal bones and the conductive apparatus housed therein mature. This increase in sensitivity is reflected in the increasing number of frequency bins that make it through cut-offs for each frequency in the age-specific filters. Figure 5-5 summarizes this pattern for all vowels measured in English; Figure 5-6 summarizes for Diné (see Appendix 5-A for tables of all data).

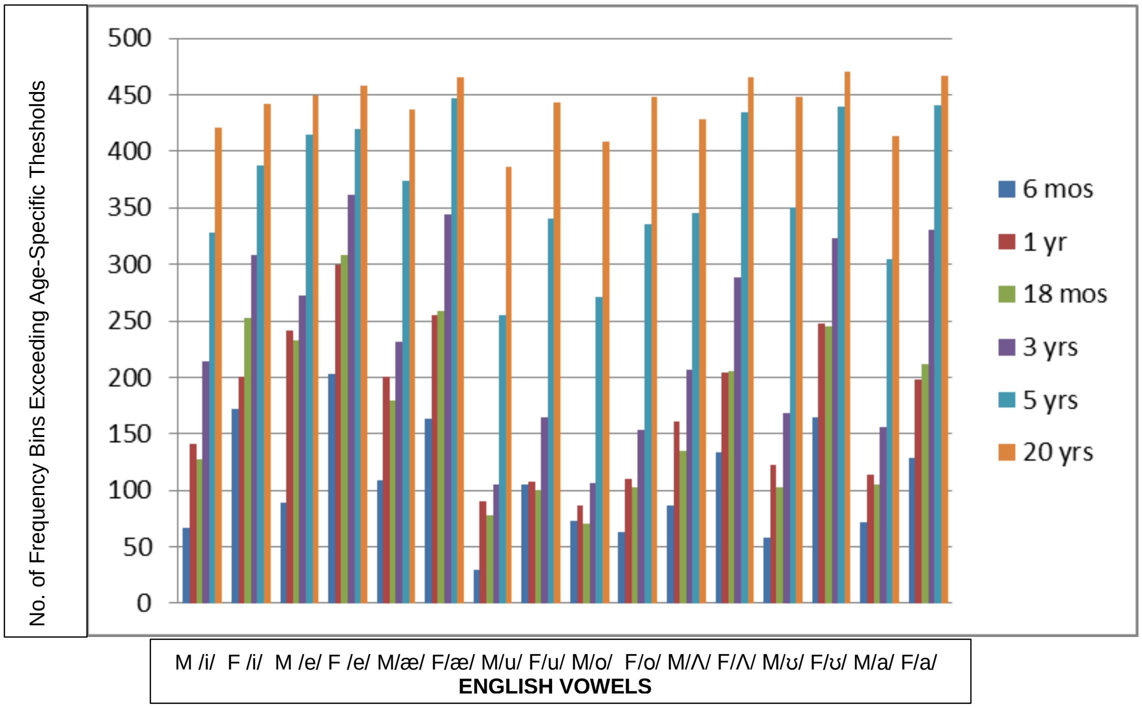


Figure 5-5. Pattern of Number of Frequency Bins in English Vowels Surviving Age Filters

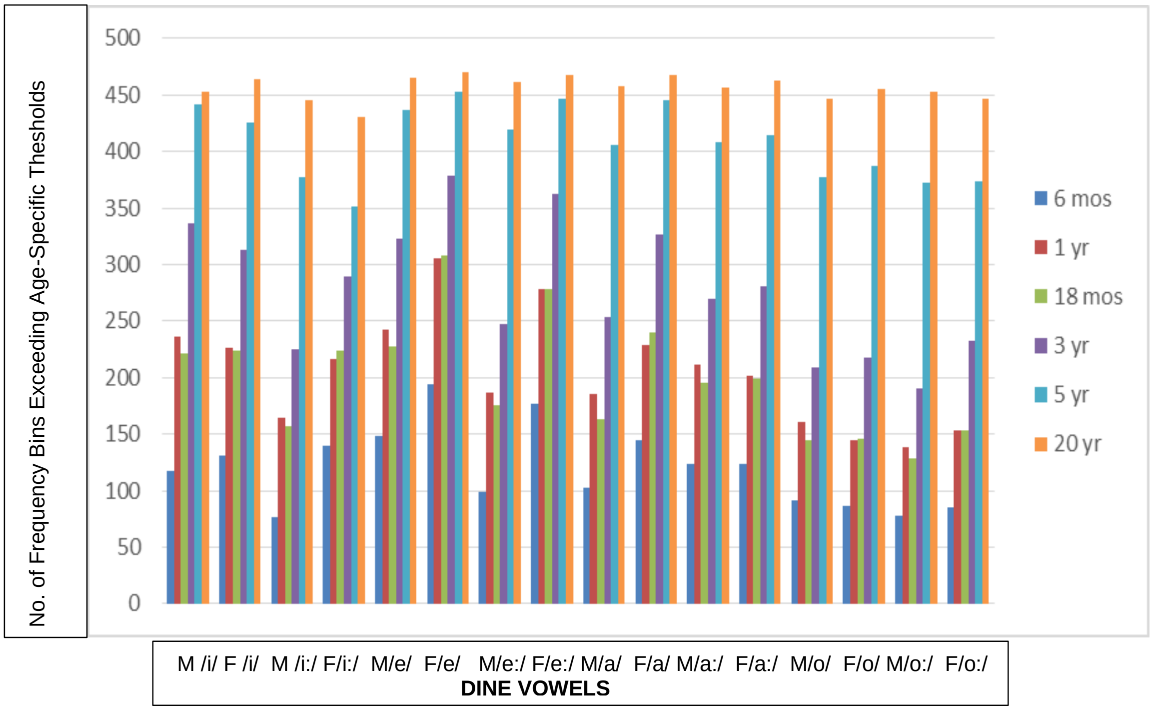


Figure 5-6. Pattern of Number of Frequency Bins in Diné Vowels Surviving Age Filters

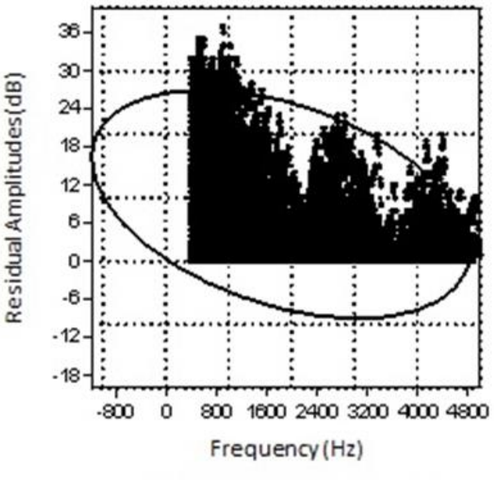
As the temporal bone develops and the conductive apparatus housed in the external and middle ear matures, children have greater access to more frequencies. However, at 18 months there is little to no gain (and indeed, a slight reversal in mid to back vowels) in terms of general overall sensitivity to vowel frequencies in male voices in both English and Diné (also see Tables 5A-1 and 5A-2 in Appendix 5A). In the female English speaker (E2), the 18-month filter affects the high front vowel /i/ such that there is a sizeable gain in sensitivity (51 more bins made it through the filter than at age 1 year), with modest gains for /e/, /u/, and /o/. The 18-month filter's effect is negligible in E2 in terms of increasing or decreasing the number of bins for /æ/, /a/, /ʌ /, and /ʊ/, located lower or more centrally than the vowels showing gains. In the female Diné speaker (D2), the 18-month filter affects /a/ the most dramatically with an increase of 11 more bins than at 1 year, followed by /i:/

with a moderate increase of 7 bins. The filter’s effect seems negligible in D2 for all other vowels. Also interesting is that at age 3 years, /i/ in both male voices showed the greatest gain in number of bins, while E2’s voice showed the greatest gain at /a/ and D2’s voice, at /i/. Perhaps it was adaptive in humans’ evolution for the immature to sacrifice great frequency discrimination and the linguistic detail of the configurations of F1, F2, and F3 for a more general picture of male voices, one that would carry the information present in F0 (e.g., the paralinguistic information available there about individual identity, social and regional groupings, and mood) rather than the complete configuration of F1, F2, and F3, at least initially. This finding might also reflect an earlier role for females in helping children to develop language.

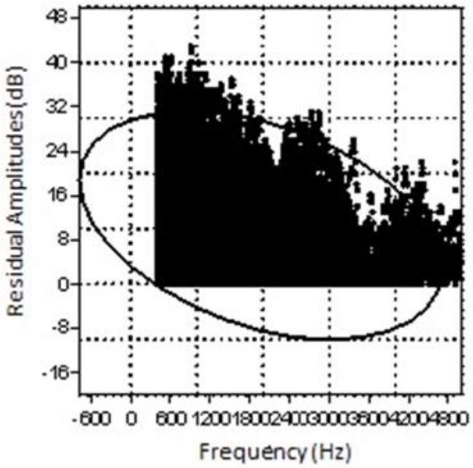
5.2 AGE EFFECTS ON ENGLISH AND DINÉ VOWELS IN AGGREGATE

Figures 5-7 and 5-8 show plots of all vowels included in this study in each of the two languages, English and Diné. (These plots were drawn in PAST and include English /i/, /e/, /æ/, /a/, /ʌ/, /o/, /ʊ/, and /u/ and Diné /i:/, /i/, /e:/, /e/, /a:/, /a/, /o:/, and /o/ from both pairs of speakers.) The two languages look generally different from one another at every age. These results are not surprising because these are two different vowel systems (nonlinguists should recognize that, for example, an /e/ in English is phonetically distinct from /e/ occurring in Diné). The ellipses indicates the 95% confidence interval, meaning 95% of the vowel's patterned information is inside the circle. Formant peaks are the 5% of the information that falls outside the circle, being the most salient part (or the most unusual part) of the formants. (A 95% confidence ellipse is usually considered a standard analysis ($\alpha < .05$) and is used to indicate the reliability of an estimate).

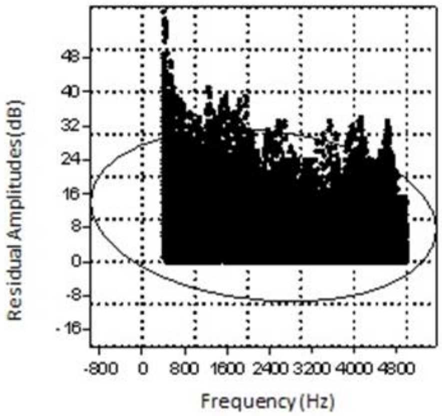
Beyond the fact that, as expected, these two languages are visually distinct is that the English vowels group into three distinct and broad formants at every age except 18 months, while Diné vowels group into four narrower regions of formants (recall Figure 5-4). At 18 months, Diné vowels recapitulate a pattern seen at 6 months, where the second aggregated formant is less intense than at other ages and includes a narrower band of frequencies than seen at 1 year, 3 years, 5 years, and 20 years. Analytically this difference between English and Diné derives from English including more steady-state vowels than Diné. In other words, the vowel space of English is more densely populated that that of Diné (see Figure 4-1). Age 18 months is the age at which the great majority of human children become dedicated bipedalists, and this event may be at the root of this effect.



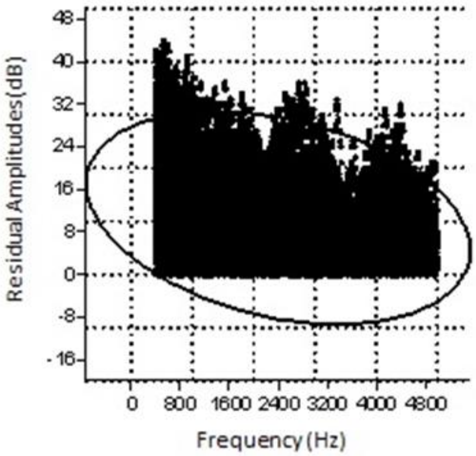
English All Vowels 6 Months



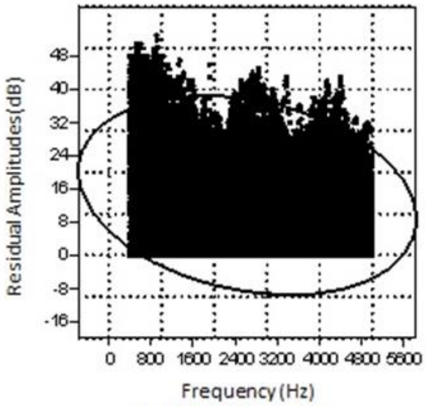
English All Vowels 1 Year



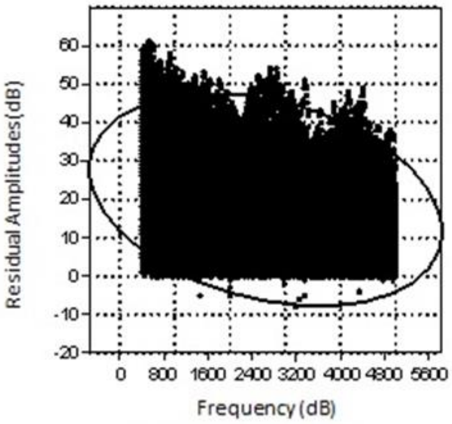
English All Vowels 18 Months



English All Vowels 3 Years

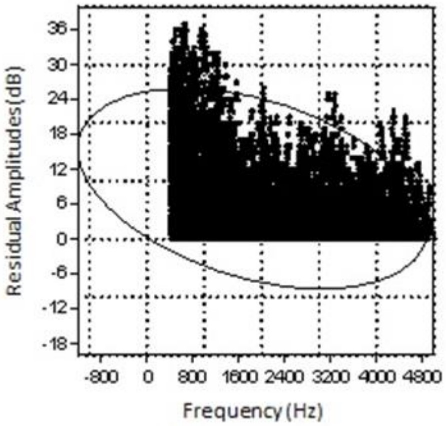


English All Vowels 5 Years

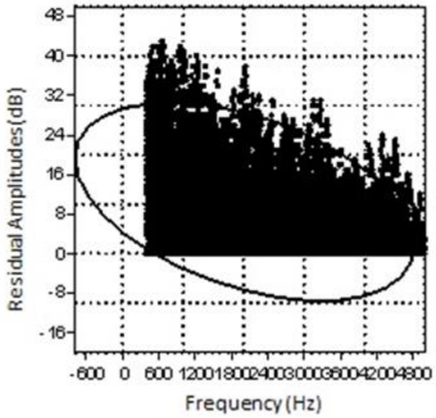


English All Vowels 20 Years

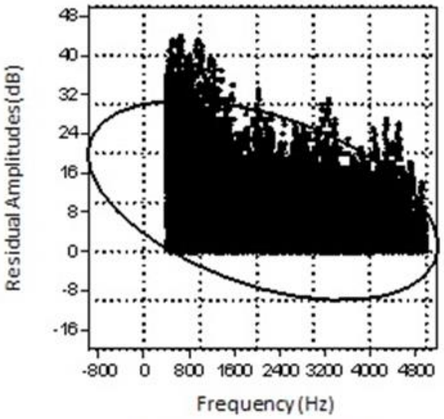
Figure 5-7. All Vowels of English Partitioned by Age



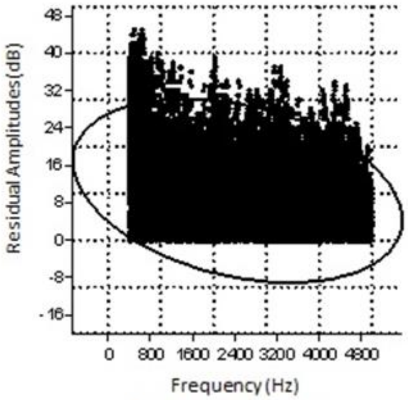
Diné All Vowels 6 Months



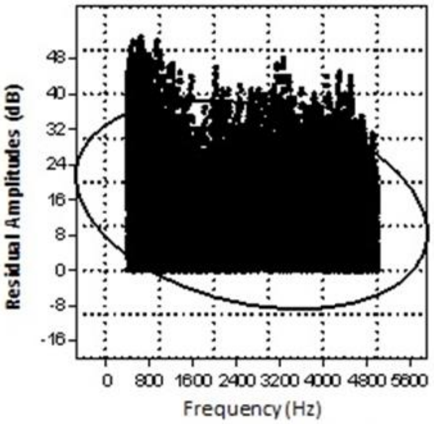
Diné All Vowels 1 Year



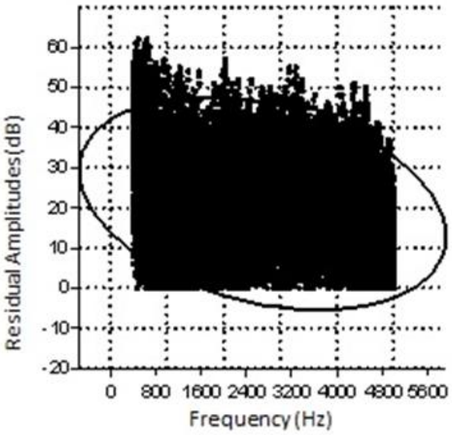
Diné All Vowels 18 Months



Diné All Vowels 3 Years



Diné All Vowels 5 Years



Diné All Vowels 20 Years

Figure 5-8. All Vowels of Diné Partitioned by Age

5.3 DINÉ VOWELS

An investigation of Diné vowels was carried out to determine if, as reported in McDonough (2003), that Diné /i/ and /i:/ differ not only in lengthening but also in terms of quality (frequency pattern), to a greater extent than /a/ and /a:/; /o/ and /o:/; and /e/ and /e:/, and to determine if there is a training effect at young ages created by the pattern of children's sensitivities to frequencies. As McDonough points out, the difference between /i/ and /i:/ is audible enough that native speakers are aware of it, unlike the other pairs of oral short and long vowels. McDonough's study is a meticulous investigation of the Diné sound system, and the study of vowels includes data from 10 female speakers and 4 male speakers who are bilingual speakers, who live on the reservation, and who use Diné daily. Results from this dissertation's study confirm McDonough's findings about the quality of /i/ and /i:/, even for in small samples. This lends credence that this dissertation's findings can be interpreted with confidence even with the even smaller sample used here; results exhibit robusticity comparable to that of McDonough; sufficient for the purposes of this study.

