Age and Phonetic Context Effects in Children vs. Adults

by

Tessa Utz

B. A. Candidate in Communication Science and Disorders, University of Pittsburgh, 2009

Submitted to the Graduate Faculty of

Health and Rehabilitation Sciences in partial fulfillment

of the requirements for the degree of

Bachelor of Philosophy

University of Pittsburgh

2009

UNIVERSITY OF PITTSBURGH

School of Health and Rehabilitation Sciences

This thesis was presented

by

Tessa Utz

It was defended on

April 16, 2009

and approved by:

John Durrant, PhD., Communication Sciences and Disorders, University of Pittsburgh

Deborah Moncrieff, PhD., Communication Sciences and Disorders, University of Pittsburgh

Lori Holt, PhD., Psychology, Carnegie Mellon University

Thesis Advisor: Sheila Pratt, PhD., Communication Sciences and Disorders, University of Pittsburgh

Age and Phonetic Context Effects in Children vs. Adults

Tessa Utz

University of Pittsburgh, 2009

In normal hearing adults, acoustic context influences perception of speech in a spectrally contrastive manner. The aim of this study is to investigate whether typically developing children, aged 5 through 6 and 7 through 9 years, demonstrate phonetic context effects in a manner and extent similar to adults. By comparing the children's responses to those of adults aged 18 to 28 years, it will be determined if the children's use of phonetic context is limited by maturity.

A total of 61 individuals participated in this study: 45 adults and 16 children. The participants listened to isolated vowels along the /ʌ/ to /ɛ/ acoustic continuum and indicated if they heard /ʌ/ or /ɛ/. They then listened to the same vowel continuum within a /d/-Vowel-/d/ syllable context and in a /b/-Vowel-/b/ syllable context. With each syllable presentation the participants identified the vowel sound that they heard. The participant's responses were assessed for shifts in the vowel perceptual boundaries relative to consonant context.

The results indicated that the older children and the adults exhibited a context effect, but as a group, the younger children did not exhibit the effect. However, some of the younger children presented an effect that was consistent with the Older Children and Adults.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This defense would not have been possible without the guidance and support of my research advisor, Dr. Sheila Pratt. Thank you for taking the time, amongst your numerous other obligations, to share your research expertise with students like myself. Thank you also to the Faculty of Falk School for being so accommodating in allowing me to work with your students.

1.0 INTRODUCTION

1.1 SPEECH PERCEPTION IN CONTEXT

When perceiving speech, listeners rely on context to help them discern words and individual sounds. For example, if a listener cannot distinguish a certain word in a sentence, semantic cues within the sentence can be used to select the correct word over a phonologically similar word. Phonetic context cues also are used in the perception of speech. Typically sounds are not produced in isolation, but rather in an overlapping and continuous stream. During continuous speech there is movement overlap resulting in acoustic overlap of speech sounds. As a consequence, the perceived identity of a phonological unit is influenced by adjacent vowels or consonants (Holt & Lotto, 2002).

For example, vowels are especially reduced in ongoing speech produced at normal and fast speaking rates. Speakers rarely reach the target formant frequencies that are produced in isolation and are influenced more by the surrounding context (Lindblom, 1963; Lindblom & Studdert-Kennedy, 1967). For example, Lindblom (1963) found that Swedish vowels produced within /b/-Vowel-/b/ (/bVb/), /d/-Vowel-/d/, (/dVd/), and /g/-Vowel-/g/ (/gVg/) syllable contexts do not reach their normal steady-state formant frequencies. The vowels are reduced due to the physical limitations of the articulatory system and the need to produce a continuous stream of speech (Lindblom & Studdert-Kennedy, 1967).

Known as assimilation, vowels, as well as other phonological units, are modified to resemble surrounding context (Lindblom, 1963; Mann, 1980) in order to preserve economy of movement for speakers. For example, vowels produced in a CVC context with anterior consonants will assume a vowel that is produced more towards the front of the mouth, and this shift in placement has acoustic and perceptual consequences. Mann (1980), using the synthetic stimuli /al-da/, /al-ga/, /ar-da/, and /ar-ga/, discovered that the syllable-final liquids /l/ and /r/ affected listeners' place perception of /d/ and /g/ when they occurred after the liquids. Because the /l/ sound has a rather forward place of articulation, the stops following an /l/ also are produced with a more forward place of articulation. In the same manner, those stops following an /r/ production, which has a more posterior place of articulation, assume a more posterior place of articulation. The /d/ and /g/ also have opposing places of articulation; /d/ having an anterior place of articulation and /g/ having a posterior place of articulation. Therefore, assimilation occurs when sounds following /al/ (anterior) better resemble a /da/-like (anterior) production and sounds following /ar/ (posterior) better resemble a /ga/-like (posterior) production. Yet, a listener's perception compensates for these coarticulatory or assimilatory effects (Lindblom & Studdert-Kennedy, 1967; Mann, 1980). Even though the place of articulation is more anterior following /al/, listeners perceive more /g/s, a posterior articulation, following /al/ than they do /d/ sounds, a more anterior articulation. Correspondingly, more /d/ sounds are perceived following an /ar/ production.

The close correspondence between speech production and perception led some researchers to hypothesize that a person's perception is actually mediated by articulatory rather than acoustic events. Motor theorists such as Liberman, Cooper, Shankweiler, and Studdert-Kennedy (1967) suggested that the perception of sound resembles articulation more precisely than the acoustics of the sound and that speech perception is different from general auditory perception. To illustrate, Liberman et al. described how the consonant /d/ in /di/ and /du/ syllable contexts could be perceived as the same sound yet be acoustically distinct. The /d/ in both syllables is characterized by the second and third formant transitions, but the transitions vary substantively in frequency and direction across the two vowel contexts. As a consequence the transitions do not sound the same when presented in isolation, but within the context of the entire syllable they are recovered as /d/. Liberman et al. suggested that the transitions resemble the intended gestures of the vocal tract and are likely processed through the motor system. Thus, despite differing transitions, the initial /d/s in the two CV syllables are perceived as such due to invariance in speech gestures.

Similar to the Motor Theory, the Theory of Direct Perception states that vocal tract gestures themselves cause the structure in the acoustic speech signals, and that listeners directly perceive these gestures (Fowler, 1996). Even though consonant and vowel gestures are co-produced, they still structure the acoustic signal independently, and listeners can recover the gestures despite the temporal overlap of the two sounds. Fowler offered several pieces of evidence to support her Direct Realist Theory, one with a task eliciting the McGurk effect. A speaker mouthed a /da/, while an acoustic signal for /ma/ was simultaneously presented to listeners. Although the acoustic stimulus was /ma/ the listeners frequently reported hearing /na/, a response indicative of auditory-visual integration. Because /na/ shares the same alveolar place of articulation as the visually articulated /da/, Fowler argued that listeners directly perceive gestures and that vision, in addition to acoustics, affects the perception of some gestures.

Unlike the Motor theorists, Direct Realists do not argue that speech perception is different from other forms of perception (Fowler, 1996). For example, Fowler acknowledged that it is unlikely that animals perceive the human vocal tract while detecting acoustic speech signals (Fowler). As a result, Fowler's Direct Realist Theory has lost popularity because it fails to account for the lack of vocal tract knowledge needed to explain the similar perception of speech in animals and humans. Further weakening the Direct Realist and Motor Theories is that acoustic signals can be modified by other means than with the vocal tract (Diehl, Lotto, & Holt, 2004).