Figure 5-9 (constructed in Minitab) overlays the formant patterns of /i/ and /i:/ at 6 months, and the vowels clearly do not match up even at that age of less discriminated hearing. Figure 5-10 draws sinusoidal regression lines on the plot displaying the 20-year model of the /i/ and /i:/ data, and the shapes of the line are very different from one another indicating a difference in formant pattern (thus vowel quality), especially as compared with the same plots drawn for /a/ and /a:/ in Figures 5-11 and 5-12.

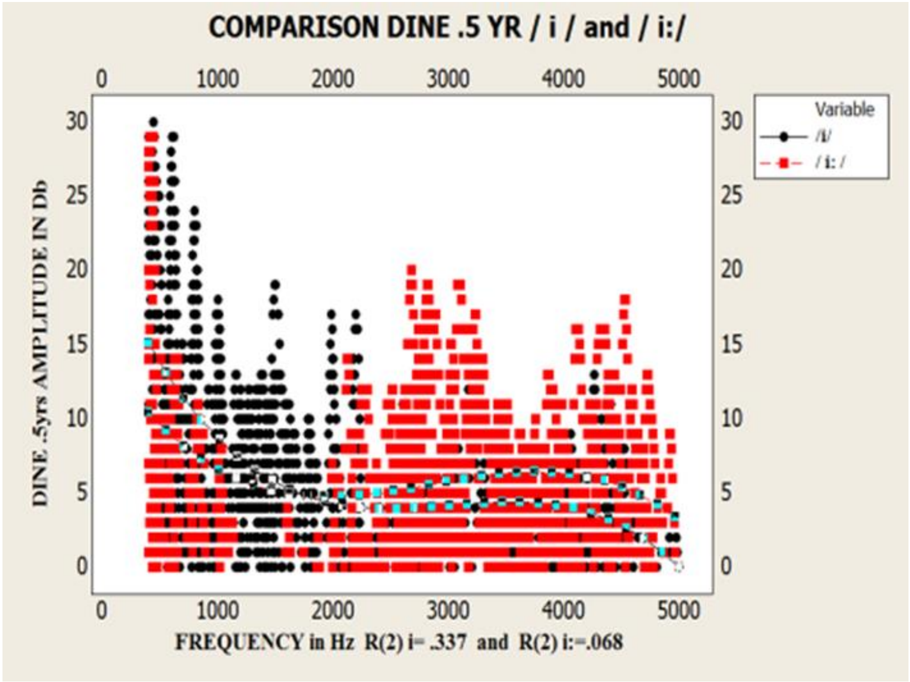


Figure 5-9. Comparison of Diné /i/ and /i:/ Reveals a Mismatch in the Formants Falling between 400 Hz and 2000 Hz at 6 Months

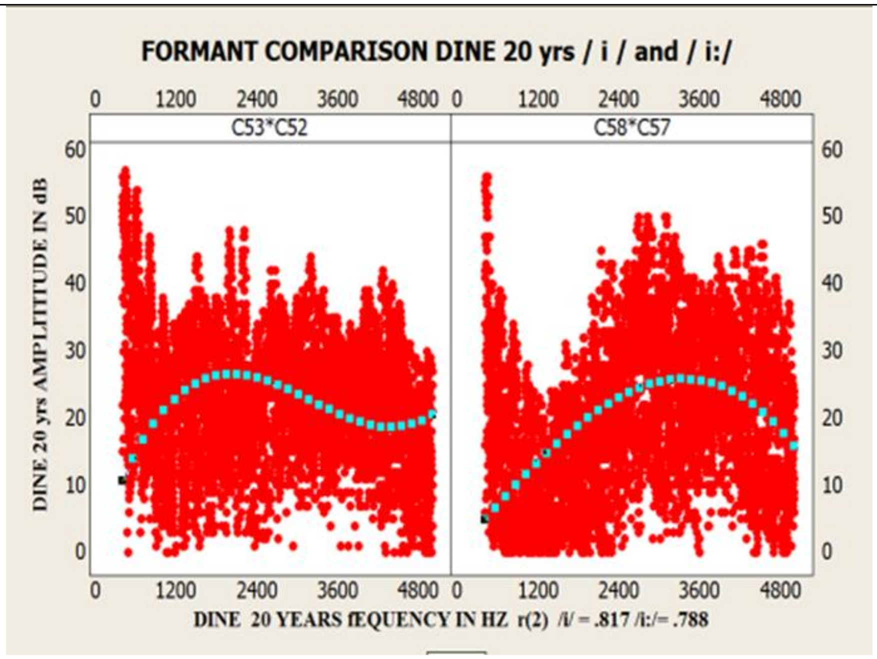


Figure 5-10. The Regression Lines Showing a Different Formant Profile for /i:/ and /i/ at 20 Years of Age

A t-test on /i/ and /i:/ at 6 months yielded a p-value of .0024 (for the chance that those vowels are the same). The r^2 in a cubic analysis on the 6-month-old model of /i/ is .337, and the r^2 for /i:/ is .168, neither of which being explanatory. The regression lines trace the path of the relationship of the frequency and amplitude, and the regression lines at 6 months for /i/ and /i:/ shows the tightest relationship in the region of F1. In other words, at 6 months, the relationship in the region of F1 is the most explanatory relationship for both /i/ and /i:/.

Interestingly, the p-value from the t-test that /i/ and /i:/ are the same vowel for the 20-year-old model is .007, indicating that at age 20 years there is again very little chance that the vowels are the same. The r^2 on the 20-year-old model of /i/ is .817 and for /i:/, the r^2 is .788, a negligible difference in terms of the lines' explanatory power. However, the shapes of these sinusoidal regression lines confirm that these are two different vowels, and that the relationship in the region of F2 and F3 is the most explanatory relationship. So, in terms of training effects, at age 6 months children appear to discriminate /i/ and /i:/ on the basis of F1 while at age 20 years they appear to rely heavily on F2 and F3.

Figures 5-11 and 5-12 plot /a/ and /a:/. At 6 months, the two vowels plot one on top of the other, and the shapes of their sinusoidal regression lines are similar. A t-test to test if /a/ and /a:/ are the different vowels yielded a p-value of .030. The r^2 for the 20-year-old model of /a/ was .667, while the r^2 for /a:/ was .674, indicating similar degrees of explanatory power for the sinusoidal regression lines which confirm that /a/ and /a:/ are the same vowel. However, the p-value for the probability that the 20-year-old model of /a/ and /a:/ are different vowels falls to .041. Thus the vowels /a/ and /a:/ are more similar at 6 months than at 20 years.

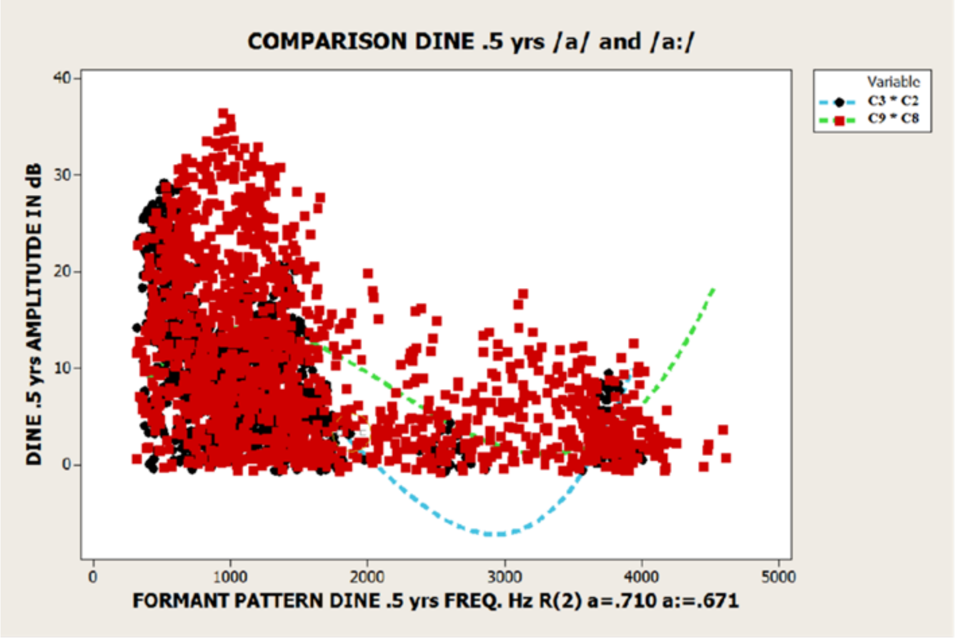


Figure 5-11. Formant Patterns for Diné /a/ and /a:/ Match Well at 6 Months

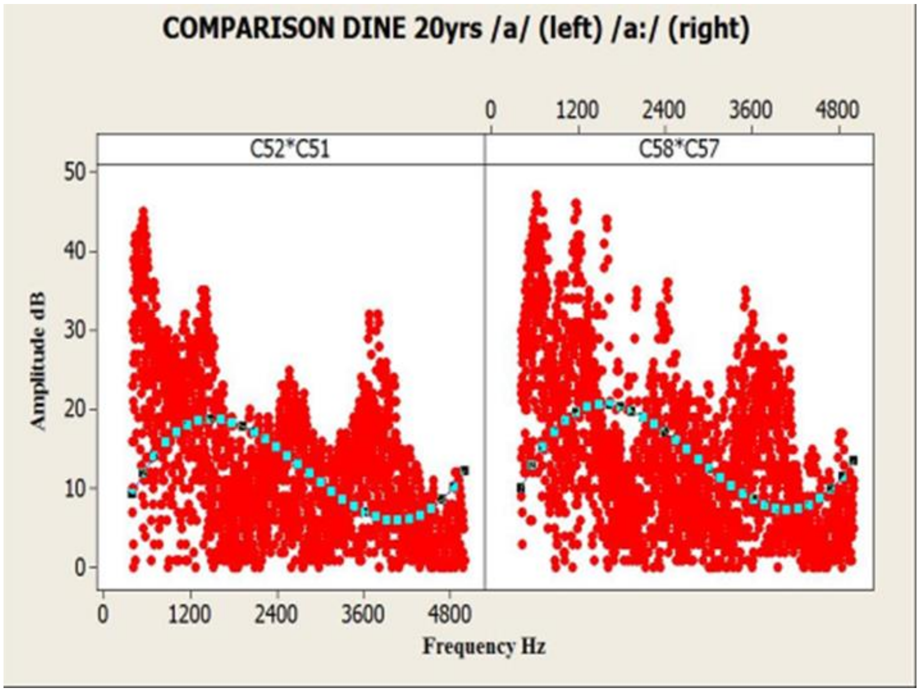


Figure 5-12. Regression Lines Showing the 20-Year Filtered Data for /a/ and /a:/ Have the Same Shape

5.4 AGE EFFECTS PARTITIONED BY SEX

The fundamental frequencies of male voices in this study are around 100–150 Hz lower than female voices, and since the filters for the frequencies do not change for the male and female inputs, we expect that children would experience the vowels produced by male and female vocal tracts differently. The general inspection of the number of frequency bins in Figures 5-5 and 5-6 as well as Tables 5A-1 and 5A-2 showed that, on balance, children experience more frequency information in the vowels produced by females than those by males. Adults generally experience less frequency information in female voices than in male voices because female voices are higher pitched so the harmonics are more widely spaced. However, the filters of immature ears subtract more frequencies from male voices than females until about age 5 years.

Nonetheless, while we expect children to have a different experience of male vowels as compared with female vowels, we would not expect that difference to affect their experience of vowel quality. In other words, we would expect that children experience the same configuration of a given vowel's formants whether that vowel was produced by a male or a female, albeit at lower frequencies for male voices, and whether that vowel was experienced by a 6-month-old or a 20-year-old, albeit more or less loudly. For example, Fry (1979) observes:

> Formant structure is important because of the role that it plays in the recognition and differentiation of speech sounds. We have seen that changes of fundamental frequency produce a shift in the exact location of the peaks in the spectrum because these are tied to the harmonics, but the formants, the true resonances of the vocal tract, will lead to spectral peaks in the same frequency region for a given configuration of the tract, regardless of changes of fundamental frequency. There are quite appreciable differences both in the range of fundamental frequencies and in the dimensions of the vocal tract as we go from one speaker to another, particularly as between men, women and children, but the general formant pattern enables listeners

> to recognize the “same” vowel when it is uttered by many different speakers. The vowel of *heed* will always have *F1* and *F2* widely spaced and in the vowel of *hard* they will be close together. (Fry 1979:78)

The initial question pursued in this section of analysis is whether or not the differential patterns of children’s sensitivities to frequencies is great enough to affect vowel quality, that is, whether the vowel of *heed* (a high front vowel) will always have F1 and F2 widely spaced and back vowels will always have F2 and F3 widely spaced (see Figures 5-13 and 5-14). The results indicate that the pattern of auditory sensitivity at each age does not at any point completely filter formants (the regions of greatest intensity) from the input signal. However, the signal does undergo simplification and the models do highlight some formants differently as compared with their adult forms.

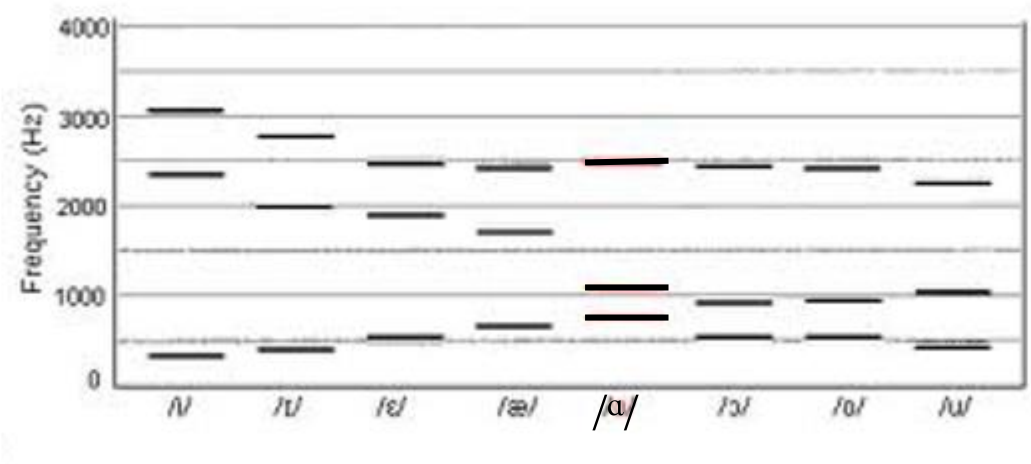


Figure 5-13. General Pattern of Average Formants in a Male English Speaker (from Lynn & Garn-Nunn 2004: 124)

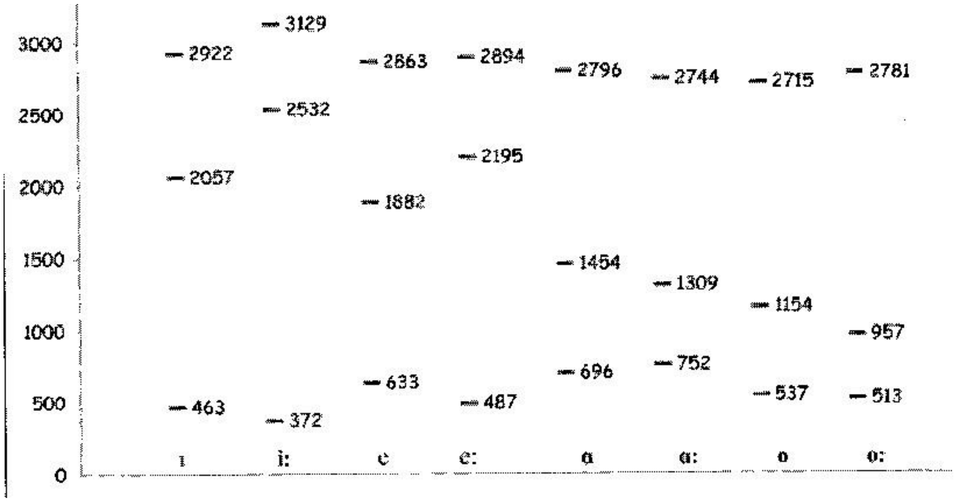


Figure 5-14. General Pattern of Average Formants in Diné speakers (from McDonough 2003: 118)

Figures 5-15 through 5-20 examine a selection of the data within each language partitioned by males and females. The peripheral vowels depicted in these plots are a compilation of six repetitions (two repetitions of three words containing the vowel), and, as in earlier sections, the amplitudes reported on the y-axis represent the amount by which the amplitude for a given frequency bin exceeded the sensitivity thresholds in each age's model. F1, F2, and F3 have been labeled with the aid of the tables in Appendix 5-B, which report the average formants for F1, F2, F3, and F4 obtained from the Get Formant command in Praat. Another point to bear in mind when viewing these plots is that these models are based on research that provided well-established sensitivity data for frequencies starting at 400 Hz; however, E1 has a deeper than average voice and F1's peak falls below 400 Hz in /i/ and /u/ for that speaker.