1.2 GENERAL AUDITORY MECHANISM

More recent research suggests that a general auditory mechanism can account for much of the research looking at perceptual compensation for coarticulation (Holt, 2005; Holt, 2006; Holt, Lotto, & Kluender, 2000; Lotto, Kluender, & Holt, 1997; Lotto, Sullivan, & Holt, 2003; Stephens & Holt, 2003). One argument for a general auditory mechanism is that if both speech and nonspeech contexts can alter the identification of an embedded or adjacent speech, then phonetic context effects arise from a general rather than a speech-specific perceptual mechanism (Holt, 2005; Stephens & Holt). Indeed, identification shifts are evident even within nonspeech and tonal contexts (Diehl et al., 2004; Holt, 2005; Holt, 2006; Holt & Lotto, 2002; Holt et al, 2000; Lindblom & Studdert-Kennedy, 1967; Stephens & Holt). For example, when the /ʌ/ - /ɛ/ vowel continuum occurs within a /dVd/ context the vowel is labeled more as /ʌ/ than /ɛ/ when compared to a /bVb/ context. The effect also is present when the context is speech or nonspeech acoustic analogues (Holt et al., 2000).

A link between the production and perception of speech cannot account for the effect with nonspeech stimuli because it lacks the spectral characteristics found in natural speech

(Stephens & Holt, 2003). The implication is that speech is perceived without respect to whether the signals originate from a vocal-tract source (Holt, 2006). Furthermore, speech-specific modules cannot elucidate the finding that vowel identification varies more within a CVC context with voiced consonants than with voiceless consonants. According to Holt et al. (2000), voiceless consonants have less of an effect on vowel perception than voiced consonants because they are lower in amplitude, and, as a result, offer less acoustic energy within context. Thus, even though voiced and voiceless consonants might share the same vocal tract formations because of the perceptual difference in amplitude, voiceless consonants do not affect neighboring vowels to the same extent as voiced consonants.

Another line of evidence supporting the role of a general auditory mechanism in compensating for coarticulation involves the testing of Japanese quail. These birds, which have no speech experience or human speech mechanisms, reveal the same results as humans in a replication of the Mann (1980) study. Two Japanese quail were trained to peck to CV syllables with a low-frequency onset, and the other two were trained to peck to CV syllables with a high-frequency onset. The birds trained to peck to high-frequency onsets pecked more vigorously for CVs following /ar/, whereas those trained to peck to low-frequency onsets pecked more vigorously for CVs following /al/ (Lotto et al., 1997). The finding that quails demonstrated results similar to humans supports the argument that compensation for coarticulation is a general auditory process and not human specific (Lotto et al.; Holt, 2006), discounting the Direct Realist and Motor Theory perspectives of speech perception being specific to humans. Further support for a general auditory mechanism is that Lotto and Kluender (1998) replicated Mann (1980) with nonspeech /al/ and /ar/ precursors resembling a male speaker and /ga/ and /da/ syllables produced by female speakers. Thus, the perception could not have been influenced solely by articulatory events (Lotto & Kluender).

1.3 CONTRAST

Holt (2005) argued that the manner in which our auditory systems compensate for coarticulation is based on spectral contrast. All sensorimotor systems, like audition and vision, recognize change (Coady, Kluender, & Rhode, 2003; Holt et al., 2000). For example, the human auditory system recognizes that longer vowels are followed by voiced consonants and shorter vowels are followed by voiceless consonants (Coady et al.). Frequencies that are not present in the precursor of an acoustic signal will be more salient relative to frequencies that are present in the precursor (Holt & Lotto, 2002). Such contrast is needed to compensate for the articulatory assimilation that occurs due to the physical constraints of speech production (Holt et al.; Lotto et al., 1997). Thus, a listener perceives a lower frequency sound following a higher frequency sound and vice versa. Holt et al. posited that neural adaptation might be responsible for these frequency contrasts. If auditory neurons fire and become adapted when excited by a preferred frequency, stimuli not matching the preferred frequency will create contrast. Holt (2006) recently argued that stimulus specific adaptation accounts for spectrally contrastive context effects and that it can occur rapidly with one stimulus or more slowly with multiple stimuli. The auditory neurons increase their response to a new stimulus, or enhance rare stimuli.

1.4 DEVELOPMENT

If our auditory systems are responsible for highlighting, or contrasting certain phonetic speech cues, then children might be limited by the immaturity of their auditory mechanisms (Mayo & Turk, 2004; Werner, 2007). Evidence suggests that children weight frequency contrasts differently than adults depending on context (Mayo & Turk, 2005; Nittrouer & Miller, 1997). Several hypotheses aim to clarify the discrepancies found in adult-child cue weighting. The Developmental Weighting Shift (DWS) hypothesis states that adult-child perceptual differences originate from different weighting strategies (Mayo & Turk, 2005; Nittrouer, 2002; Nittrouer & Miller; Nittrouer & Studdert-Kennedy, 1987). Conversely, Sussman (2001) argued that discrepancies in adult-child performance also could arise from general sensory/auditory processing differences between adults and children.

According to the DWS hypothesis, children initially process speech in large, global units, and over time learn to process speech in smaller units (Mayo & Turk, 2004; Nittrouer, 2002; Nittrouer, 2006; Nittrouer & Miller, 1997; Nittrouer & Studdert-Kennedy, 1987). Children attend to the dynamic resonance patterns produced by the vocal tract first because these cues are sufficient for speech perception. With experience children develop the ability to discriminate sounds by better utilizing acoustic information and recognizing more salient or static cues (Nittrouer, 2002; Nittrouer, 2004; Nittrouer, 2005; Nittrouer & Miller; Werner 2007). That is, they may initially struggle to process cues that lack physical distinctiveness (Mayo & Turk) by only focusing on the slow acoustic changes produced within the vocal tract. According to Nittrouer (2002), determining the phonetic structure of an acoustic signal is a skill not refined until about 10 years of age. Only as a child becomes familiar with anticipating coarticulated streams of speech does ongoing speech become more easily perceived.

With increased familiarity with the speech in their native language, children learn which acoustic cues are most beneficial. Miller and Nittrouer (1997) noted that children's acoustic cue weighting, in comparison to that of adult's, is less variable across different phonetic environments. Adult's cue weighting varies within different environments because they know which cues aid perception in different phonetic contexts and acoustic environments, and as a result are able to trade cues more easily. In contrast, children have to learn the informativeness or relevance of a cue before knowing when it can be applied.

In line with the DWS, it has been suggested that children weight formant transitions more heavily than adults (Hicks & Ohde, 2005; Nittrouer, 2002; Nittrouer, 2004; Nittrouer, 2005; Nittrouer, 2006; Nittrouer & Miller, 1997; Nittrouer & Studdert-Kennedy, 1987). Formant transitions capture the dynamics of the vocal tract, providing information about both the consonant and the vowel (Nittrouer, 2005; Nittrouer, 2006). According to Lindblom and Studdert-Kennedy (1967), formant transitions might serve as vital cues in processing speech. They reported that vowel stimuli sharing the same formant frequencies are labeled differently depending on the transitions to and from adjacent phones. Nittrouer and Studdert-Kennedy concluded that, overall, children assign greater weight to transitional cues than adults, even though groups of both children and adults weight frication noise more heavily than transitional cues in general. In their study, they measured the perceptual weight of vowel-onset formant transitions for the syllable contrasts /su/-/ʃu/ and /si/-/ʃi/. The transitions between /su/ and /ʃu/ can be described as informative or salient because the frequencies of the two syllables differ greatly, creating a distinctive transition. However, the formant transition differences between /si/ and /ʃi/ are not as informative because the onset frequencies are similar and create less of a sharp transition. Despite this informative differential, children were found to weight transitional cues more than adults for both sets of stimuli. According to Nittrouer and Studdert-Kennedy, spectral informativeness did not have any influence on the children's perceptions; supporting the idea that cue informativeness is something learned. Overall, the results suggested that children use different perceptual strategies than adults.