The cubic regression lines are highly explanatory at the majority of ages for nearly every vowel (English /ʌ/ at 6 months for both male and female voices and English /u/ at 1 year and 1.5 years for the male voice are slight exceptions). The highest points of the colored regression lines show the range of frequencies for which the relationship between the frequency and the residual amplitude are the tightest—where frequencies have been completely filtered out, the regression line is black and is continuing its trajectory from the preceding frequencies into the next. Interestingly, the r^2 shows a tendency to rise as the filters move toward age 20 years, indicating a better fit for the regression lines at older ages, except for a slight dip at 18 months. The effect of frequencies being completely dropped or greatly reduced in amplitude at younger ages as compared to their adult profile is at the root of the difference. This effect may reflect a learning strategy in the earlier years (18 months and younger, before humans become dedicated bipedalists) that is more focused on individual formants rather than a strategy focused on the overall shape of the vowel. Swingley (2005), for example, is among linguists who posit that children switch learning during the acquisition of their native language(s).

In English /i/ and Diné /i:/, which are similarly high front vowels, we see simplification of the frequency information between F1 and the rest of the formants (see Figures 5-15 and 5-16). In the model of 6-month hearing, the frequencies between F1 and the other formants are filtered out, in both languages and in all speakers, and, if not filtered out entirely, in the models of 1 year, 18 months, and 3 years the frequencies between F1 and the other formants are diminished. In the male input, F2 and F3 are additionally given more prominence relative to F1 starting at 1 year and continuing through 5 years.

As a rule, back vowels such as English /u/ and Diné /o:/ are generally inherently less energetic than front vowels, although the data used in this study don't show this tendency, probably because they were elicited in a word list (however, infant-directed speech is usually more similar to a citation style of speech than it is to usual adult conversation.) Figures 5-17 and 5-18 show that both these vowels in male and female voices are greatly affected by the 6-month-model with the frequencies between F2 and F3 being completely filtered out. In back vowels, F1 and F2 are close together. In the auditory models, there is a slightly bigger trough between F1 and F2 in the 6-months model of both English /u/ and Diné /o:/ in the female voices than there is at age 20 years, but for male voices this effect of a deeper trough between F1 and F2 is decidedly more marked, and appears up until age 5 years. This effect of deeper troughs in occurs between F3 and F4 in both males and females until age 20 years.

Lacerda (1994) studied the perception of infants aged 6 months to 1 year of variants in a schwa vowel and concluded there is a perceptual asymmetry favoring F1 differences during the early stages of vowel perception. The auditory models of those ages for English /ʌ / and Diné /a:/ show that F1 is generally more prominent than other formants especially in English /ʌ/ (see Figures 5-19 and 5-20). Similar to back vowels, there is also separation of F1 and F2 from F3 and F4 particularly in the male voices and most markedly at 6 months and again at 18 months.

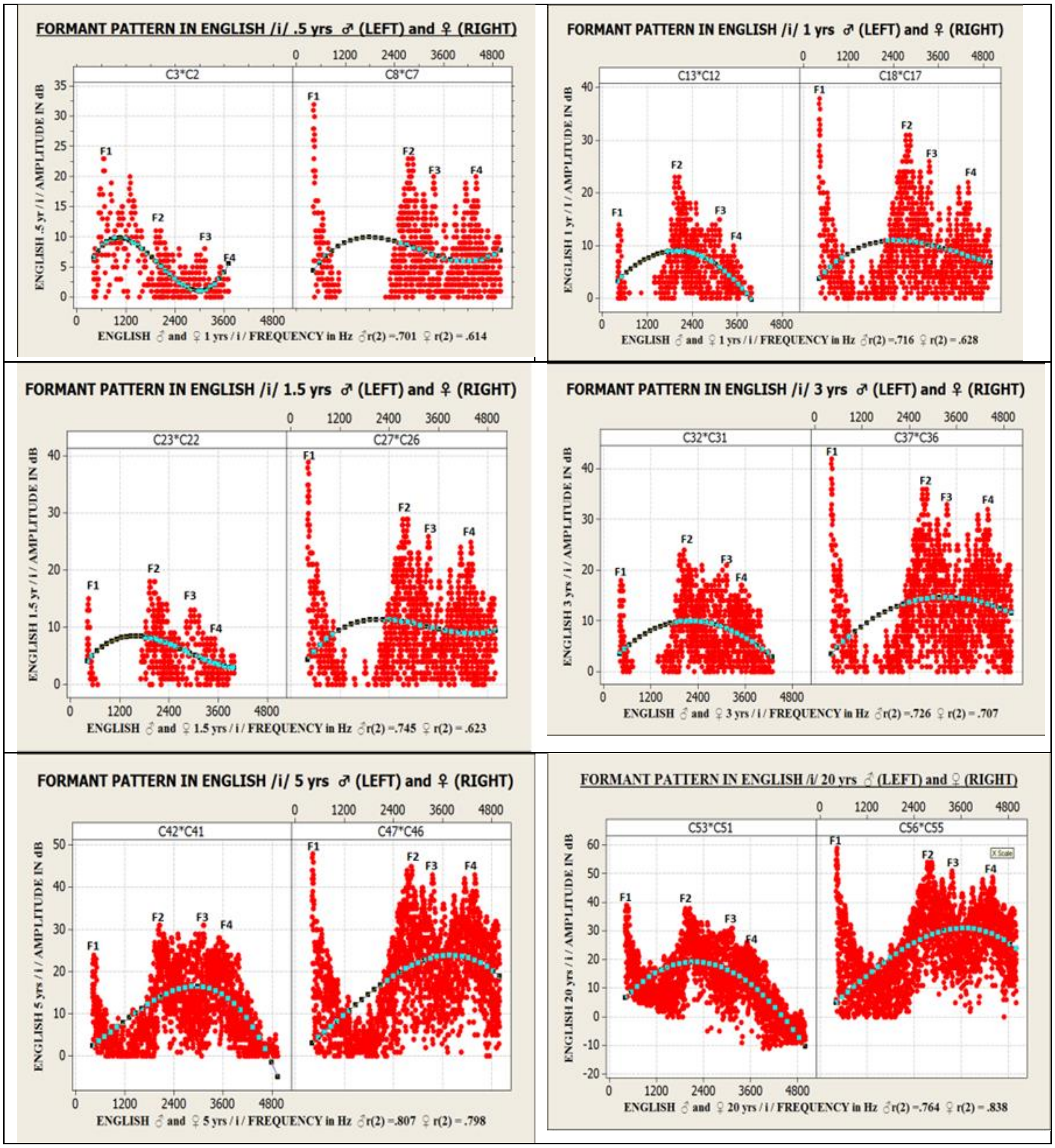


Figure 5-15. The English Vowel /i/ Partitioned by Males and Females with Regression Lines

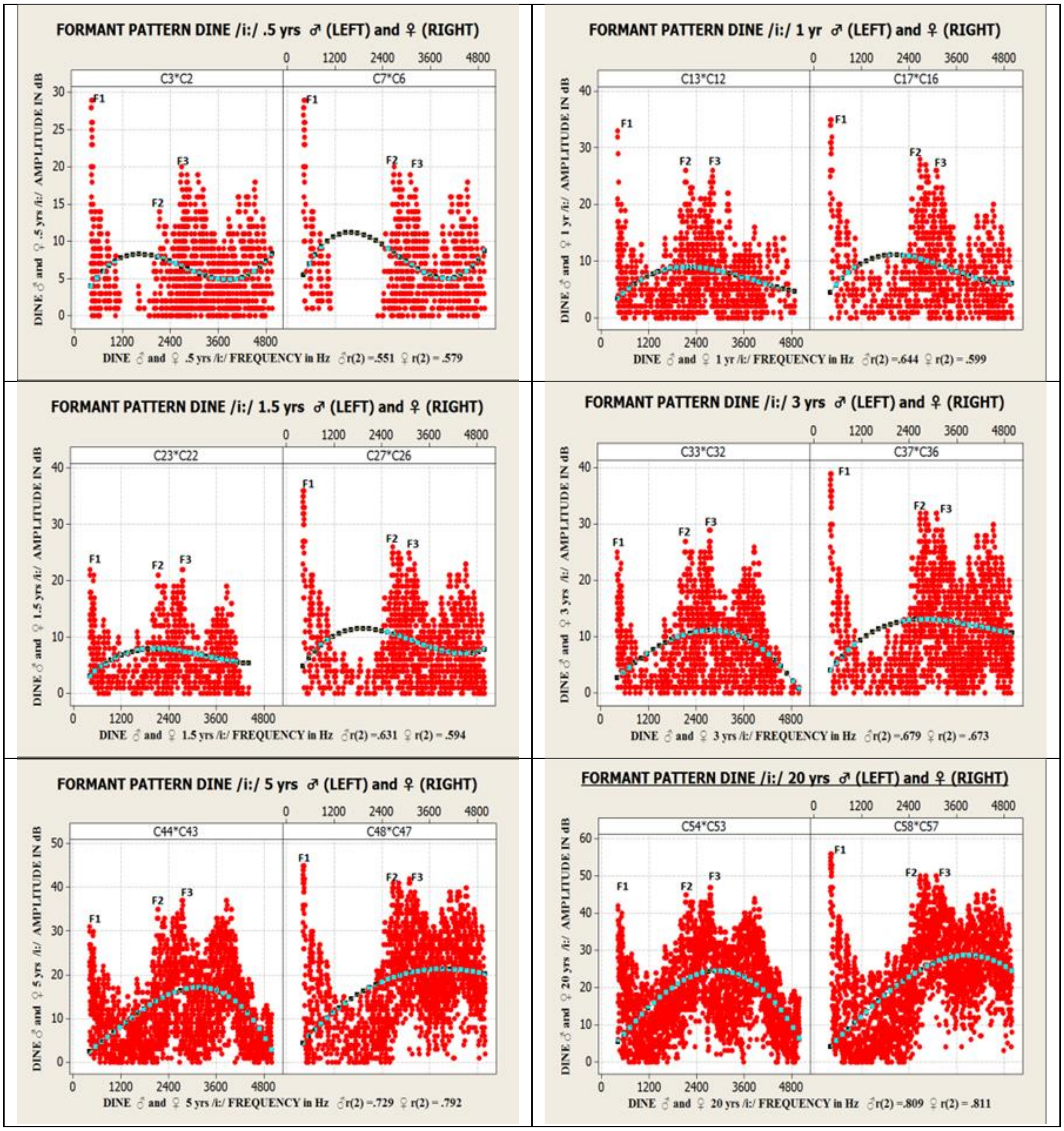
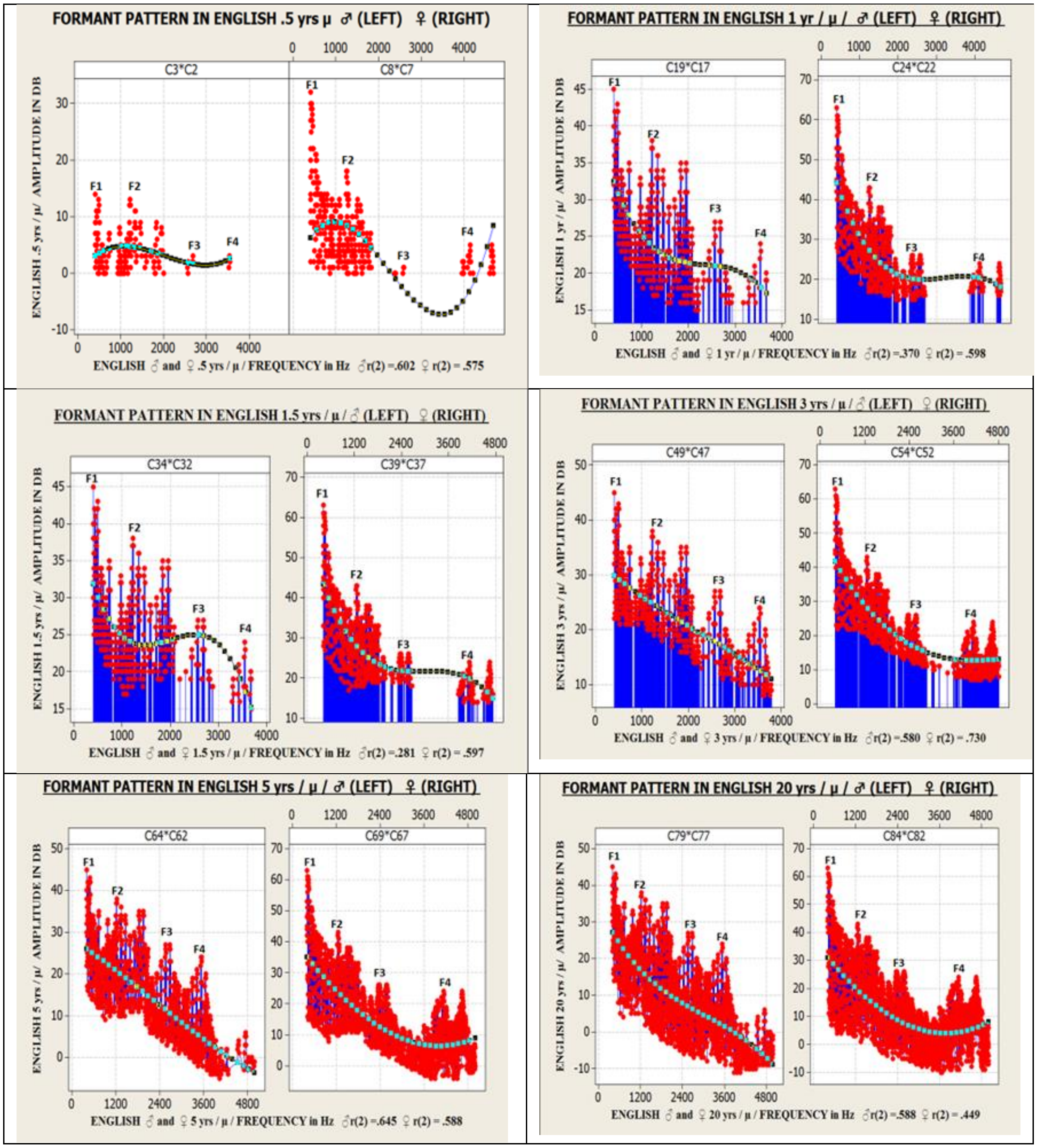


Figure 5-16. The Diné Vowel /i:/ Partitioned by Males and Females with Regression Lines


FORMANT PATTERN IN ENGLISH .5 yrs µ ♂ (LEFT) ♀ (RIGHT)

FORMANT PATTERN IN ENGLISH 1 yr / µ / ♂ (LEFT) ♀ (RIGHT)

FORMANT PATTERN IN ENGLISH 1.5 yrs / µ / ♂ (LEFT) ♀ (RIGHT)

FORMANT PATTERN IN ENGLISH 3 yrs / µ / ♂ (LEFT) ♀ (RIGHT)

FORMANT PATTERN IN ENGLISH 5 yrs / µ / ♂ (LEFT) ♀ (RIGHT)

FORMANT PATTERN IN ENGLISH 20 yrs / µ / ♂ (LEFT) ♀ (RIGHT)

Figure 5-17. The English Vowel /u/ Partitioned by Males and Females with Regression Lines