Similarly, Nittrouer (2004) found that when using synthetic stimuli children rely on formant transitions more than vocalic duration when discerning the voicing of final stops. Vowel length is a reliable cue for voicing of final stops: a longer vowel is followed by a voiced stop, and a shorter vowel is followed by a voiceless stop. So, vocalic duration can be described as an informative cue for voicing. Adults weighted the vocalic duration heavier than the 6-8 year old children in the study, who relied more on the F1 transitions. Overall, the weighting of vocalic duration increases with age. Nittrouer's experiment also suggested that learning plays a role in cue weighting. She applied the experiment to non-native language learners who don't recognize final stops or vocalic length cues in their native language. The non-native language learners, like the children in the experiment, didn't rely on vocalic length as much as native speakers. The heavy weighting of transitional cues by children might be characteristic of a tendency to attune to more global units of speech earlier in life or an immaturity in the ability to discern the informativeness of cues.

Nittrouer (2005) further investigated the weighting of acoustic cues using more naturalistic stimuli, namely, noise. She found no increase in the weighting of vocalic duration in noise at any age. So, even when the formant transition was somewhat masked by noise, the weighting did not shift to vocalic duration. The 6 and 8 year olds in the study could only attend to vocalic duration in quiet, but not in noise. Overall, this finding supports the notion that children first attend to the global aspects of speech. Although adults did not show a shift in their cue weighting with different listening conditions, weight shifted developmentally.

The intricacies of auditory processing are not fully developed until late childhood or adolescence (Nittrouer & Miller, 1997; Sussman, 2001; Werner, 2007), but despite support found for the Developmental Weighting Shift Hypothesis, others believe that the cue weighting differences between adults and children arises from general auditory processing differences instead of different cue weighting strategies. Opponents to the DWS suggest that children do not initially use global cues such as formant transitions, and that informativeness may not have to be learned but actually contributes to initial perception (Sussman, 2001). Some studies have illustrated that children do not prefer to use transitions in every situation (Mayo & Turk, 2004; Mayo & Turk, 2005; Sussman, 2001). Mayo and Turk (2004) compared adults' and children's cue weighting for the contrasts /saɪ/-/ʃaɪ/, /de/-/be/, /ta/-/da/, and /ti/-/di/. They determined that children weight transitions more heavily than adults on half of the stimuli; the /saɪ/-/ʃaɪ/ and /ta/-/da/ contrasts, but not the /de/-/be/ or the /ti/-/di/ contrasts.

Furthermore, Sussman (2001), in a study of adults and children aged 4-5 years, ascertained that children were more accurate in vowel identification when the vowel was presented in a steady-state center than when in a transition. In fact, the children in her study actually used the steady-states to identify the vowels more than did the adults, and they also used the steady-states to identify vowels more than they used transitional cues. Therefore, children's weighting tendencies might be attributed to the informativeness of the cues and not to the dynamics of the vocal tract during speech. Children might require longer, more stable cues, such as steady states, or more spectral information than adults. Nittrouer (2005), however, offered another possible explanation for Sussman's findings, arguing that the time associated with the steady-state portion of the stimuli overwhelmed the time associated with the formant transitions. Overall, children might require a greater amount of acoustic cues than adults because the immaturity of their auditory systems might inhibit them from distinguishing between ambiguous acoustic signals.

A problem arises when comparing support for the DWS and Sussman position because Nittrouer and Sussman used different types of speech sounds in their studies. The DWS Hypothesis is based on results from studies looking at fricative perception, whereas Sussman studied vowel contrasts. Thus, the results of the aforementioned studies lead to no definitive conclusion: a child might rely on transitions (Hicks & Ohde, 2005; Nittrouer, 2002; Nittrouer, 2004; Nittrouer, 2006; Lindblom & Studdert-Kennedy, 1967; Nittrouer & Miller, 1997), especially for the identification of fricative contrasts (Nittrouer & Studdert-Kennedy, 1987) and on steady-state cues for the identification of vowels (Sussman, 2001).

Mayo and Turk (2005) conducted a study to further investigate the conclusions of these previous experiments. Specifically, they assessed the role of cue informativeness in children's perception of speech. Because the DWS hypothesis was largely based on studies of fricative contrasts, and Sussman's hypothesis resulted from vowel contrasts, Mayo and Turk combined aspects of both experiments, focusing on the informativeness of different consonant-vowel contrasts. Their experiment used spectrally distinct stimuli (/no/-/mo/, /do/-/bo/, and /ta/-/da/) and spectrally similar stimuli (/ni/-/mi/, /de/-/be/, and /ti/-/di/) to compare the weighting results of children ages 3-7 to those of adults. That is, the vowel-onset transitions in /no/ and /mo/ are very distinct, making them especially informative and the nasals easy to identify. The vowel-onset transitions in /ni/ versus /mi/ are very similar, making the transition less informative and the nasal more difficult to identify.

As expected, the spectrally distinct transitions (/no/-/mo/, /do/-/bo/, and /ta/-/da/) influenced both children and adults more than the spectrally similar transitions (/ni/-/mi/, /de/-/be/, and /ti/-/di/). That is, the distinct cues lead to more influence on consonant labeling than did the spectrally similar cues. In accordance with Sussman's (2001) finding, children did weight the spectrally similar cues in /ni/-/mi/ and /de/-/be/ less than adults. In other words, they were less able to distinguish one consonant from another in the context of spectrally similar transitional cues, and therefore, they might not demonstrate the ability to utilize those less informative cues. However, this pattern did not apply to all of the contrasts. The findings for the spectrally similar /ti/-/di/ contrast revealed no differences in weighting between children and adults. Furthermore, children weighted the spectrally informative transitions of /ta/-/da/ less than adults, despite Sussman's hypothesis that children should weight more informative cues more heavily. So, cue informativeness alone cannot explain children's weighting strategies in every phonetic context. Nevertheless, Nittrouer's DWS hypothesis also was not successful in accounting for all of the results of the experiment. The children did not weight the transitional cues in /ti/-/di/ or /ta/-/da/ any more heavily than adults. Therefore, solely transition-based hypotheses for adult-child cue weighting strategies failed to explicate the results of this study (Mayo & Turk, 2005). It is possible that a mechanism more basic, such as spectral or phonetic contrast, might account for the observed developmental patterns observed in children.

To that end the following preliminary work was conducted to see if children demonstrate the phonetic contrast effect in the same manner as young adults and whether children differ by age. Although the previous research did not test the phonetic context effect directly, it did show that children are sensitive to acoustic contrast and that along many speech continua they differ from adults and by age in their responsiveness to it.

2.0 HYPOTHESIS

Adults are expected to identify vowels within CVC syllables in a spectrally contrastive manner. Children aged 5-6 and 7-9 years will not differ from the young adults or from each other. More specifically, in the context of /dVd/ syllables, adults will perceive more /ʌ/ vowels along the /ʌ/- /ɛ/ continuum than when presented in a /dVd/ syllable context than in the /bVb/ context, despite having identical acoustical information for the steady portion of the vowels in each context. The perceptual shifts introduced by the phonetic contexts are not expected to differ across the three groups.

Such results would suggest that children are sensitive to the formant transitions in a manner similar to the adults and are influenced by the level of spectral contrast produced by the transitions. If, on the contrary, children do not exhibit phonetic context effects or exhibit less of an effect, it would suggest that children are paying more attention to the steady-state vowel than to the formant transitions and that they are less influenced by the spectral contrast. The children might produce a mixture of results, which could be indicative of their inability to consistently use informative acoustic cues.