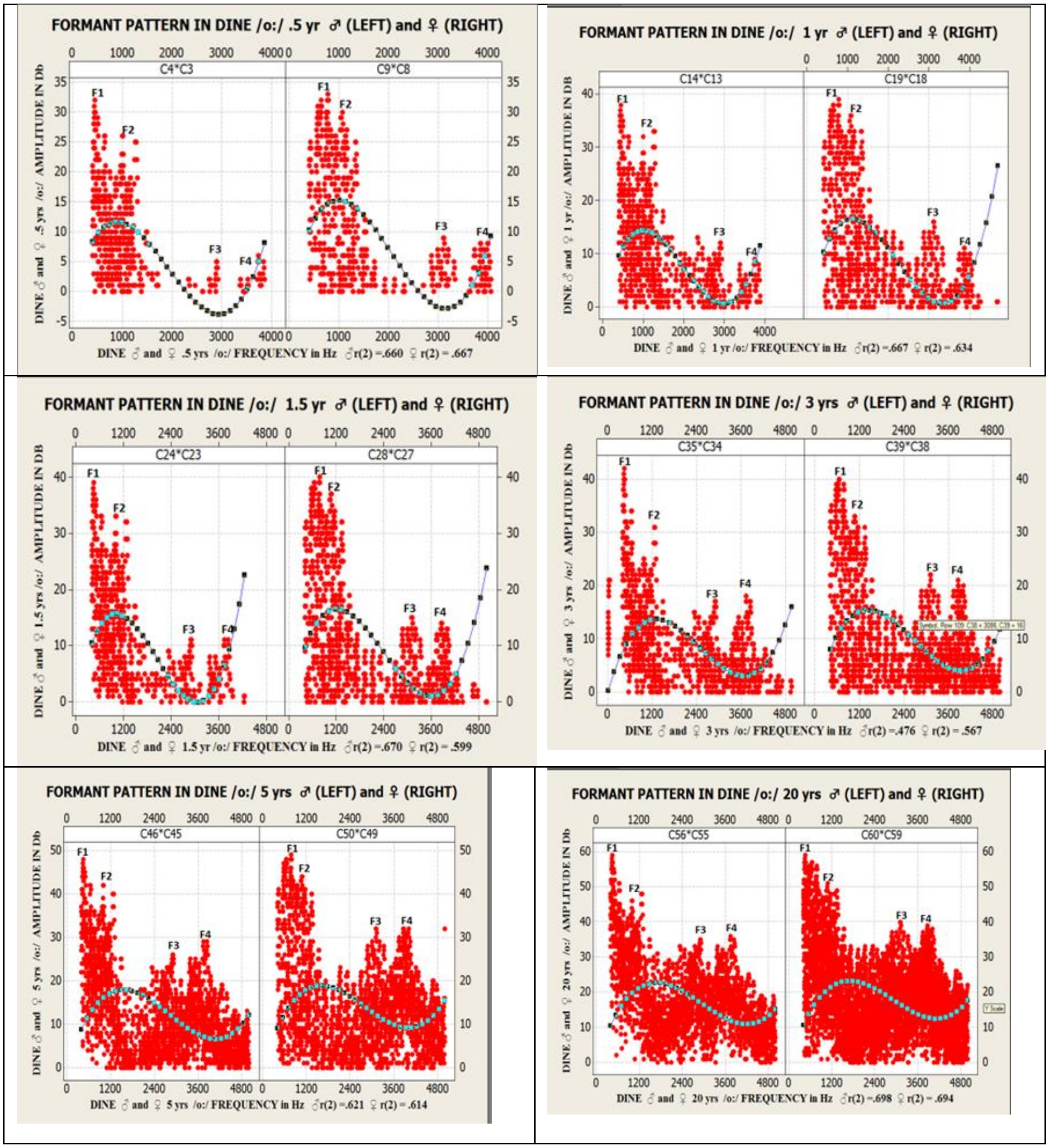


Figure 5-18. The Diné Vowel /o:/ Partitioned by Males and Females with Regression Lines

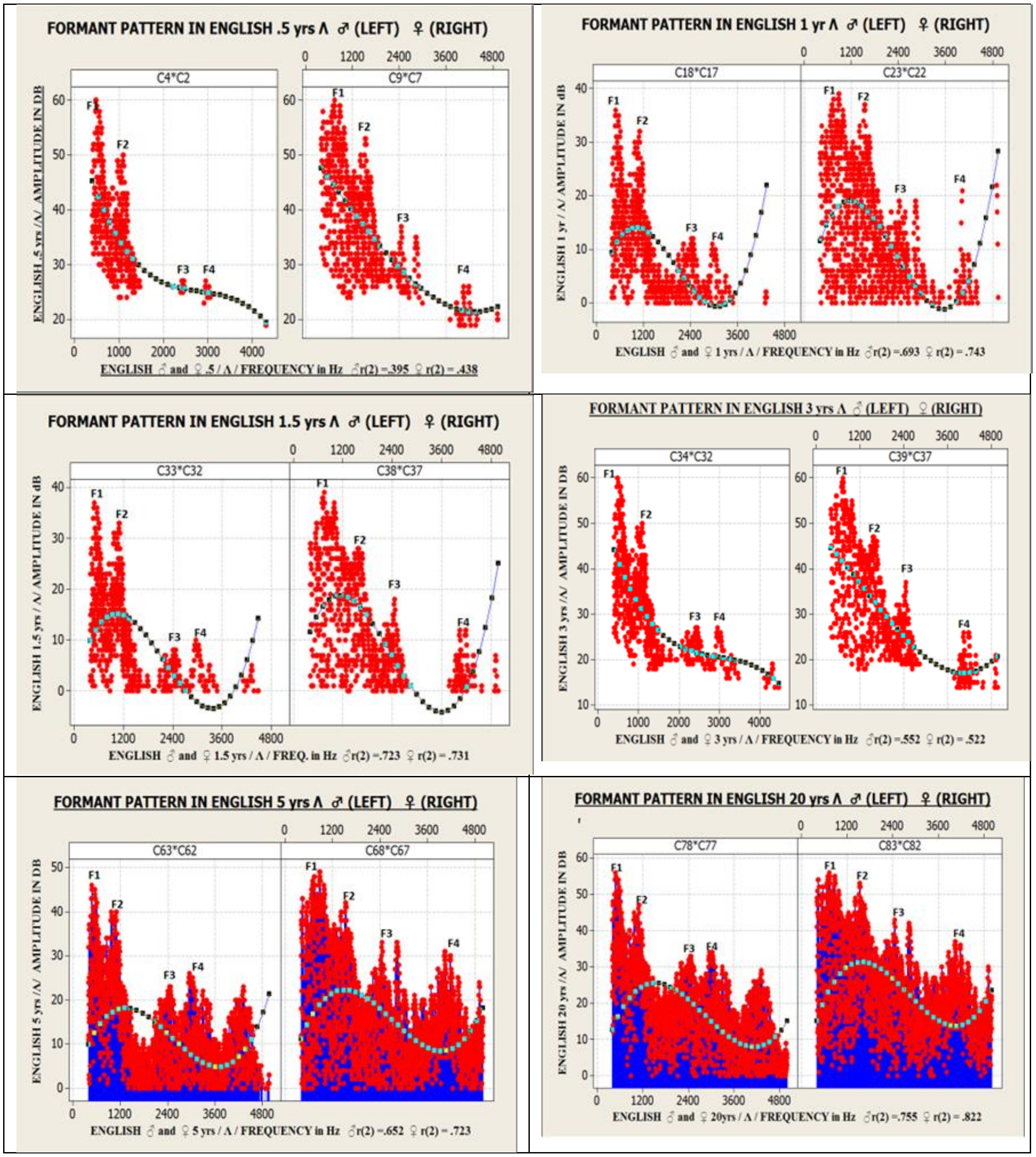


Figure 5-19. The English Vowel /ʌ/ Partitioned by Males and Females with Regression Lines

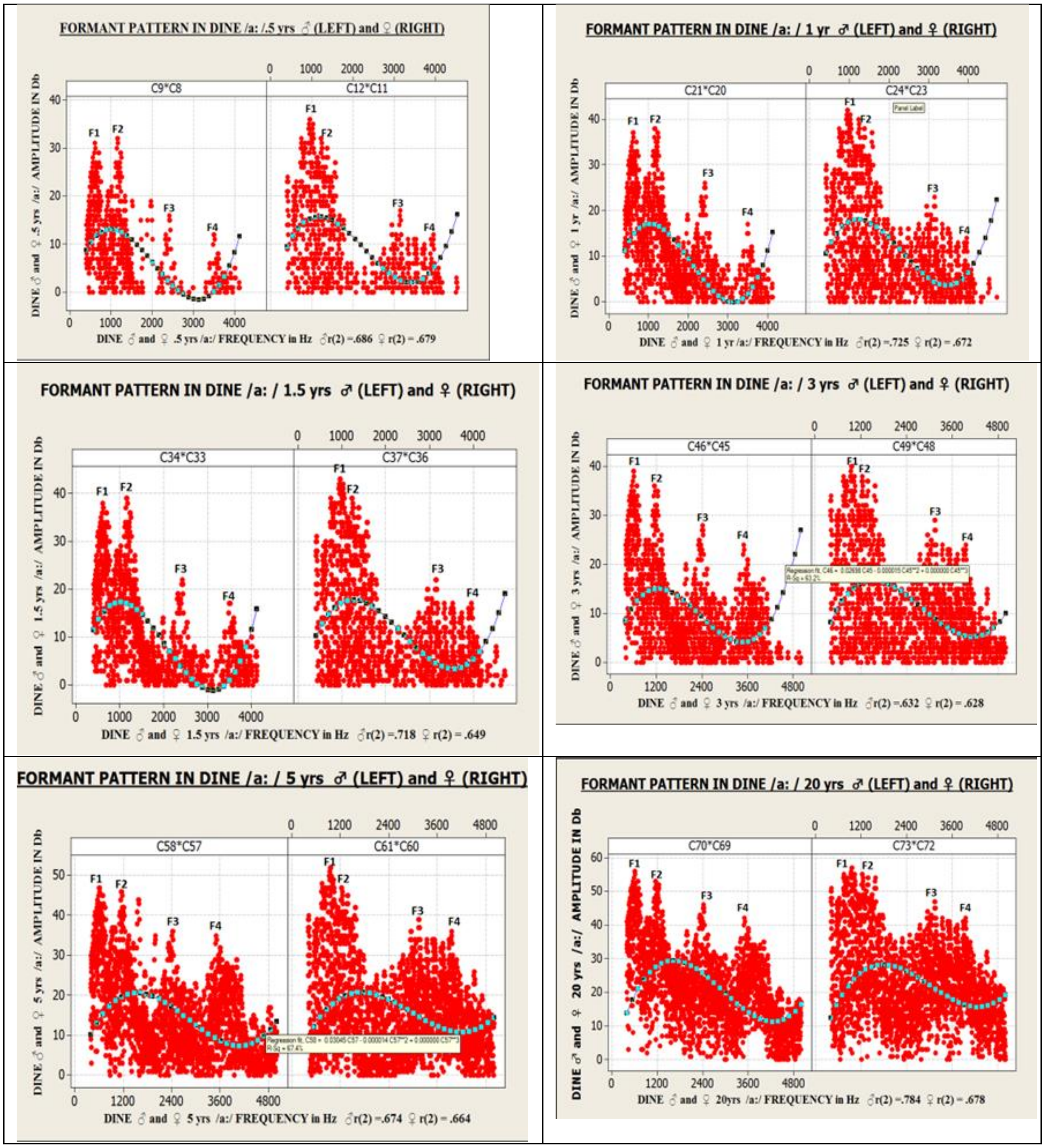


Figure 5-20. The Diné Vowel /a:/ Partitioned by Males and Females with Regression Lines

5.5 CONCLUSIONS ABOUT TRAINING EFFECTS

Chistovich (1985) and subsequent researchers have found that adult listeners perceive the first and second formants in back vowels as an integration of the two rather than as two separate formants, and in front vowels, they perceive the second and third formants as an integration. This investigation did not find that there was a training effect at younger ages such that listeners learn to integrate these formants according to their immature perception so much as that the auditory modeling of all ages tended to compress the formants that are close together in front and back vowels. However, this investigation does find a training effect in that formants are simplified at younger ages as compared to 20-year-olds because of the auditory filtering, specifically through the elimination of distracting detail or the highlighting of troughs between formants.

At 18 months, children lose a degree of sensitivity to the frequencies of male voices that they had at 1 year, which causes another period of time where formants receive highlighting and emphasis because of the filtering. For female voices at 18 months, there is not a gross loss of sensitivity as seen in male voices, but in front vowels F1 is set off from the other formants, in back vowels F1 and F2 are separated from other formants, and in mid-low vowels F1, F2, and F3 are set off from F4 and other formants. The pattern of sensitivity at 18 months is, in terms of general sensitivity, a time of more extreme filtering of male voices than occurs at 1 year and a shift of filtering of female voices in the F2 region (approximately 2–3 kHz) that in adults is enhanced by the external auditory meatus. At 18 months, children are beginning to produce two-word utterances from an approximately 200-word vocabulary (Boysson-Bardies 1999) in what Blount (1975) called “pivot grammars” (e.g., utterances like “more wet” and “allgone lettuce”). In terms of motor development by 18 months children are generally committed to bipedalism and learning how to run and move objects around. Obviously this apparent step backward in terms of gains in hearing sensitivity is an area ripe for further analysis and study.

APPENDIX 5-A

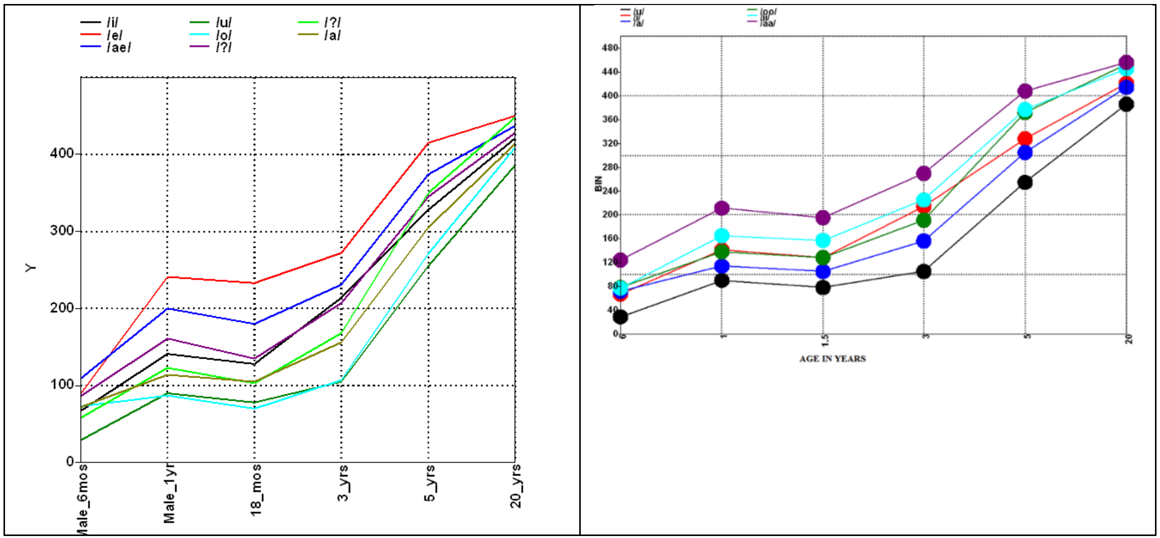
COUNTS OF FREQUENCY BINS FOR MALE AND FEMALE VOICES IN DINÉ AND ENGLISH

Table 5A-1. Average Frequency Bin Counts after Filtering for Each Age for Each Vowel of English as a General Measure of Sensitivity

English Vowel	Male 6mos	Male 1yr	Male 18 mos	Male 3 yrs	Male 5 yrs	Male 20 yrs	Female 6 mos	Female 1 yr	Female 18 mos	Female 3 yrs	Female 5 yrs	Female 20 yrs
/i/	67	141	128	214	328	421	172	201	252	308	387	442
/e/	89	241	233	272	415	450	203	299	308	361	420	458
/ae/	109	200	180	231	374	437	163	255	259	344	447	466
/u/	29	90	78	105	255	386	105	108	100	165	340	443
/o/	73	87	70	107	271	409	63	110	103	154	336	448
/ʌ/	86	161	135	207	345	428	134	204	206	289	434	465
/ʊ/	58	123	103	168	350	448	165	247	245	323	439	470
/a/	72	114	105	156	305	414	129	198	212	330	441	467

Table 5-2. Average Frequency Bin Counts after Filtering for Each Age for Each Vowel of Diné as a General Measure of Sensitivity

Diné Vowel	Male 6mos	Male 1yr	Male 18 mos	Male 3 yrs	Male 5 yrs	Male 20 yrs	Female 6 mos	Female 1 yr	Female 18 mos	Female 3 yrs	Female 5 yrs	Female 20 yrs
/i/	118	236	221	336	441	453	131	226	224	313	426	464
/i:/	77	165	157	225	377	445	140	217	224	289	351	431
/e/	148	242	228	323	436	465	194	305	308	379	453	470
/e:/	99	187	175	247	419	461	177	278	278	362	446	467
/a/	102	186	163	253	406	458	145	229	240	326	445	467
/a:/	124	211	195	270	408	456	124	201	199	281	414	463
/o/	92	161	145	209	377	446	86	145	146	218	387	455
/o:/	78	138	128	191	372	453	85	153	153	233	373	447



E1 Vowels:
Black = /i/
Red = /e/
Blue = /æ/
Olive green = /u/
Turquoise = /o/
Purple = /ʊ/
Neon green = /Λ/
Grey = /a/

D1 Vowels:
Black = /o/
Red = /i/
Blue = /a/
Olive green = /o:/
Turquoise = /i:/
Purple = /a:/

Figure 5A-1. This figure compares patterns of frequency bins in E1 and D1. The plots show the numbers of frequency bins that survive a 20-year filter are much more similar for all vowels than they are at different ages for different vowels.