3.0 METHODS

3.1 PARTICIPANTS

This study was approved by the University of Pittsburgh IRB. Two groups of adults, both aged 18-28 years, and two groups of children, aged 5 through 6 (Younger Children) and 7 through 9 years (Older Children), were recruited to this study. The adults were recruited through course announcements at the University of Pittsburgh and the children were recruited through the Falk School, which is a laboratory school affiliated with the University of Pittsburgh. A letter from the school principal, study information and consent forms were sent home to the children's parents, and if the parents were interested in having their children participate they returned signed consent forms to the school. Extra credit was offered to the adults but no compensation was provided to the children for participating in the study. In all, 49 adults and 18 children produced signed consent forms. The first group of adults (Group A) consisted of 22 undergraduate students and the second group (Group B) consisted of 27 graduate students. All of the adult and pediatric participants were native English-speakers with no history of speech, language, or hearing problems.

Prior to completing the experiment the participants received an otoscopic exam, a screening tympanogram (226 Hz, WlechAllyn Mincortymp 2) and a puretone hearing screening in each ear (ASHA, 1997). The puretone screening was completed in a quiet room with a portable audiometer (Maico MA25) and ER3 insert earphones. The screening signals consisted of 500, 1000, 2000, 4000 and 8000 Hz tones presented at 25 dB HL. The participants also completed the Edgerton and Danhauer *Nonsense Syllable Test* (Edgerton & Danhauer, 1979) administered at 65 dB HL, and to be included in the study they had to perform at or above 75% correct. In addition, a non-word repetition task (Dollaghan & Campbell, 1998) was completed to account for auditory working memory skills. Group performances on the Edgerton and Danhauer *Nonsense Syllable Test* and nonword repetition task are provided in Table 1 and Table 2, respectively.

Table 1. Mean percent correct on the Edgerton and Danhauer Nonsense Syllable Test

Group	Ear	
	Right	Left
Younger Children	83.50	86.10
Older Children	82.90	84.80
Adults, Group A	85.95	91.00
Adults, Group B	88.08	89.67

Four adult and 2 pediatric participants were excluded from the study. One participant from each adult group was eliminated due to incomplete data storage by the computer during the experiment, and three adults from the second adult group did not meet the inclusion criteria. Two of the 18 children also were excluded: One child was terminated from the study during the screening procedures due to loss of consent, and a second child was unable to complete all phases of the experiment because of illness. Complete group data is provided in Tables 3 and 4.

Table 2. Average number of errors on Nonword Repetition Task

Group	Number of Syllables			
	1	2	3	4
Younger Children	3.10	2.60	2.30	7.60
Older Children	2.19	2.83	3.83	7.67
Adults, Group A	1.57	1.95	0.29	6.38
Adults, Group B	1.08	0.79	0.63	3.71

Table 3. Adult Demographic Information

Participant	Demographic Characteristics		
Adults, Group A	Age	Race	Sex
A1	21	Caucasian	F
A2	21	Caucasian	F
A3	21	Caucasian	F
A4	22	Caucasian	M

A5	21	Caucasian	F
A6	21	Caucasian	M
A7	19	Caucasian	M
A8	21	Caucasian/Asian	F
A9	19	Caucasian	F
A10	20	Caucasian	F
A11	20	Caucasian	F
A12	21	Caucasian	F
A13	21	Caucasian	F
A14	22	Caucasian	F
A15	22	Caucasian	F
A16	21	Caucasian	M
A17	25	Caucasian	F
A18	21	Caucasian	F
A19	21	Caucasian	M
A20	22	Caucasian	M
A21	22	Caucasian	M
	Mean=21.14		F=14, M=7
Adults, Group B			
B1	22	Caucasian	F
B2	22	Caucasian	F
B3	22	Caucasian	F

B4	22	Caucasian	F
B5	23	Caucasian	F
B6	22	Caucasian	F
B7	25	Asian Indian	F
B8	24	Caucasian	F
B9	23	Caucasian	F
B10	23	Caucasian	F
B11	23	Caucasian	F
B12	23	Caucasian	F
B13	22	Caucasian	F
B14	24	African American	F
B15	23	Caucasian	F
B16	23	Caucasian	F
B17	22	Caucasian	F
B18	22	Caucasian	F
B19	23	Caucasian	F
B20	23	Caucasian	F
B21	22	Caucasian	F
B22	22	Caucasian	F
B23	22	Caucasian	F
B24	22	Caucasian	F
B25	24	Caucasian	F
	Mean = 22.83		F=25, M=0

Table 4. Pediatric Demographic Information

Participant	Demographic Information		
Older Children (7-9 yrs)	Age	Race	Sex
O1	7	Caucasian	M
O2	7	Caucasian	F
O3	7	Caucasian	F
O4	7	Caucasian	M
O5	8	Caucasian	M
O6	9	Caucasian	F
	Mean=7.5		F=3, M=3
Younger Children (5-6 yrs)			
Y7	5	Caucasian	M
Y8	5	Caucasian	F
Y9	6	Caucasian	F
Y10	6	Caucasian	M
Y11	6	Caucasian	F
Y12	6	Caucasian	F
Y13	6	Caucasian	F
Y14	6	Caucasian	M
Y15	6	Caucasian	F

Y16	6	Caucasian	F
	Mean=5.8		F=7, M=3

3.2 STIMULI

This project used synthesized syllables constructed with a Klatt parametric formant synthesizer (Klatt, 1980; Klatt & Klatt, 1987) embedded within the HLsyn formant synthesizer program (Sensimetrics, v2.2). A 12-step series of vowels across the /ʌ/ - /ɛ/ continuum was constructed with the F2 center frequency increasing from 1210 to 1760 in 50 Hz steps. The vowel series was constructed in isolation with no transitions, and in /bVb/ and /dVd/ syllables. The parameters used to construct the stimuli were consistent with those used by Holt et al. (2000) except that the continuum was extended at each end by an additional step. The extra steps were added to allow slightly larger category ranges, because it was not clear where on the continuum the children would locate their categorical boundaries, and how stable their responses would be within category. The primary distinction between the /bVb/ and /dVd/ series was the F2 onset and offset frequency. For the /bVb/ stimuli, the F2 onset and offset was relatively low in frequency (800 Hz), whereas for /dVd/ stimuli a higher frequency transition locus (2270 Hz) was used.

3.3 PROCEDURES

The participants were first trained to identify isolated /ɛ/ and /ʌ/ vowels from the endpoints of the continuum. The participants were asked to press one of two buttons on an enlarged computer mouse corresponding to which vowel they perceived. The buttons were labeled “eh” for /ɛ/ and “uh” for /ʌ/. Each training block included 10 trials, and participants were provided visual feedback on each trial: “Correct” or “Sorry, Not Correct”. All of the participants were first trained on the endpoint vowels in isolation and then within /bVb/ and /dVd/ syllables, with the training order for the /bVb/ and /dVd/ syllables counterbalanced across participants within each age-group.

Following the training, the participants identified target vowel stimuli presented from along the entire vowel continuum. The stimuli were presented in three blocks: vowels in isolation, /bVb/ syllable, and /dVd/ syllables. The isolated vowels were presented first and the /bVb/ and /dVd/ blocks followed in counterbalanced order per group. Within each block each stimulus was presented 10 times in random order without replacement, for a total of 120 trials per block. In all, each participant completed 360 experimental trials.

The stimulus presentation and response record was controlled by SuperLab software (Cedrus, v 2.0) residing on a laptop computer. The acoustic stimuli were routed from the computer audio port to an earphone amplifier (Behringer MiniAMP, AMP800) and presented at 65 dB SPL under circumaural earphones (Radio Shack, 04A08). During the training the visual display on the laptop was used to provide feedback, but during the experimental tasks the display was blank. The identification tasks lasted approximately 30 minutes, not including breaks. The adults were tested in a quiet laboratory space at the University of Pittsburgh and the children were tested in a quiet room at their school.

All participants within the first group of adults (Group A) completed the isolated vowel, /bVb/, and /dVd/ identification tasks in one session with no breaks. The children were tested with breaks between tasks. The breaks were approximately 10 minutes in duration if completed in the same day, although a number of the children required two sessions due to attention and behavioral issues.

The adult participants in Group B, however, completed the isolated vowel identification task during the first session and the /bVb/ and /dVd/ identification tasks in a second session. Breaks were given between the /bVb/ and /dVd/ identification tasks for all but four of the Group B adults.

4.0 ANALYSES

Each participant's responses were totaled for each vowel token within each block. A Probit analysis (Finney, 1971), with a base-10 log transform (SPSS, v 16) was applied to the individual data in order to estimate the 50% point on the categorical boundaries (identification threshold). ANOVAs were then applied to these identification threshold data to test for context and group effects. As a follow-up measure the number of participants showing the anticipated effect in each group also was assessed with a Pearson's X^2.

5.0 RESULTS

5.1 ADULTS- GROUP A

Upon preliminary inspection the mean vowel identification curves failed to show the expected context effects for vowel identification in the /bVb/ and /dVd/ syllables. In general, the adult listeners demonstrated no effect or the opposite effect than expected (Figure 1). That is, some listeners reported hearing more low-frequency /ʌ/ vowels in the low-frequency /bVb/ context and more high- frequency /ɛ/ vowels in a high-frequency /d/ consonant context. In contrast, a preliminary view of the children's data showed that they were exhibiting the desired effects. The most obvious difference between the adults and children was that the children had been given breaks between blocks. A concern was that the isolated vowel identification task had influenced the adult responses in the two subsequent blocks. So the second group of adults was enrolled with a forced break between the isolated vowel block and vowels in context blocks. Enrolling a new group of adult participants was considered critical because the adult group served as a developmental control in the study.

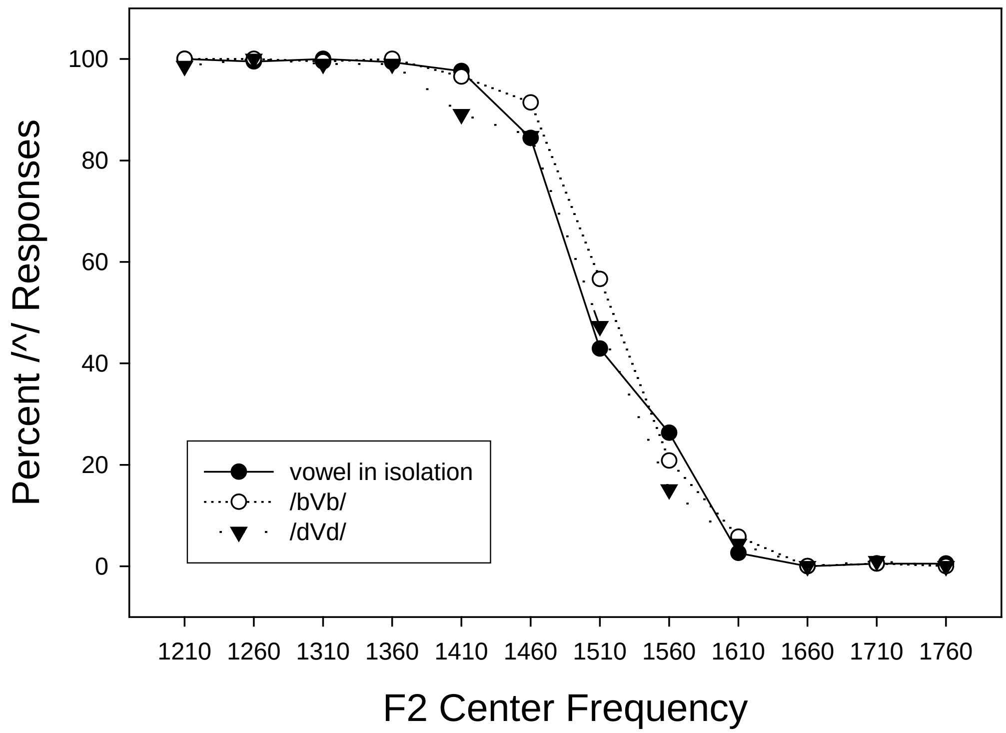

Figure 1. Mean vowel identification functions for the Group A Adults

5.2 ADULTS (GROUP B) AND PEDIATRIC RESULTS

In order to first document the extent to which the three groups perceived vowels equivalently, the identification thresholds for the vowels in isolation were compared across the three groups with a one-way ANOVA. No group difference was observed, F=0.05, df=2,38, p=.951, and the mean thresholds were quite similar across the groups: Adults - 1494 Hz, Older Children - 1492 Hz, and Younger Children - 1491 Hz (See Figure 2). A two-way ANOVA with repetition on one factor (context) was then applied to the identification thresholds for the vowel continuum within the

/bVb/ and /dVd/ contexts. Context was significant, *F*=10.778, *df*=1,38, *p*=002, but a context by group interaction was not observed, *F*=1.349, *df*=2,38, *p*=.272. The group factor did not reach significance at the .05 level, *F*=2.915, *df*=2,38, *p*=.066, but given the limited number of pediatric participants, the group results warranted further investigation (See Figures 3-4). Post hoc tests showed differences between the Young Children and the Adults (*p*=.029) but not between the two pediatric groups (*p*=.069) or the Older Children and the Adults (*p*=.802). In addition, paired t-tests showed that the /dVd/ identification thresholds were higher than /bVb/ thresholds for the Adults, *t*=3.585, *df*=1,24, *p*=.001, and Older Children, *t*=3.20, *df*=1,5, *p*=.024, but not for the Younger Children, *t*=.429, *df*=1,9, *p*=.678. Visual examination of the mean thresholds for each contrast suggested that group differences might be associated with the /dVd/ context, so the groups were compared separately within each context with one-way ANOVAs. The ANOVA results were not significant (/dVd/ context: *F*=2.917, *df*=2,38, *p*=.066 and /bVb/ context: *F*=.723, *df*=2, 38), but post hoc comparisons showed that the thresholds for the Younger Children in the /dVd/ context were significantly lower than for the Older Children (*p*=.038) and the Adults (*p*=.047). The Older Children and the Adults were not different (*p*=.457).

It should be noted that not all of the Adults or the Younger Children demonstrated the desired context effect. That is, the vowel identification threshold for the /dVd/ context was expected to be higher than the threshold for the /bVb/ context. To examine this further, each participant was classified as either having a higher threshold for the /dVd/ context than the /bVb/ context or a higher threshold for the /bVb/ context. The counts for each classification per group are shown in Figure 5, and although the figure is suggestive of group differences, a Pearson's X^2 did not confirm this observation (*exact p*=.224). To further investigate the participants demonstrating the expected context effect, the threshold data were re-examined with the exclusion of participants failing to show the effect. As can be seen in Figure 6, the mean thresholds for the three groups were very similar and no group differences were observed for either context. Furthermore, all three modified groups produced significant context effects. The implication is that if the expected effect is present the location of the categorical boundary and the extent of the shift are similar regardless of age.