APPENDIX 5-B

AVERAGE FORMANT FREQUENCIES OBTAINED WITH PRAAT'S GET FORMANT COMMAND

Diné /i/	dibé r1	dibé r2	tsin r1	tsin r2	łitso r1	łitso r2	Average /i/
D1							
F1	422	421	411	426	365	405	408
F2	1386	1313	1535	1499	1468	1348	1422
F3	2620	2643	2564	2721	2619	2629	2633
F4	3458	3423	3782	3565	3668	3783	3613
D2							
F1	575	539	709	664	547	574	601
F2	1568	1977	2202	2029	1809	1289	1812
F3	2466	3124	3243	3221	3136	3262	3075
F4	3526	4250	4367	4337	4378	3621	4080

Diné /i:/	ałk'iistł'in r1	ałk'iistł'in r2	neidiiłts'in r1	neidiiłts'in r2	nábidiiłne' r1	nábidiiłne' r2	Average /i:/
D1							
F1	362	373	283	290	307	300	319
F2	1692	1664	2257	2060	1991	2000	1944
F3	2519	2477	2699	2667	2714	2680	2626
F4	3857	2753	3677	3628	3730	3670	3553
D2							
F1	416	381	421	415	418	434	414
F2	1934	2767	2715	2223	2729	1202	2261
F3	2838	3230	3074	2807	2990	2946	2981
F4	3346	4414	3642	3227	3239	3531	3567

Diné /e/	bine'déé'o r1	bine'déé'o r2	kone' r1	kone' r2	nihinełí r1	nihinełí r2	Average /e/
D1							
F1	498	486	500	515	538	502	507
F2	1691	1490	1496	1341	1661	1551	1538
F3	2672	2764	2603	2793	2721	2562	2686
F4	3717	3719	3682	3534	3763	3762	3696
D2							
F1	779	748	763	765	735	772	760
F2	1931	1230	1594	1942	1922	1771	1732
F3	2140	2203	2106	3318	3389	2006	2525
F4	3345	3396	3367	4273	4313	3247	3657

Diné /e:/	náádeest'íí' r1	náádeest'íí' r2	deez'įį' r1	deez'įį' r2	héideez'íí' õr1	héideez'íí' õ r2	Average /e:/
D1							
F1	497	460	456	402	442	421	446
F2	1447	1484	1652	1651	1573	1555	1560
F3	2760	2582	2598	2639	2581	2599	2627
F4	4006	3774	3883	3861	3883	3895	3884
D2							
F1	697	741	654	663	651	627	672
F2	2057	1089	2096	1597	2056	1593	1748
F3	2088	2019	2671	2110	3059	1947	2316
F4	3331	3416	3306	3170	3261	3229	3286

Diné /a/	báhachį r1	báhachį r2	dah násk'id r1	dah násk'id r2	ch'il látah hozóón r1	ch'il látah hozóón r2	Average /a/
D1							
F1	492	465	554	510	527	500	508
F2	1445	1425	1414	1319	1155	1139	1316
F3	2514	2487	2534	2498	2290	2347	2445
F4	3520	3616	3844	3709	3545	3582	3636
D2							
F1	761	774	811	807	827	756	789
F2	1887	1767	1691	1769	1442	1378	1656
F3	3179	2938	2952	3169	2709	2637	2931
F4	4089	4182	3927	4241	3900	3901	4040

Diné /a:/	bighaa' r1	bighaa' r2	naaki r1	naaki r2	binaa' r1	binaa' r2	Average /a:/
D1							
F1	571	554	571	540	592	597	571
F2	1224	1156	1352	1349	1281	1165	1255
F3	2456	2282	2489	2594	2321	2525	2445
F4	3494	3493	3742	3668	3592	3795	3631
D2							
F1	844	882	914	894	890	821	874
F2	1297	1329	1486	1460	1281	1213	1344
F3	3148	2760	3113	2484	1348	2163	2503
F4	4092	2980	3732	3156	3688	3545	3532

Diné /o/	ałhosh r1	ałhosh r2	tsosts'id r1	tsosts'id r2	hozhõ r1	hozhõ r2	Average /o/
D1							
F1	433	446	437	428	437	577	459
F2	1081	1018	1152	1183	1508	1579	1254
F3	2529	2482	2681	2602	2759	2829	2647
F4	3301	3356	3667	3615	3328	3967	3539
D2							
F1	599	625	644	644	588	577	613
F2	1097	1150	1456	1447	1580	1579	1385
F3	2835	2314	3001	3225	2751	2829	2826
F4	3908	3416	3133	4220	3764	3967	3735

Diné /o:/	ałts'oozí r1	ałts'oozí r2	dootł'izhgo r1	dootł'izhgo r2	naaltsoos r1	naaltsoos r2	Average /o:/
D1							
F1	403	423	415	402	542	446	439
F2	1134	1010	1116	1215	1230	1034	1123
F3	2619	2704	2832	2744	2473	2611	2664
F4	3778	3515	3535	3520	3664	3533	3591
D2							
F1	632	629	659	669	618	642	642
F2	1106	1090	1290	1204	1086	1056	1139
F3	3134	3200	2875	2571	3001	2652	2906
F4	3917	3620	4049	3875	4033	3191	3781

English /i/	sees r1	sees r2	sheep r1	sheep r2	keep r1	keep r2	Average /i/
E1							
F1	337	306	277	346	267	267	300
F2	1961	2036	1951	2023	2069	2100	2023
F3	2590	2598	2584	2604	2471	2559	2568
F4	3620	3555	3312	3381	3461	3483	3519
E2							
F1	388	405	393	370	361	405	387
F2	2664	2634	2673	2532	2787	2653	2657
F3	2967	2899	3230	3113	3378	3339	3154
F4	4203	4378	4361	3520	3535	4331	4055

English /e/	shakes r1	shakes r2	takes r1	takes r2	taking r1	taking r2	Average /e/
E1							
F1	410	408	380	387	354	406	391
F2	1727	1800	1805	1807	1940	1905	1831
F3	2455	2590	2404	2425	2454	2482	2468
F4	3202	3317	3354	3332	3315	3373	3316
E2							
F1	531	528	504	530	436	401	488
F2	2323	2090	2531	2481	2592	2521	2423
F3	2495	2478	2916	2951	3043	2621	2751
F4	3023	2953	3221	4313	4309	2890	3452

English /æ/	back r1	back r2	nap r1	nap r2	mad r1	mad r2	Average /æ/
E1							
F1	605	596	618	632	566	568	598
F2	1295	1251	1350	1333	1333	1400	1327
F3	2200	2171	2356	2342	2311	2364	2291
F4	3305	3185	3084	2962	3110	3181	3138
E2							
F1	942	938	903	885	889	943	917
F2	1776	1784	1359	1983	1954	1817	1779
F3	2655	2619	2351	2818	2703	2756	2650
F4	4086	4197	3526	4300	4260	4227	4099

English /u/	food r1	food r2	mood r1	mood r2	rude r1	rude r2	Average /u/
E1							
F1	315	351	312	280	299	345	317
F2	1251	1344	1293	1422	1531	1538	1397
F3	2442	2425	2379	2211	2372	2503	2389
F4	3472	3435	3473	3178	3174	3478	3368
E2							
F1	413	389	368	361	418	372	389
F2	1259	1447	1242	1217	1519	1560	1374
F3	2560	2203	2265	2316	2444	2105	2149
F4	4423	4220	3970	3917	3804	3794	4021

English /0/	pokes r1	pokes r2	nose r1	nose r2	goes r1	goes r2	Average /o/
E1							
F1	413	421	383	435	431	415	416
F2	1018	1035	1216	1168	1274	1237	1158
F3	2416	2366	2489	2499	2458	2459	2448
F4	3443	3476	3400	3264	3422	3405	3401
E2							
F1	497	517	431	475	483	427	472
F2	1112	1128	1026	1071	1063	1050	1075
F3	2491	2226	2335	2255	2557	2608	2412
F4	4014	4089	3778	3926	4250	4184	4040

English /ʌ/	stuff r1	stuff r2	covered r1	covered r2	comes r1	comes r2	Average /ʌ/
E1							
F1	522	534	485	462	568	562	522
F2	1125	1158	1044	1074	1116	1133	1108
F3	2501	2464	2131	2201	2458	2432	2365
F4	3484	3378	3202	3275	3090	3062	3249
E2							
F1	858	874	692	721	795	864	767
F2	1549	1541	1484	1565	1533	1516	1531
F3	2571	2463	2428	2399	2867	3011	2623
F4	4236	4058	4179	4088	4160	4176	4150

English /ʊ/	book r1	book r2	good r1	good r2	took r1	took r2	Average /ʊ/
E1							
F1	466	466	451	412	444	438	446
F2	1093	1017	1556	1414	1250	1257	1265
F3	2290	2328	2376	2394	2367	2191	2324
F4	3444	3437	3319	3299	3290	3271	3343
E2							
F1	533	535	590	528	560	559	551
F2	1115	1141	1760	1481	1586	1380	1411
F3	2457	2382	2558	2369	2330	2241	2390
F4	3839	3880	4224	4048	3907	3764	3944

English /ɔ/	lots r1	lots r2	pops r1	pops r2	fall r1	fall r2	Average /ɔ/
E1							
F1	606	585	627	652	532	549	592
F2	991	1032	969	994	751	767	917
F3	2452	2419	2295	2472	2614	2641	2482
F4	3436	3362	3496	3418	3527	3311	3425
E2							
F1	906	787	913	837	722	772	823
F2	1408	1265	1322	1266	1077	1065	1234
F3	2913	2910	2820	2842	2838	2902	2871
F4	4081	3764	2943	3156	3612	4098	3609

This Page Left Intentionally Blank

6

CONCLUSIONS

> “For Whorf, language is a guide to behavior because its logic is transferred, through analogy, to other domains of human cognitive activity” Duranti (2000: 221).

In order to sort through the significance of the results in Chapter 5, this chapter returns to several predictions made in Chapter 2 and suggests several avenues of future research.

6.1 RETURN TO PREDICTIONS

One of the predictions in Chapter 2 took up the question of whether or not the direction and nature of children’s language learning is solely an interaction of neural and behavioral mechanisms. If that were the case, there would be no effects found in this study. However, it is clear that immature hearing as caused by the temporal bones’ development does affect children’s experience of vowels. The temporal bones do appear to be a non-neural mechanism involved in pre-processing the speech signal.

This study has shown that vowels are simplified during early language learning and that this simplification changes in form over time. This result makes it seem more probable that the vowel systems of languages can be acquired with procedures for general-purpose learning rather than with the genetically controlled neural language acquisition device posited by generative linguists. It appears that the developing temporal bones, not the brain alone, create a bias toward particular types of information carried in vowels. As mentioned earlier, effects on speech sounds caused by the temporal bones remain interesting because they would have evolved in connection with behavioral and biological adaptations (including language) to certain strategies thought to be important in human evolution. Additionally this result supports the work of such linguists as Swingley (2005) proposes that statistical learning is operational early in infant language learning, but that children later rely on mechanisms such as tracking allophonic variation, phonotactics, and stress patterns.

Also mentioned in Chapter 2 was the speculation by Lindblom and colleagues (1992: 365) that the acoustic variability they found in various speech styles “might be considerably reduced if the measurements were translated into more perceptually relevant dimensions.” In other words, they suggested that, while an acoustic analysis of the data might reduce nonsystematic (random) variability, infant-directed speech cannot be the only mechanism responsible for providing the best instances around which phonetic categories are presumably built in native language magnet theory. This dissertation’s results suggest that the development of the temporal bones may be another such mechanism.

6.2 FUTURE RESEARCH

If there ever was an area that needed more research, the preprocessing that the temporal bones contribute to the entire speech signal would be it. The most immediate direction for future research would be to include the variability that the words collected during the speakers’ narration of the eight story pictures bring to the data and see if the results in Chapter 5 still hold.

The foramen of Huschke persists in some small percentage of adults, occurring bilaterally and unilaterally. Moreover, research from several quarters indicates that the nature of the stimulation conducted to the auditory nerve during development is important to the healthy maturation of the auditory structures of the brain. Experimenting on mice, Webster & Webster (1977, 1979) show that auditory neurons need acoustic stimulation to mature, anatomically and functionally; further, the effects on auditory neurons caused by temporary deficits in acoustic stimulation within a critical period do not appear to be reversible. Several studies of humans reveal a significant relationship between conductive problems caused by recurrent otitis media (middle ear infection) during childhood and impairments in linguistic development, including impairments in hearing speech in quiet and noise, auditory discrimination, phonemic synthesis, dichotic listening, and reading comprehension (e.g., Eimas & Kavanagh 1974; Lewis 1976; Luotonen, Uhari, Aitola, Lukkaroinen, Luotonen, Uhari & Korkeamäki 1996; see Bluestone & Klein 2001 for an excellent review). These studies show the effects of abnormal experience on hearing and language skills. Additionally, there appears to be growing speculation that a foramen of Huschke persisting into adulthood might not be as benign as previously thought (Cecire, Austin, & Ng 1991; Hawke et al. 1988; Heffez, Anderson, & Mafee 1989; Rabinov et al. 1997; Wang et al. 1991). Given the advances in recent years in noninvasive imaging (cone beam CTs as described in, e.g., Jaju 2012 and Tozoglu, Caglayan, & Harorli 2012) and contrast techniques (as described in Moreno et al. 2005) for detecting the foramen of Huschke in living individuals it is becoming possible to discover if there are language development problems in people living with a persisting foramen of Huschke.

Yet another area for more research would be to test whether adults would benefit from experiencing vowels in a new language with the ears of the young. Daphne Maurer (2005: 1) introduces a special issue on “critical periods” (renamed as “sensitive periods” among the contributors to this special issue) with this passage:

> Several decades ago, evidence for critical periods emerged in animal research on imprinting and on visual development. In both cases, input early in life altered the behavioral responses of the animal to later input (e.g., to conspecifics, to fine detail) and induced observable changes in the nervous system. These effects occurred during a critical period shortly after birth, such that the comparable manipulation later in life had no effect. Based on these findings, some scientists drew the inference that the nervous system loses plasticity after infancy; however, recent studies of humans with a variety of sensory deficits indicate considerable residual plasticity even in adulthood (e.g., Levi, Polat, & Hu, 1997; Ramachandran & Rogers-Ramachandran, 2000; Taub, Uswatte, & Elbert, 2002).

The plasticity in adult brains described above may mean that it would be helpful to adult language learners to experience vowels and other speech sounds as children do for a time as an aid to mastering sound systems, perhaps through a filtering device.

One potentially significant aspect of development that has been set aside in this dissertation’s auditory modeling is that the right and left temporal bones usually ossify asymmetrically (Ossenberg 1981; Torgersen 1951). This asymmetrical development may mean that the right and left hemispheres of the child’s developing brain receive two versions of the same sounds that differ from one another in ways not presently recognized by auditory researchers. Humans exploit differences in timing and intensity at each of the ears to localize sources of sound (e.g., Pickles 1988); even newborns within hours of birth will turn their heads toward sources of sound (Butterworth & Castillo 1976; Wertheimer 1961). Humans also exploit differences in input arriving from each of the two eyes in depth perception (e.g., Simons 1993), and barn owls exploit the asymmetry of the height of the ears on either sides of their skulls to locate prey (Payne 1968). Is it possible that children are exploiting differences in auditory spectra produced by each of the ears to categorize or triangulate on a location for a speech sound in a map of the phonetic space containing the sounds of their native languages? Native language magnet theory, H&H theory, and auditory enhancement theories (which assume people do form mental representations) would predict that if training effects exist that they would have a role in shaping the prototypes, maps of sound constellations, or distinctive regions posited in their respective theories. This research addresses those theories as framed in biological process: the transformation of earlier simple to later complex states must account for constraint as a factor. Language is biology, thus never completely unconstrained: children cannot perceive what they cannot yet hear. The array of rates of change, directions of change, and differential organizational schema that forms biological constraint lessens as adult state is reached; the initial noise that somehow means something becomes the transformed pattern it is constrained to—these transformed vowels need to be examined as a set in the context of the formation of intermediate maps of phonetic space.

This Page Left Intentionally Blank

7

REFERENCES

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans Automatic Control, 19,* 716–723.

Ambridge, B. & Lieven, E. V. M. (2011). *Child Language Acquisition: Contrasting Theoretical Approaches*. Cambridge: Cambridge University Press,

Anson, B. J., Davies, J. & Duckert, L. G. (1991). Developmental anatomy of the ear. In Paparella, M. M. & Shumrick, D. A. (eds), *Otolaryngology, Volume I Basic Sciences and Related Disciplines,* 3 ed, pp. 3–21. Philadelphia: W. B. Saunders.

Anson, B. J. & Donaldson, J. A. (1981). *Surgical Anatomy of the Temporal Bone.* Philadelphia: W. B. Saunders.

Armstrong, D. F., Stokoe, W. C. & Wilcox, S. E. (1995). *Gesture and the Nature of Language*. Cambridge, England: Cambridge University Press.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L. & Perey, A. J. (1981). Discrimination of voice onset time by human infants: new findings and implications for the effects of early experience. *Child Development*, *52*, 1135–1145.

Aslin, R. N., Pisoni, D. B. & Jusczyk, P.W. (1983). Auditory development and speech perception in infancy. In Haith, M. M. & Campos, J. J. (eds), *Infancy and the Biology of Development,* pp. 573–687. New York: Wiley.

Assmann, P., Nearey, T. & Hogan, J. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America, 71,* 975–989.

Baker, B. J., Dupras, T. L. & Tocheri, M. W. (2005). *The Osteology of Infants and Children*. College Station: Texas A&M University Press.