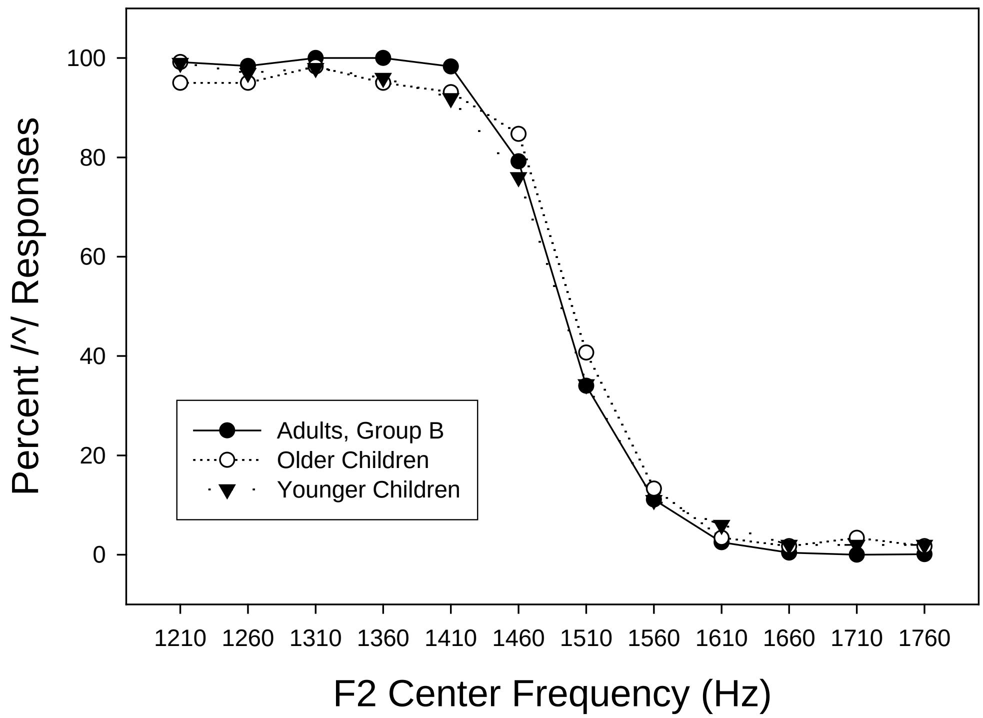

Figure 2. Identification /ʌ/ in isolation

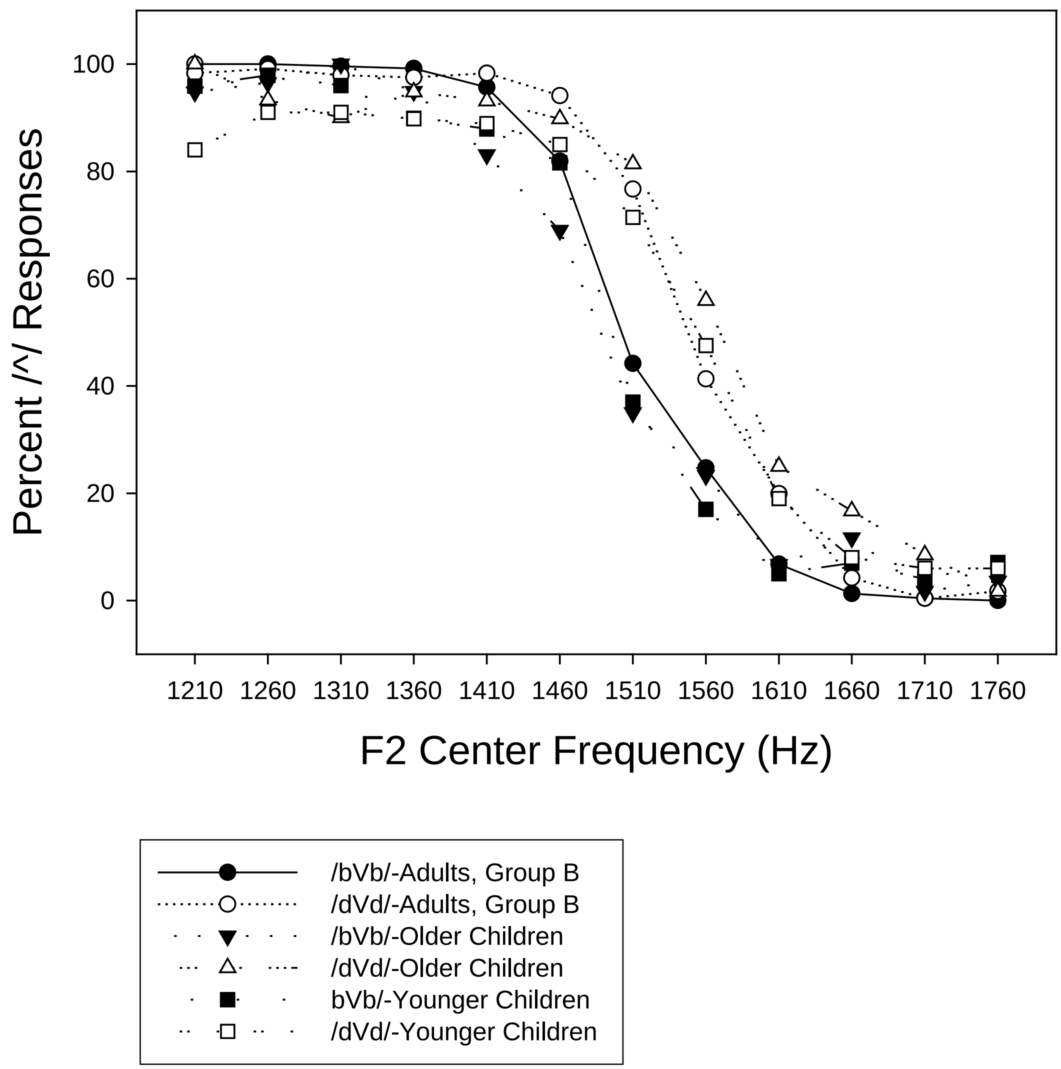

Figure 3. Identification /ʌ/ in /bVb/ and /dVd/ Syllables

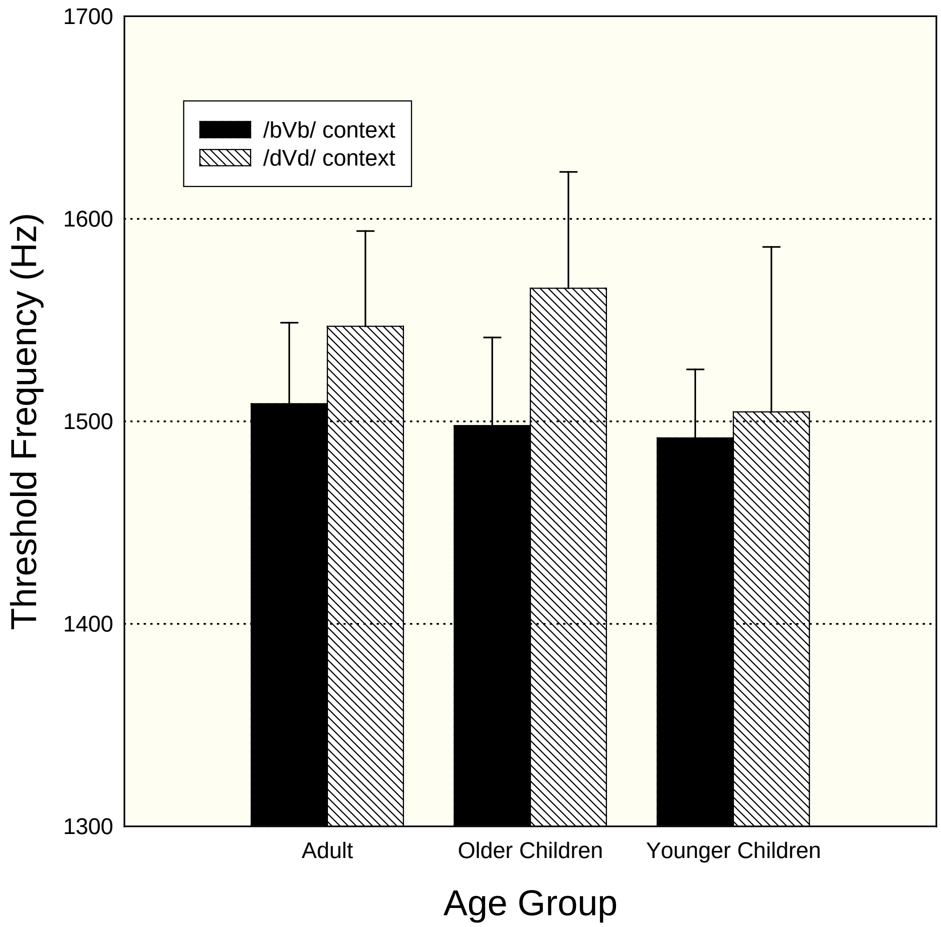

Figure 4. Mean vowel identification thresholds

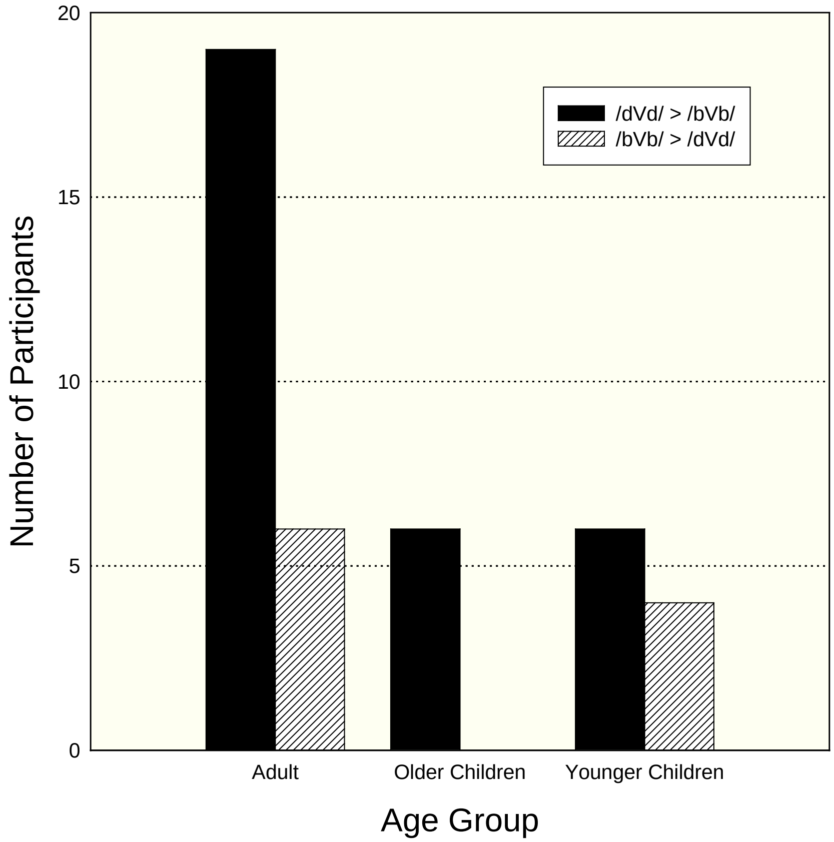

Figure 5. Number of participants with the expected context effect

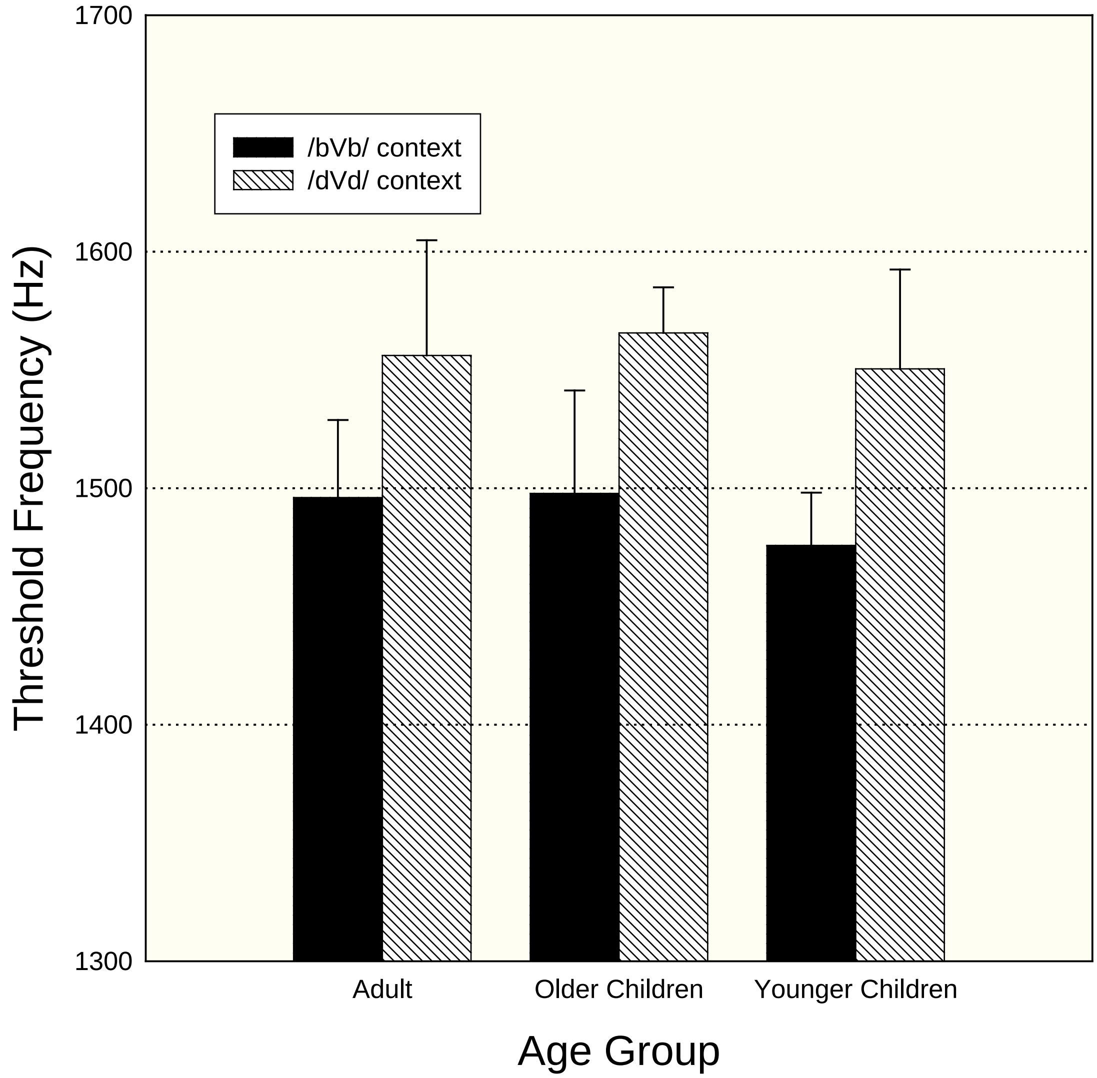

Figure 6. Mean vowel identification thresholds (participants showing expected context effects)