Ballachanda, B. (1995). *The Human Ear Canal.* San Diego, CA: Singular Publishing Group, Inc.

Barrett, L. (2011). *Beyond the Brain: How Body and Environment Shape Animal and Human Minds.* Princeton, NJ: Princeton University Press.

Bass, W. M. (1995). *Human Osteology: A Laboratory and Field Manual.* 4 ed. Columbia, MO: Missouri Archaeological Society.

Bastir, M. & Rosas, A. (2004). Comparative ontogeny in humans and chimpanzees: Similarities, differences and paradoxes in postnatal growth and development of the skull. *Annals of Anatomy, 186,* 503–509.

Bateson, G. (1972). *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology*. Chicago: University of Chicago Press.

Behne, T., Liszkowski, U., Carpenter, M. & Tomasello, M. 2012. Twelve-month-olds' comprehension and production of pointing. *British Journal of Developmental Psychology, 30(3),* 359–375.

Benguerel, A.-P. & McFadden, T. (1989). The effect of coarticulation on the role of transitions in vowel perception. *Phonetica, 46*, 80–96.

Bentler, R. A. (1991). The resonance frequency of the external auditory canal in children: Letters to the editor. *Ear and Hearing, 12(2),* 89–90.

Bentley, D. R. & Hoy, R. R. (1996). Genetic control of the neuronal network generating cricket (*Teleogryllus gryllus*) song patterns. In Houck, L. D. & Drickamer, L. E. (eds), *Foundations of Animal Behavior: Classic Papers with Commentaries,* pp. 343–357. Chicago: University of Chicago Press.

Best, C. T. & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech, 46(2),* 183–216.

Best, C. T., McRoberts, G. W., LaFleur, R. & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant Behavior and Development, 18,* 339–350.

Birch, H. G. (1962). Dyslexia and the maturation of visual function. In Money, J. (ed), *Reading Disability: Progress and Research Needs in Dyslexia*, pp. 161–169. Baltimore: Johns Hopkins Press.

Bladon, A. (1983). Two-formant models of vowel perception: Shortcomings and enhancements. *Speech Communication, 2,* 305–313.

Bladon, A. & Lindblom, B. (1981). Modeling the judgement of vowel quality differences. *Journal of the Acoustical Society of America, 69,* 1414–1422.

Blount, B. (1975). Studies in child language: An anthropological review. *American Anthropologist, 77,* 580–599.

Bluestone, C. D. & Klein, J. O. (2001). *Otitis media in infants and children.* 3 ed. Philadelphia: Harcourt Brace.

Boas, F. (1911 [1966]). *Introduction to Handbook of American Indian Languages.* Lincoln: University of Nebraska Press.

Bogin, B. (1999). *Patterns of Human Growth.* 2ed. Cambridge, England: Cambridge University Press.

de Boysson-Bardies, B. (1999). *How Language Comes to Children: From Birth to Two Years.* Cambridge, MA: The MIT Press.

Bramblett, C. (2004). Evolution of primates. Retrieved from http://uts.cc.utexas.edu/~bramblet/ant301/thirteen.html.

Bredberg, G. (1968). Cellular pattern and nerve supply of the human organ of Corti. *Acta Otolaryngologica, 236 (supplement),* 1–135.

Brent, M. R. & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61,* 93–125.

Bruner, J. (1983). *Child's Talk: Learning to Use Language.* New York: W. W. Norton.

Bundgaard-Nielsen, R. L., Best, C. T., Kroos, C. & Tyler, M. D. (2012). Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility. *Applied Psycholinguistics, 33(3),* pp. 643–664.

Butterworth, G. & Castillo, M. (1976). Coordination of auditory and visual space in newborn human infants. *Perception, 5,* 155–160.

Byrne, R. (1995). *The Thinking Ape: Evolutionary Origins of Intelligence.* Oxford, England: Oxford University Press.

Caramazza, A., Chialant, D., Capasso, R., & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature, 403*, 159–160.

Cecire, A. A., Austin, B. W., & Ng, P. K. (1991). Polyp of the external ear canal arising from the temporomandibular joint: A case report. *Journal of Otolaryngology 20,* 168–170.

Charles-Luce, J. & Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language, 17,* 205–215.

Chiel, H. J. & Beer, R. D. (1997). The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in Neuroscience, 20(12),* 553–557.

Chistovich, L. A. (1985). Central auditory processing of peripheral vowel spectra. *Journal of the Acoustical Society of America, 77,* 789–805.

Chomsky, N. (2000). *New Horizons in the Study of Language.* Cambridge, England: Cambridge University Press.

Chomsky, N. (1988). *Language and Problems of Knowledge: The Managua Lectures*. Cambridge, MA: The MIT Press.

Chomsky, N. (1965). *Aspects toward a Theory of Syntax.* Cambridge, MA: The MIT Press.

Chomsky, N. (1959). A review of B. F. Skinner's *Verbal Behavior*. *Language, 35,* 26–58.

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action and Cognitive Extension*. Oxford, England: Oxford University Press.

Clark, E. (2001). Emergent categories in first language acquisition. In Bowerman, M. & Levinson, S. C. (eds), *Language Acquisition and Conceptual Development,* pp. 379–405. Cambridge, England: Cambridge University Press.

Clark, E. (1979). Building a vocabulary: Words for objects, actions and relations. In Fletcher, P. & Garman, M. (eds), *Language Acquisition: Studies in First Language Development,* pp. 149–160. Cambridge, England: Cambridge University Press.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America, 24,* 597–606.

Dahm, M. C., Shepherd, R. K. & Clark, G. M. (1993). The postnatal growth of the temporal bone and its implications for cochlear implantation in children. *Acta Laryngologia, 505 (supplement),* 5–40.

Davis, B. L. & Lindblom, B. (2001). Phonetic variability in baby talk and development of vowel categories. In Lacerda, F., von Hofsten, C. & Heimann, M. (eds), *Emerging*

Cognitive Abilities in Early Infancy, pp. 135–171. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Deacon, T. W. (1997). *The Symbolic Species: The Co-Evolution of Language and the Brain.* New York: W. W. Norton.

Deane, P. D. (1991). Syntax and the brain: Neurological evidence for the spatialization of form hypothesis. *Cognitive Linguistics, 2,* 361–367.

de Beer, G.R. (1958). *Embryos and Ancestors,* 3 ed. London: Oxford University Press.

de Boer, B. (2000). Self-organization in vowel systems. *Journal of Phonetics, 28,* 441–465.

DeCasper, A. J. & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science, 208,* 1174–1176.

DeCasper, A. J. & Spence, M. J. (1986). Prenatal maternal speech influences newborn's perception of speech sounds. *Infant Behavior and Development, 9,* 133–150.

Degusta, D., Gilbert W. H. & Turner, S. P. (1999). Hypoglossal canal size and hominid speech. *Proceedings of the National Academy of the Sciences, 96(4),* 1800–1804.

Dempster, J. H. & Mackenzie, K. (1990). The resonance frequency of the external auditory canal in children. *Ear and Hearing, 11(4),* 296–298.

Dempster, J. H. & Mackenzie, K. (1991) Reply to Bentler: Letters to the editor. *Ear and Hearing, 12(2),* 90.

Denes, P. B. & Pinson, E. N. (1993). *The Speech Chain: The Physics and Biology of Spoken Language.* 2 ed. New York: W. H. Freeman and Company.

Diehl, R. L. (2000). Searching for an auditory description of vowel categories. *Phonetica, 57*, 267–274.

Diehl, R. L., Kluender, K. R. & Walsh, M. A. (1990). Some auditory bases of speech perception and production. *Advances in Speech, Hearing & Language Processing, 1,* 243–267.

Dillon, G. L. (2003, April 7). *The vowel quadrilaterals for stereotypical national dialects* [WWW Document]. URL http://faculty.washington.edu/dillon/PhonResources/vowels.html (visited 2003, May 27).

Duranti, A. (2000). Relativity. *Journal of Linguistic Anthropology 9(l-2),* 220–222.

Durrant, J. D. & Lovrinic, J. H. (1977). *Bases of Hearing Science.* Baltimore: Williams and Wilkins.

Eimas, P. D. (1975). Speech perception in early infancy. In Cohen, L. B. & Salapatek, P. (eds), *Infant Perception: From Sensation to Cognition, Volume 2 Perception of Space, Speech, and Sound,* pp. 193–231. New York: Academic Press.

Eimas, P. D. & Kavanagh, J. F. (1974). Otitis media, hearing loss, and child development. *Public Health Reports, 101,* 289–293.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. & Vigorito, J. (1971). Speech perception in infants. *Science* [New Series], *171(3968),* 303–306.

Elliot, L.L. & Hammer, M.A. (1993). Fine-grained auditory discrimination: Factor structures. *Journal of Speech & Hearing Research, 36(2),* 396–409.

Elman, J., Bates, E., Karmiloff-Smith, A. Parisi, D., Johnson, M., & Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development.* Cambridge, MA: The MIT Press.

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., Monaco, A. P., and Pääbo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature, 418,* 869–872.

Everitt, B. S. & Dunn, G. (2001). *Applied Multivariate Data Analysis*. 2 ed. London, England: Arnold.

Falk, D. & Clarke, R. (2007). New reconstruction of the Taung endocast [Brief Communication]. *American Journal of Physical Anthropology, 134(4),* 529–534.

Falk, D., Zollikofer, C.P.E., Morimoto, M. & Ponce de León, M.S. (2012). Metopic suture of Taung (*Australopithecus africanus*) and its implications for hominin brain evolution. *Proceedings of the National Academy of the Sciences, 109(22),* 8467–8470.

Fant, G. (1956). On the predictability of formant levels and spectrum envelopes from formant frequencies. In Halle, M., Lunt, H., and MacLean, H. (eds), *For Roman Jakobson,* pp. 109–120. The Hague: Mouton.

Farina, M. (2012). Louise Barrett, beyond the brain: how body and environment shape animal and human minds [Book Review]. *Phenomenology and the Cognitive Sciences, 11,* 415–421.

Fauconnier, G. (2002). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities.* New York: Basic Books.

Fauconnier, G. (1997). *Mappings in Thought and Language.* Cambridge, England: Cambridge University Press.

Fauconnier, G. (1985). *Mental Spaces: Aspects of Meaning Construction in Natural Language.* Cambridge: MIT Press.

Fellman, V. & Huotilainen, M. (2006). Cortical auditory event-related potentials in newborn infants. *Seminars in Fetal and Neonatal Medicine, 11(6),* 452–458.

Fernald, A. (2000). Speech to infants as hyperspeech: Knowledge-driven processes in early word recognition. *Phonetica, 57,* 242–254.

Fisher, C. & Tokura, H. (1995). The given-new contract in speech to infants. *Journal of Memory and Language, 34,* 287–310.

Flemming, E. (1996). Evidence for constraints on contrast: The dispersion theory of contrast. In Hsu, C. S. (ed), *UCLA Working Papers on Phonology 1,* pp. 1–23. Los Angeles: UCLA.

Fletcher, H. & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *Journal of the Acoustical Society of America, 5,* 82–108.

Fogerty, D. & Humes, L. E. (2012). The role of vowel and consonant fundamental frequency, envelope, and temporal fine structure cues to the intelligibility of words and sentences. *Journal of the Acoustical Society of America, 131(2),* 1490–1501.

Foley, W. A. (1997). *Anthropological Linguistics.* Malden, MA: Blackwell Publishers.

Forrest, K., Weismer, G., Milenkovic, P. & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America, 84*, 115–123.

Foushée, R. D. & Lickliter, R. (2002). Early visual experience affects postnatal auditory responsiveness in bobwhite quail (*Colinus virginianus*). *Journal of Comparative Psychology, 116,* 369–380.

Fowler, C. (1994). Speech perception: Direct realist theory. In Asher, R. E. (ed), *The Encyclopedia of Language and Linguistics*, pp. 4199–4203. Oxford, England: Pergamon Press.

Fowler, C. (1986). Event approach to the study of speech perception. *Journal of Phonetics, 14,* 3–28.

Friend, C. (2000). *Silly Ruby* [illustrated by Merriman, R.]. Cambridge, MA: Candlewick Press.

Fry, D. B. (1979). *The Physics of Speech.* Cambridge, England: Cambridge University Press.

Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C. R., Conboy, B. T., Romo, H., Klarman, L., Ortiz, S. & Kuhl, P. K. (2011). Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics, 39(4),* 546–557.

Geertz, C. (1973). *The Interpretation of Cultures*. New York: BasicBooks.

Gerrits, E. & deBree, E. (2009). Early language development of children at familial risk of dyslexia: Speech perception and production. *Journal of Communication Disorder*, *42(3*), 180–194.

Ghitza, O. (1993). Adequacy of auditory models to predict human internal representation of speech sounds. *Journal of the Acoustical Society of America, 93,* 2160–2171.

Gibson, J. J. (1971 [1982]). A preliminary description and classification of Affordances. In Reed, E. & Jones, R. (eds), *Reasons for Realism: Selected Essays of James J. Gibson,* pp. 403–406. Hillsdale, NJ: Lawrence Erlbaum.

Gibson, J. J. (1950). *The Perception of the Visual World.* Boston: Houghton Mifflin.

Gibson, K. R. (1996). The biocultural human brain, seasonal migrations, and the emergence of the Upper Palaeolithic. In Mellars, P. and Gibson, K. R. (eds), *Modelling the Early Human Mind,* pp. 33–46. Cambridge, England: The McDonald Institute for Archaeological Research.

Gibson, K. R. (1991). Myelination and behavioral development: A comparative perspective on questions of neoteny, altriciality and intelligence. In Gibson, K. R. & Petersen, A. C. (eds), *Brain Maturation and Cognitive Development: Comparative and Cross-Cultural Perspectives*, pp. 29–63. New York: Aldine de Gruyter.

Goffman, E. (1974). *Frame Analysis: An Essay on the Organization of Experience.* New York: Harper & Row, Publishers.

Golestani, N & Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain and Language, 109(2–3),* pp. 55–67.

Gordon, I. & Slater, A. (1998). Nativism and empiricism: The history of two ideas. In Slater, A. (ed), *Perceptual Development: Visual, Auditory, and Speech Perception in Infancy*, pp. 73–103. East Sussex, England: Psychology Press Ltd., Publishers.

Gottlieb, G. (1971). Ontogenesis of sensory function in birds and mammals. In Tobac, E., Aronson, L. & Shaw, E. (eds), *The Biopsychology of Development*, pp. 67–128. New York: Academic Press.

Gottlieb, G., Tomlinson, W. T. & Radell, P. L. (1989). Developmental intersensory interference: Premature visual experience suppresses auditory learning in ducklings. *Infant Behavior and Development*, *12,* 1–12.

Gould, E. (1975). Experimental studies of the ontogeny of ultrasonic vocalization in bats. *Developmental Psychobiology, 8,* 333–346.

Gould, E. Cooley, B. & Barnick, P. (1981). Echolocation by young bats on their initial and subsequent flights. *Developmental Psychobiology, 14,* 41–54.

Gould, S. J. (1977). *Ever Since Darwin.* New York: W. W. Norton.

Grieser, D. & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology, 24,* 577–588.

Guild, S. R. (1936). Hearing by bone conduction: The pathways of transmission of sound. *Annals of Otology, Rhinology, and Laryngology, 45,* 736–738.

Gumperz, J. J. (1972). Introduction. In Gumperz, J. J. and Hymes, D., Directions in Sociolinguistics: The Ethnography of Communication, pp. 1–25. New York: Holt, Rinehart and Winston.

Hainline, L. (1998). The development of basic visual abilities. In Slater, A. (ed), *Perceptual Development: Visual, Auditory, and Speech Perception in Infancy,* pp. 5–50. East Sussex, England: Psychology Press Ltd., Publishers.

Hainline, L. & Abramov, I. (1992). Assessing visual development: Is infant vision good enough? In Rovee-Collier, C. & Lipsitt, L. P. (eds), *Advances in Infancy Research, Volume 7,* pp. 39–102. Norwood, NJ: Ablex.

Halina, M., Rossano, F. & Tomasello, M. (2013). The ontogenetic ritualization of bonobo gestures. *Animal Cognition,* online publ., 31 January 2013, 14pp.

Hammer, Ø., Harper, D.A.T. & Ryan, P. D. 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontologia Electronica 4(1), online publ., 9pp.

Harrison, M. J. (2001). Theory of auditory thresholds in primates: Why monkeys can't hear human speech as well as humans. Paper presented at the 2001 American Physical Society March Meeting, Seattle. [abstract and lay language version available http://www.aps.org]

Harrison, M. J. (1998). The roles of statistical fluctuation in particle density and thermal excitation of resonant ear canals in establishing auditory thresholds in primates, including *H. Sapiens*. Paper presented at the 20th International Conference on Statistical Physics, International Union of Pure and Applied Physics, Paris.

Hashimoto, T., Ojiri, H., & Kawai, Y. (2011). The foramen of Huschke: Age and gender specific features after childhood. *International Journal of Oral Maxillofacial Surgery, 40,* 743–746.

Hawke, M., Kwok, P., Shanker, L., & Wang, R. G. (1988). Spontaneous temporomandibular joint fistula into the external auditory canal. *Journal of Otolaryngology, 17,* 29–31.

Hawkins, S. (1999a). Auditory capacities and phonological development: Animal, baby, and foreign listeners. In Pickett, J. M., *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*, pp. 183–197. Boston: Allyn and Bacon.

Hawkins, S. (1999b). Looking for invariant correlates of linguistic units: Two classical theories of speech perception. In Pickett, J. M., *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*, pp. 198–231. Boston: Allyn and Bacon.

Hawkins, S. (1999c). Reevaluating assumptions about speech perception: Interactive and integrative theories. In Pickett, J. M., *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*, pp. 232–288. Boston: Allyn and Bacon.

Heath, S. B. (1983). *Ways with Words: Language, Life and Work in Communities and Classrooms.* Cambridge, England: Cambridge University Press.

Heffez, L., Anderson, D., & Mafee, M. (1989). Developmental defects of the tympanic plate: case reports and review of the literature. *Journal of Oral Maxillofacial Surgery, 47,* 1336–1340.

Herzog, S. & Fiese, R. (1989). Persistent foramen of Huschke: Possible risk factor for otologic complications after arthroscopy of the temporomandibular joint. *Oral Surgery, Oral Medicine and Oral Pathology, 68,* 267–270.

Hillenbrand, J., Getty, L. A., Clark, M. J. & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97,* 3099–3111.

Holdsworth, J., Schwartz, J. L., Berthommier, F. & Patterson, R. D. (1992). A multi-representation model for auditory processing of sounds. In Cazals, Y., Horner, K.

& Demany, L. (eds), *Auditory Physiology and Perception*, pp. 447–453. Oxford, England: Pergamon Press.

Hollinshead, W. H. (1962). *Textbook of Anatomy*. New York: Harper & Row.

Hollinshead, W. H. (1954). *The Head and Neck: Anatomy for Surgeons, Volume 1.* New York: Harper Brothers.

Hoyte, D. A. N. (1997). Growth of the ear capsule. In Dixon, A. D., Hoyte, D. A. N. & Rönning, O. (eds), *Fundamentals of Craniofacial Growth,* pp. 335–362. Boca Raton, FL: CRC Press.

Humphrey, L.T. & Scheurer, L. (2006). Age of closure of the foramen of Huschke: An osteological study. *International Journal of Osteoarchaeology, 16,* 47–60.

Hurford, J. R., Studdert-Kennedy, M. & Knight, C., eds (1998). *Approaches to the Evolution of Language.* Cambridge, England: Cambridge University Press.

Hymes, D. H. (1974). *Foundations in Sociolinguistics: An Ethnographic Approach.* Philadelphia: University of Pennsylvania Press.

Iverson, P. & Kuhl, P. K. (2000). Perceptual magnets and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Perception and Psychophysics, 62,* 874–886.

Janky, K. L., Zuniga, M. G., Carey, J. P., & Schubert, M. (2012). Balance dysfunction and recovery after surgery for superior canal dehiscence syndrome. *Archives of Otolaryngology--Head & Neck Surgery, 138(8),* 723–730.

Jakobsen, R., Fant, G. & Halle, M. (1963). *Preliminaries to Speech Analysis.* Cambridge, MA: The MIT Press.

Johnson, K. (1997). *Acoustic and Auditory Phonetics.* Cambridge, MA: Blackwell Publishers.

Johnson, K. (1992). Acoustic and auditory analysis of Xhosa clicks and pulmonics. *UCLA Working Papers in Phonetics, 83,* 33–45.

Johnson, K. (1989). Vowel contrast and normalization. *Journal of Phonetics, 18,* 229–254.

Johnson, K., Flemming, E. & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language, 69,* 505–528.

Johnson, M. & Lakoff, G. (2002). Why cognitive linguistics requires embodied realism. *Cognitive Linguistics, 13,* 246–263.

Jusczyk, P. W., Cutler, A. & Redanz, N. J. (1993). Infants’ preference for the predominant stress patterns of English words. *Child Development, 64,* 675–687.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y. & Jusczyk, A. M. (1993). Infants’ sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32,* 402–420.

Jusczyk, P. W., Houston, D. & Goodman, M. (1998). Speech perception during the first year. In Slater, A. (ed), *Perceptual Development: Visual, Auditory, and Speech*

Perception in Infancy, pp. 357–388. East Sussex, England: Psychology Press Ltd., Publishers.

Kay, R. F., Cartmill, M. & Balow, M. (1998). The hypoglossal canal and the origin of human vocal behavior. *Proceedings of the National Academy of Sciences, 95,* 5417–5419.

Keefe, D. H., Bullen, J. C., Campbell, S. L. & Burns, E. M. (1994). Pressure transfer function and absorption cross section from the diffuse field to the human infant ear canal. *Journal of the Acoustical Society of American, 95,* 355–371.

Kenney, P. A. & Turkewitz, G. (1986). Effects of unusually early visual stimulation on the development of homing behavior in the rat pup. *Developmental Psychobiology, 19,* 57–66.

Kewley-Port, D. & Atal, B. (1989). Perceptual differences between vowels located in a limited phonetic space. *Journal of the Acoustical Society of America, 85,* 1726–1740.

Kewley-Port, D., Burkle, T. Z., & Lee, J. H. (2007). Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *Journal of the Acoustical Society of America, 122(4),* 2365–2375.

King, B. J. (1994). *The Information Continuum: Evolution of Social Information Transfer in Monkeys, Apes, and Hominids.* Santa Fe, NM: School of American Research Press.

Klatt, D. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, *7,* 279–312.

Kluender, K. R. (1998). A special scientist's vertical view. *Contemporary Psychology, 43,* 26–27.

Kluender, K. R., Lotto, A. J., Holt, L. L., & Bloedel, S. L. (1998). Role of experience for language-specific functional mappings of vowel sounds. *Journal of the Acoustical Society of America, 104,* 3568–3582.

Konner, M. (1991). Universals of behavioral development in relation to brain myelination. In Gibson, K. R. & Petersen, A. C. (eds), *Brain Maturation and Cognitive Development: Comparative and Cross-Cultural Perspectives*, pp. 181–223. New York: Aldine de Gruyter.

Kuhl, P. K. (2001). Speech, language, and developmental change. In Lacerda, F., von Hofsten, C. & Heimann, M. (eds), *Emerging Cognitive Abilities in Early Infancy*, pp. 111–131. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences, 97,* 11850–11857.

Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology, 4,* 812–822.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics, 50,* 93–107.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U. & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277,* 684–686.

Kuhl, P. K. & Miller, J. D. (1978). Speech perception by the chinchilla. Identification functions for synthetic VOT stimuli. *Journals of the Acoustical Society of America, 63,* 905-917.

Kuhl, P. K. & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science, 190,* 69–72.

Kuhl, P. K. & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America, 73,* 1003–1010.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N. & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255,* 606–608.

Künzel, H.J., Masthoff, H.R. & Köster, J.P. (1995). The relation between speech tempo, loudness, and fundamental frequency: an important issue in forensic speaker recognition. *Science and Justice*, 35(4): 291–295.

Lacerda, F. (1994). The asymmetric structure of the infant’s perceptual vowel space. *Journal of the Acoustical Society of America,* 95(5): 3016.

Lacout, A., Marsot-Dupuch, K., Smoker. W. R., & Lasjaunias, P. (2005). Foramen tympanicum, or foramen of Huschke: pathologic cases and anatomic CT study. *American Journal of Neuroradiology, 26(6),* 1317–1323.

Ladefoged, P. (1956). The classification of vowels. *Lingua, 5,* 113–127.

Lakoff, G. (1990). The invariance hypothesis: Is abstract reason based on image-schemas? *Cognitive Linguistics, 1,* 39–74.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind.* Chicago: University of Chicago Press.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought.* New York: Basic Books.

Lakoff, G. & Johnson, M. (1980). *Metaphors We Live By.* Chicago: The University of Chicago Press.

Langacker, R. (2001). Dynamicity in grammar. *Axiomathes, 12,* 7–33.

Langacker, R. (2000). *Grammar and Conceptualization.* New York: Mouton de Gruyter.

Langacker, R. (1991). *Foundations of Cognitive Grammar, Volume 2*. Stanford, CA: Stanford University Press.

Langacker, R. (1987). *Foundations of Cognitive Grammar, Volume 1*. Stanford, CA: Stanford University Press.

Laver, J. (1994). *Principles of Phonetics.* Cambridge, England: Cambridge University Press.

Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge, England: Cambridge University Press.

Lecanuet, J.-P. (1998). Fœtal responses to auditory and speech stimuli. In Slater, A. (ed), *Perceptual Development: Visual, Auditory, and Speech Perception in Infancy*, pp. 317–355. East Sussex, England: Psychology Press Ltd.

Lecanuet, J.-P., Granier-Deferre, C., DeCasper, A. J., Maugeais, R., Andrieu A. J., Busnel, M.-C. (1987). Perception et discrimination foetales de stimuli langagiers; mise en évidence à partir de la réactivité cardiaque; résultats préliminaires. *Comptes Rendues de l'Academie des Sciences. Serie III, Sciences de la Vie, 305*, 161–164.

Lee, K. J. (2003). *Essential Otolaryngology: Head and Neck Surgery.* New York: McGraw Hill.

Leigh, S. R. (2012). Brain size growth and life history in human evolution [Synthesis Paper]. *Evolutionary Biology, 39,* 587–599.

Leigh, S. R. (2004). Brain growth, life history, and cognition in primate and human evolution. *American Journal of Primatology, 62*, 139–164.

Leutenegger, W. (1982). Encephalization and obstetrics in primates with particular reference to human evolution. In Armstrong, E. and Falk, D (eds), *Primate Brain Evolution: Methods and Concepts*, pp. 85–95. New York: Plenum Press.

Lévi-Strauss, C. (1963). *Structural Anthropology.* New York: Basic Books.

Levi, D., Polat, U., & Hu, Y.-S. (1997). Improvement in vernier acuity in adults with amblyopia. *Investigative Ophthalmology & Visual Science, 38,* 1493–1510.

Lewis, N. (1976). Otitis media and linguistic incompetence. *Archives of Otolaryngology, 102*, 387–390.

Liberman, A. (1996). *Speech: A Special Code.* Cambridge, MA: The MIT Press.

Liberman, A. & Mattingly, I. (1985). The motor theory of speech perception revised. *Cognition*, *21,* 1–36.

Liberman, A. M., Cooper, F. S., Shankweiler, D. B., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431–461.

Liebal, K., Carpenter, M. & Tomasello, M. (2013). Young children's understanding of cultural common ground. *British Journal of Developmental Psychology, 31(1),* 88–96.

Lieberman, P., Laitman, J. T., Reidenberg, J. S. & Gannon, P. J. (1992). *Journal of Human Evolution, 23,* 447–467.

Lickliter, R. (1996). Structured organisms and structured environments: Developmental systems and the construction of learning capacities. In Valsiner, J. & Voss, H. G. (eds), *The Structure of Learning Processes,* pp. 86–107. Norwood, NJ: Ablex Publishing.

Lightfoot, D. (1991). *How to Set Parameters.* Cambridge, MA: The MIT Press.

Lindblom, B. (2000). Developmental origins of adult phonology: The interplay between phonetic emergents and the evolutionary adaptations of sound patterns. *Phonetica, 57,* 297–314.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In Hardcastle, W. J. & Marchal, A. (eds), *Speech Production and Speech Modelling,* pp. 403–439. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Lindblom, B., Brownlee, S., Davis, B. & Moon, S.-J. (1992). Speech transforms. *Speech Communication, 11,* 357–368.

Lindblom, B., Perkell, J. S. & Klatt, D. H. (1986). Preface. In Perkell, J. S. & Klatt, D. H., (eds), *Invariance and Variability in Speech Processes*, pp. i–iv. Hillsdale, NJ: Lawrence Erlbaum Associates, Publisher.

Lindblom, B. & Engstrand, O. (1989). In what sense is speech quantal? *Journal of Phonetics, 17,* 107–121.

Lindgren, L. (1977). Influence of pressure upon fetal head during labor. *Acta Obstetricia et Gynecologica Scandinavica, 56,* 303–309.

Lum, D. A. & Braida, L. D. (2000). Perception of speech and non-speech sounds by listeners with real and simulated sensorineural hearing loss. *Journal of Phonetics, 28,* 343–366.

Luotonen, M., Uhair, M., Aitola, L., Lukkaroinen, A., Luotonen, J., Uhari, M. & Korkeamäki, R. (1996). Recurrent otitis media during infancy and linguistic skills at the age of nine years. *The Pediatric Infectious Disease Journal, 15,* 854–858.

Lynn, J. M & Garn-Nunn, P. G. (2004). *Calvert's Descriptive Phonetics.* 3 ed. New York: Thieme Medical Publishers.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk, Volume 1 The Database.* Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Maddison, I. (2003). Vowel spacing in four-vowel systems. *The Journal of the Acoustical Society of America, 113(4),* 2331.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge, England: Cambridge University Press.

Mann, V. A. & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition, 14,* 211–235.

Marchman, V.A. & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science, 11(3),* F9–F16.

Massaro, D. W. (1994). Speech perception. In Asher, R. E. (ed), *The Encyclopedia of Language and Linguistics, Volume 8,* pp. 4186–4199. Oxford, England: Pergamon Press.

Masterton, R. B. (1992). Role of the central auditory system in hearing: The new direction. *Trends in Neuroscience, 15,* 280–285.

Maurer, D. (2005). Introduction to the special issue on critical periods reexamined: Evidence from human sensory development. *Wiley Interscience [online publication www.interscience.wiley.com]*, 1.

McClelland, J. L. & Elman, J. L. (1986). Interactive processes in speech perception: The TRACE model. In McClelland, J. L., Rumelhart, D. E., & the PDP Research Group (eds), *Parallel Distributed Processing: Exploration in the Microstructure of Cognition, Volume 2 Psychological and Biological Models,* pp. 58–121. Cambridge, MA: The MIT Press.

McDonough, J. (2003). *The Navajo Sound System.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

McDonough, J. (1994). A report on the vowels of traditional speakers of Western Navajo and the implications for a theory of vowel dispersion. *Journal of the Acoustical Society of America, 95(5),* 2925.

McDonough, J., Ladefoged, P. & George, H. (1992). Navajo vowels and universal phonetic tendencies. *Journal of the Acoustical Society of America*, 92, 2416.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J. & Amiel-Tison, C. (1988). A precursor of language acquisition in infants. *Cognition, 29,* 143–178.

Mehler, J., Sebástian-Gallés, N. & Nespor, M. (2004). Biological foundations of language: Language acquisition, cues for parameter setting and the bilingual infant. In Gazzaniga, M. (ed), *The New Cognitive Neuroscience,* 3 ed, pp. 825–836. Cambridge, MA: MIT Press.

Meisel, J. M. (1995). Parameters in acquisition. In Fletcher, P. & MacWhinney, B. (eds), *The Handbook of Child Language*, pp. 10–35. Cambridge, MA: Blackwell Publishers.

Michel, G.F. & Tyler, A.N. (2005). Critical period: a history of the transition from questions of when, to what, to how. *Developmental Psychobiology, 46*, 156–162.

Miller, J. L. & Eimas, P. D. (1996). Internal structure of voicing categories in early infancy. *Perception & Psychophysics, 58,* 1157–1167.

Minor, L. B., Solomon D., Zinreich, J. S., & Zee, D. S. (1998). Sound- and/or pressure-induced vertigo due to bone dehiscence of the superior semicircular canal. *Archives of Otolaryngology--Head and Neck Surgery, 124(3)*, 249–258.

Molis, M. R. (1999). Perception of vowel quality in the F_2/F_3 plane. Proceedings of the International Congress of the Phonetic Sciences, pp. 191–194, San Francisco.

Moore, B. C. J. & Glasberg, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America, 74*, 750–753.

Moreno, R. C., Chilvarquer, I ., Hayek, J. E., & Seraidarian, P. I. (2005). Anatomic and radiograph study of the persistence of foramen of Huschke. *Brazilian Journal of Otorhinolaryngology, 71(5)*, 676–679.

Moser, H. M., Michel, J. F., & Fotheringham, W. C. (1961). Intelligibility and confusability of vowels and diphthongs under various conditions of quiet and noise. *USAF Command and Control Development Division Technical Note,* No. 61-33, 19 pp.

Munson, B., Edwards, J. & Beckman, M. (2005). Phonological knowledge in typical and atypical speech-sound development. *Topics in Language Disorders*, *25(3),* 190–206.

Myers, J., Jusczyk, P., Nelson, D. G. K., Charles-Luce, J., Woodward, A. L., Hirsh-Pasek, K. (1996). Infants' sensitivity to word boundaries in fluent speech. *Journal of Child Language, 23,* 1–30.

Nearey, T. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America, 85*, 2088–2113.

Nespor, M. (1990). On the rhythm parameter in phonology. In Roca, I. (ed), *Logical Issues in Language Acquisition,* pp. 157–175. Dordrecht, The Netherlands: Foris.

Nittrouer, S. (1994). Speech perception by normal children. *Journal of the Acoustical Society of America, 95(5)*, 2997.

Noble, W. & Davidson, I. (1996). *Human Evolution, Language and Mind: A Psychological and Archaeological Inquiry*. Cambridge, England: Cambridge University Press.

Nozza, R. J., Rossman, R. N. F. & Bond, L. C. (1990). Infant-adult difference in unmasked thresholds for the discrimination of consonant-vowel systems. *Audiology, 30*, 102–111.

Ochs, E. (1988). *Culture and Language Development: Language Acquisition and Language Socialization in a Samoan Village.* Cambridge, England: Cambridge University Press.

Ochs, E. & Schieffelin, B. (1995). The impact of language socialization on grammatical development. In Fletcher, P. & MacWhinney, B. (eds), *The Handbook of Child Language*, pp. 73–94. Cambridge, MA: Blackwell Publishers.

Ochs, E. & Schieffelin, B. (1979). Introduction: What child language can contribute to pragmatics. In Ochs, E. & Schieffelin, B. B. (eds), Developmental Pragmatics, pp. 1–17. New York: Academic Press.

Olsho, L. W. (1985). Infant auditory perception: Tonal masking. *Infant Behavior and Development, 8,* 371–384.

Osheron, D. N., Stob, M. & Weinstein, S. (1986). *Systems That Learn: An Introduction to Learning Theory for Cognitive and Computer Scientists.* Cambridge, MA: The MIT Press.

Ossenberg, N. S. (1981). *Discontinuous Morphological Variation in the Human Cranium.* Unpublished dissertation, University of Toronto.

Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *Journal of the Acoustical Society of America, 59*, 640–654.

Patterson, R. D., Allerhand, M. H. & Giguère, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *Journal of the Acoustical Society of America, 98,* 1890–1894.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L. & Milroy, R. (1982). The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *Journal of the Acoustical Society of America, 72,* 1788–1803.

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C. & Allerhand, M. (1992). Complex sounds and auditory images. In Cazals, Y., Horner, K. & Demany, L. (eds), *Auditory Physiology and Perception*, pp. 429–446. Oxford, England: Pergamon Press.

Payne, R.S. 1968. *How Barn Owls Hunt*. New York: McGraw Hill.

Penin, X., Berge, C., & Baylac, M. (2002). Ontogenetic study of the skull in modern humans and the common chimpanzees: Neotenic hypothesis reconsidered with a tridimensional Procrustes analysis. *American Journal of Physical Anthropology, 118*, 50–62.

Perkins, W. H. & Kent, R. D. (1986). *Functional Anatomy of Speech, Language, and Hearing: A Primer*. Austin, TX: Pro-Ed.

Peterson, G. E., Wang, S-Y. & Sivertsen, E. (1958). Segmentation techniques in speech synthesis. *Journal of the Acoustical Society of America, 30,* 739–742.

Pickett, J. M. (1999). *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*. Boston: Allyn and Bacon.

Pickles, J. O. (1988). *An Introduction to the Physiology of Hearing*. 2 ed. London: Academic Press.

Pisoni, D. B., Lively, S. E. & Logan, J. S. (1994). Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In Goodman, J. C. & Nusbaum, H. C., *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, pp. 121–166. Cambridge, MA: The MIT Press.

Plunkett, K. (1995). Connectionist approaches to language acquisition. In Fletcher, P. & MacWhinney, B. (eds), *The Handbook of Child Language*, pp. 36–72. Cambridge, MA: Blackwell Publishers.

Plunkett, K. (1998). *Language Acquisition and Connectionism.* East Sussex, England: Psychology Press, Ltd.

Polka, L. & Werker, J. F. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 421–435.

Proctor, B. (1989). *Surgical Anatomy of the Ear and Temporal Bone.* New York: Thieme Medical Publishers.

Pujol, R. & Lavigne-Rebillard, M. (1985). Early stages of innervation and sensory cell differentiation in the human fetal organ of Corti. *Acta Otolaryngologica, 423 (supplement),* 43–50.

Pye, *C.* (1992). The acquisition of K'iche' Maya. In Slobin, D. (ed), *The Crosslinguistic Study of Language Acquisition, vol. 3,* pp. 221–308. Hillsdale, NJ: Lawrence Erlbaum Associates.

Quartz, S. R. & Sejnowski, T. J. (1997). The neural basis of cognitive development: A constructivist manifesto. *Behavioral and Brain Sciences, 20,* 537–596.

Querleu, D. (1985). Perception auditive du fœtus in utero. *Journal de Gynécologie et Obstétrique et Biologie Reproductrice, 14 (supplement),* 58–62.

Querleu, D., Renard, X., Boutteville, C. & Crépin, G. (1989). Hearing by the human fetus? *Seminars in Perinatology, 13,* 430–433.

Querleu, D., Renard, X., & Versyp, F. (1981). Les perceptions auditives du fœtus , humain. *Médecine et Hygiène,* 39, 2101–2110.

Rabinov, C. R., Alavi, S., Canalis, R. F. & Lee, E. J. (1997). Recurrent pleomorphic adenoma of the parotid gland involving the osseous external auditory canal. With a note on the foramen of Huschke [Pt 1]. *The Annals of Otology, Rhinology, and Laryngology, 106(7),* 589–593.

Rakova, M. (2002). The philosophy of embodied realism: A high price to pay? *Cognitive Linguistics, 13(3),* 215–244.

Ramachandran, V., & Rogers-Ramachandran, D. (2000). Phantom limbs and neural plasticity. *Archives of Neurology, 57,* 317–320.

Rosch, E. & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605.

Rosowski, J. J. (1996). Models of external- and middle-ear function. In Hawkins, H. L., McMullen, T. A., Popper, A. N. & Fay, R. R. (eds), *Auditory Computation*, pp. 15–61. New York: Springer-Verlag.

Rosowski, J. J. (1994). Outer and middle ears. In Fay, R. R. & Popper, A. N. (eds), *Comparative Hearing: Mammals*, pp. 172–247. New York: Springer-Verlag.

Rossano, F., Carpenter, M. & Tomasello, M. (2012). One-year-old infants follow others' voice direction. *Psychological Science, 23(11),* 1298–1302.

Rowlands, M. (2010). *The New Science of the Mind: From Extended Mind to Embodied Phenomenology.* Cambridge, MA: MIT Press.

Ruben, R. J. (1995). The ontogeny of human hearing. *International Journal of Pediatric Otorhinolaryngology, 32 (supplement),* S199–S204.

Ruben, R. J. (1992). The ontogeny of human hearing. *Acta Laryngologia, 112,* 192–196.

Saffran, J. R., Aslin, R. N. & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.

Sapir, E. (1949). *Selected Writings in Language, Culture and Personality*. Berkeley, CA: University of California Press.

Sapir, E. (1921 [1949]). *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace & World, Inc.

Saunders, J. C., Kaltenbach, J. A. & Relkin, E. M. (1983). The structural and functional development of the outer and middle ear. In Romand, R. (ed), *Development of Auditory and Vestibular Systems,* pp. 3–25. New York: Academic Press.

Scheuer, L. & Black, S. (2000). *Developmental Juvenile Osteology*. San Diego, CA: Academic Press.

Schiefflin, B. B. (1990). *The Give and Take of Everyday Life: Language Socialization of Kaluli Children.* Cambridge, England: Cambridge University Press.

Schneider, B. A. & Trehub, S. E. (1992). Sources of developmental change in auditory sensitivity. In Werner, L. A. & Rubel, E. W. (eds), *Developmental Psychoacoustics*, pp. 3–46. Washington, D.C.:

Schneider, B. A., Trehub, S. E. & Thorpe, L. (1991). Developmental perspectives on the localization and detection of auditory signals. *Perception and Psychophysics, 49,* 10–20.

Schroeder, M. R., Atal, B. S. & Hall, J. L. (1979). Objective measures of certain speech signal degradations based on masking properties of human auditory perception. In Lindblom, B. & Öhman, S. (eds), *Frontiers of Speech Communication Research,* pp. 217–229. London, England, U.K.: Academic Press.

Schultz, A. H. (1941). Growth and development of the orangutan. *Contributions to Embryology, 29(182),* 57–110.

Schultz, A. H. (1940). Growth and development of the chimpanzee. *Contributions to Embryology, 28(170),* 1–63.

Seidenberg, M. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275,* 1599–1603.

Seneff, S. (1988). A joint synchrony/mean-rate model of auditory speech processing. *Journal of Phonetics, 16,* 55–76.

Shaw, E. A. G. (1974). The external ear. In Keidel, W. D. & Neff, W. D. (eds), *Handbook of Sensory Physiology, Volume 5,* pp. 455–490. Berlin: Springer.

Shaw, E. A. G. & Stinson, M. R. (1983). The human external and middle ear: Models and concepts. In de Boer, E. & Viergever, M. A. (eds), *Mechanics of Hearing,* pp. 3–10. The Hague: Martinus Nijhoff Publishers.

Silverstein, M. (1981). *The Limits of Awareness.* Austin, TX: Southwestern Educational Laboratory.

Silverstein, M. (1976). Shifters, linguistic categories, and cultural description. In Basso, K. & Selby, H. (eds), *Meanings in Anthropology,* pp. 11–55. Albuquerque: University of New Mexico Press.

Simons, K. (1993). *Early Visual Development, Normal and Abnormal.* New York: Oxford University Press.

Sinclair, D. (1973). *Human Growth after Birth.* 2 ed. London, England: Oxford University Press.

Sinnott, J. M. & Aslin, R. N. (1985). Frequency and intensity discrimination in human infants and adults. *Journal of the Acoustical Society of America, 78,* 1986–1992.

Skinner, B. F. (1957). *Verbal Behavior*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Slaney, M. (1999). *Auditory Toolbox, Version 2.* Technical Report #1998-010, Interval Research Corporation, Palo Alto, CA.

Slater, A. (1998). The competent infant: Innate organisation and early learning in infant visual perception. In Slater, A. (ed), *Perceptual Development: Visual, Auditory, and Speech Perception in Infancy,* pp. 105–130. East Sussex, England: Psychology Press Ltd., Publishers.

Smith-Hefner, B. (1988). The linguistic socialization of Javanese children. *Anthropological Linguistics, 30,* 166–198.

Spence, M. J. & Freeman, M. S. (1996). Newborn infants prefer the maternal low-pass filtered voice, but not the maternal whispered voice. *Infant Behavior and Development, 19,* 199–212.

Spetner, N.B., & Olsho, L. W. (1990). Auditory frequency resolution in human infancy. *Child Development, 61,* 632–652.

Stein, J. & Walsh, V. (1997). To see but not to read; the magnocellular theory of dyslexia. *Trends in Neuroscience, 20,* 147–152.

Stevens, K. N. (2000). Diverse acoustic cues at consonantal landmarks. *Phonetica, 57,* 139–151.

Stevens, K. N. (1998). *Acoustic Phonetics.* Cambridge, MA: The MIT Press.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics, 1/2,* 3–46.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In Denes, P. B. & David, Jr., E. E. (eds), *Human Communication: A Unified View*, pp. 51–66. New York: McGraw Hill.

Strange, W. (1989). Evolving theories of vowel perception. *Journal of the Acoustical Society of America, 85,* 2081–2087.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology, 50(1),* 86–132.

Syrdal, A. & Gopal, H. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America, 79*, 1086–1100.

Szalay, F. S. & Delson E. (1979). *The Evolutionary History of the Primates.* New York: Academic Press.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12,* 49–100.

Talmy, L. (1983). How language structures space. In Pick, H. L. & Acredolo, L. P. (eds), *Spatial Orientation: Theory, Research, and Application*, pp. 225–282. New York: Plenum Publishing Corporation.

Tanner, J. M. (1989). *Fetus into Man: Physical Growth from Conception to Maturity*. Revised and enlarged. Cambridge, MA: Harvard University Press.

Taub, E., Uswatte, G., & Elbert, T. (2002). New treatments in neurorehabilitation founded on basic research. *Nature Reviews Neuroscience, 3*, 228–236.

Tharp, A. M. & Ashmead, D. H. (2001). A longitudinal investigation of infant auditory sensitivity. *American Journal of Audiology*, *10,* 104–112.

Tomasello, M. (2001). Perceiving intentions and learning words in the second year of life. In Bowerman, M. and Levinson, S. (eds), *Language Acquisition and Conceptual Development,* pp. 132–158. Cambridge, England: Cambridge University Press.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74(3),* 209–253.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition.* Cambridge, MA: Harvard University Press.

Tomasello, M. (1998). Introduction: A cognitive-functional perspective on language structure. In Tomasello, M. (ed), *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure,* pp. vii–xxiii. Mahwah, NJ: Lawrence Erlbaum Associates.

Torgersen, J. (1951). Asymmetry and skeletal maturation. *Acta Radiologica, 36,* 521–523.

Torre, C., Giacobini, G., & Ardito, G. (1978). Skeletal development of an orang-utan premature newborn: A comparative study with man. *Journal of Human Evolution, 7(2),* 143–149.

Tozoglu, U., Caglayan, F., & Harorli, A. (2012). Foramen tympanicum or foramen of Huschke: Anatomical cone beam CT study. *Dento Maxillo Facial Radiology*, 41(4), 294–297.

Traunmüller, H. (1984). Articulatory and perceptual factors controlling the age-and sex-conditioned variability in formant frequencies of vowels. *Speech Communication, 3,* 49–61.

Tsao, F-M, Liu, H-M, & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development, 75(4),* 1067–1084.

Turkewitz, G. & Devenny, D. (1993). Timing and the shape of development. In Turkewitz, G. & Devenny, D. (eds), *Developmental Time and Timing.* Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Turkewitz, G. & Kenny, P. (1986). The role of developmental limitations of sensory input on sensory/perceptual organization. *Journal of Developmental and Behavioral Pediatrics, 6,* 302–306.

Turkewitz, G. & Kenny, P. (1982). Limitations on input as a basis for neural organization and perceptual development: A preliminary theoretical statement. *Developmental Psychobiology, 15*, 357–368.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience.* Cambridge, MA: The MIT Press.

Vellutino, F. (1987 [1991]). Dyslexia. In Wang, W. S.-Y. (ed), *The Emergence of Language Development and Evolution*, pp. 159–170. New York: W. H. Freeman and Company.

Verbrugge, R., Strange, W., Shankweiler, D. & Edman, T. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America*, *60,* 198–212.

Wang, R. G., Bingham, B., Hawke, M., Kwok, P., & Li, J. R. (1991). Persistence of the foramen of Huschke in the adult: an osteological study. *The Journal of Otolaryngology, 20(4),* 251–253.

Weber, A. & Smits, R. (2003). Consonant and vowel confusion patters by American English listeners. In *Proceedings of the International Congress of Phonetic Sciences*, 1437–1440. Barcelona, Spain: Palau de Congressos.

Webster, D. B. & Webster, M. (1979). Effects of neonatal conductive hearing loss on brain stem auditory nuclei. *Annals of Otology, Rhinology, and Laryngology*, *88,* 684-688.

Webster, D. B., & Webster, M. (1977). Neonatal sound deprivation affects brain stem auditory nuclei. *Archives of Otolaryngology, 103,* 392–396.

Weir, C. (1979). Auditory frequency sensitivity of human newborns: Some data with improved acoustic and behavioral controls. *Perception and Psychophysics, 26,* 287–294.

Werker, J. F. & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24,* 672–683.

Werker, J. F. & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49–63.

Werner, L.A. (2007). Issues in human auditory development. *Journal of Communication Disorders, 40,* 275–283.

Werner, L. A., Fay, R. R. & Popper, A. N. (2012). *Human Auditory Development.* New York: Springer-Verlag.

Werner, L. A. & Gray, L. (1998). Behavioral studies of hearing development. In Rubel, E. W., Popper, A. N. & Fay, R. R. (eds), *Development of the Auditory System,* pp. 12–79. New York: Springer-Verlag.

Werner, L. A. & Marean, G. C. (1996). *Human Auditory Development.* Madison, WI: Brown & Benchmark, Publishers.

Werner, L. A. & Rubel, E. W., eds (1992). *Developmental Psychoacoustics.* Washington, D. C.: American Psychological Association.

Wertheimer, M. (1961). Psychomotor coordination of auditory and visual space at birth. *Science, 134,* 1692.

Westermann, G., Ruh, N. & Plunkett, K. (2009). Connectionist approaches to language learning. *Linguistics, 47(2),* 413–452.

Wexler, K. & Culicover, P. W. (1980). *Formal Principles of Language Acquisition.* Cambridge, MA: The MIT Press.

Williams, P. L., Warwick, R., Dyson, M. & Bannister, L. H., eds (1982). *Gray's Anatomy*. 37 ed. Edinburgh, Scotland: Churchill Livingstone.

Whorf, B. (1956). *Language, Thought, and Reality.* Cambridge, MA: The MIT Press.

Yakovlev, P. I. & Lecours, A.-R. (1967). The myelogenetic cycles of regional maturation of the brain. In Minkowski, A. (ed), *Regional Development of the Brain in Early Life.* Philadelphia: F. A. Davis Company.

Yost, W. A. (1994). *Fundamentals of Hearing: An Introduction*. 3 ed. San Diego: Academic Press.

Zahorian, S. & Jagharghi, A. (1993). Spectral-shape features versus formants as acoustic correlates for vowels. *Journal of the Acoustical Society of America, 94,* 1966–1982.

Zuercher, J. C., Carlson, E. V. & Killion, M. C. (1988). Small acoustic tubes: new approximations including isothermal and viscous effects. *Journal of the Acoustical Society of America, 83,* 1653–1660.