6.0 DISCUSSION

The Older Children performed in a manner similar to the Adults in Group B, but despite the lack of an overall group effect, the Younger Children appeared to differ from the Older Children and the Adults as a group. That is, the Older Children and the Adults demonstrated the expected context effect, but the Younger Children did not. Because no group differences were observed between the identification thresholds for the vowels in isolation and the thresholds were similar, differences or similarities between a group's identification of vowels in context can be attributed to the effect of the context itself, and not to differences in vowel labeling. As seen in Holt et al.'s (2000) study, context had a significant effect on the identification of vowels. Specifically, participants identified vowels in a spectrally contrastive manner. Participants in Adult Group B and the Older Children reported hearing more low-frequency /ʌ/ vowels in a high-frequency /d/ context when compared to the more low-frequency /b/ context. As a group, the Young Children failed to show this effect although many of the children in this group did demonstrate it. In comparing the pediatric participants to the Adults in Group B, the results from most of the pediatric participants support the idea that children can attend to formant transitions in different CVC syllable contexts in a manner similar to adults. However in the younger children, auditory maturity might have reduced responsiveness to acoustic contrast in the /dVd/ syllables.

Post-hoc examination of the data showed that as a group the Young Children did not demonstrate spectrally contrastive context effects to the same extent as the Older Children or the

Adults. Specifically, the thresholds for the Younger Children in the /dVd/ context were significantly lower than for both the Older Children and the Adults. That the Young Children in the post hoc comparison produced results different from both the Older Children and the Adults suggests that at least some of the Young Children might still have difficulty in perceiving spectral contrast or consistently using the same cues within each context. Although this study did not test for cue weighting strategies directly, the results suggest that a few of the Young Children could have possibly been attending more to the steady-state vowel in the /dVd/ context than to the formant transitions, as seen with Sussman's (2001) results. Alternatively, and more likely, these differences might reflect the lesser attention spans of the Young Children and their inability to focus on the labeling tasks for the approximate 8 minutes that each identification block took to complete. Furthermore, being mindful of which response button corresponded to the /ɛ/ or /ʌ/ vowels might have been too difficult for some of the Young Children.

As stated in the Results section, not all participants demonstrated the desired context effect. This may be due, again, to the limitations of the children's attention or cognitive skills, or to methods differences from Holt et al. (2000) study. Holt et al. used a 10-step series of vowel stimuli, ranging from 1260 Hz for /ʌ/ and to 1710 for /ɛ/, while the current study implemented a 12-step series of vowel stimuli from 1210 Hz for /ʌ/ and 1760 for /ɛ/. The participants in Holt et al.'s (2000) study identified each stimulus 20 times, for a total of 400 responses, while participants in the current study identified each stimulus 10 times, for a total of 360 responses. Importantly, the total reported 360 stimuli in the current study included 120 stimuli of the vowel identification in isolation, a block not included in the Holt et al.'s (2000) study. Thus, each participant identified 240 vowels within a /b/ and /d/ context in the current study, but each participant in Holt et al. (2000) identified 400 vowels across the /b/ and /d/ contexts. Participants not demonstrating the desired context effect might have provided different results if tested over more trials.

Although not all participants in this study showed the anticipated effect, differences between groups in the number of participants showing the desired effect did not differ significantly. Thus, most Young Children did, in fact, show the expected results, and aforementioned group differences may be attributed to the few Young Children whose results were especially skewed, possibly due to lack of proficiency on the task itself and not the lack of context effects. In fact, in analyzing only the data of those participants illustrating the desired effect, no significant group differences occur in either the /b/ or the /d/ contexts. In conclusion, the children are not limited by their ability to highlight or contrast phonetic speech cues. Children do not differ from adults in their responsiveness to acoustic contrast in a /bVb/ and /dVd/ syllable context, and they, indeed, are influenced by the level of spectral contrast in the formant transitions. Supporting the idea of global speech processing as purported by the Developmental Weighting Shift Hypothesis, children are able to consistently weight the formant transitions in a manner similar to adults in each context.

The unexpected results seen in the study was the result produced by the Group A adults. They might have been influenced by their completion of the /ʌ/- /ɛ/ identification task prior to the following /b/ and /d/ identification tasks. When the adults in Group B completed the /ʌ/- /ɛ/ identification task separate from the /b/ and /d/ identification tasks, they demonstrated the previously supported findings of Holt et al. (2000). Thus, forcing a break between the /ʌ/- /ɛ/ identification task and the /b/ and /d/ identification tasks likely accounted for the discrepancy between the results of Adult Group A and Adult Group B. Order of presentation of the /b/ and /d/ identification tasks was likely not a factor, as it was counterbalanced across participants.

BIBLIOGRAPHY

American Speech-Language-Hearing Association (1997). Guidelines for Audiologic Screening [Guidelines]. Available from www.asha.org/policy.

Coady, J.A., Kluender, K.R. & Rhode, W.S. (2003). Effects of contrast between onsets of speech and other complex spectra. *Journal of the Acoustical Society of America, 114,* 2225-2235.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *The Annual Review of Psychhology, 55*, 149-179.

Dollaghan, C. & Campbell, T. (1998). Nonword repetition and language impairment. *Journal of Speech, Language, & Hearing Research, 41*, 1136-1146.

Edgerton, B. & Danhauer, J. (1979). Nonsense Syllable Test. St. Louis, MO: Auditec of St. Louis.

Finney, D. J. (1971). Probit Analysis. Cambridge: Cambridge University Press.

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America, 99 (3),* 1730-1741.

Hicks, C. B., & Ohde, R. N. (2005). Developmental role of static, dynamic, and contextual cues in speech perception. *Journal of Speech, Language, and Hearing Research, 48,* 960-974.

Holt, L. L., (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *The American Psychological Society, 16 (4),* 305-312.

Holt, L. L. (2006). The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization. *The Journal of the Acoustical Society of America, 120 (5),* 2801-2817.

Holt, L. L., & Lotto, A. J. (2002). Behavioral examinations of the level of auditory processing of speech context effects. *Hearing Research, 167,* 156-169.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America, 108 (2),* 710-722.

Klatt, D.H. (1980). Software for a cascade/parallel formant synthesizer, *Journal of the Acoustical Society of America, 67*, 971-995.

Klatt, D.H. & Klatt, L.C. (1987). Analysis, synthesis and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of American, 87*, 820-857.

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

Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America, 35 (11),* 1773- 1781.

Lindblom, B. E. F., & Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. *The Journal of the Acoustical Society of America, 42 (4),* 830- 843.

Lotto, A. J. & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics, 60,* 602-619.

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica). Journal of the Acoustical Society of America, 102*, 1134-1140.

Lotto, A., Sullivan, S., & Holt, L. L. (2003). Central locus for nonspeech context effects on phonetic identification. *Journal of Acoustical Society of America, 113 (1),* 53-56.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics, 28 (5),* 407-412.

Mayo, C. & Turk, A. (2004). Adult-child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions. *The Journal of the Acoustical Society of America, 115 (6),* 3184- 3194.

Mayo, C., & Turk, A. (2005). The influence of spectral distinctiveness on acoustic cue weighting in children's and adults' speech perception. *Journal of the Acoustical Society of America, 118 (3),* 1730-1741.

Nittrouer, S. (2002). From ear to cortex: A perspective on what clinicians need to understand aboust speech perception and language processing. *Language, Speech, and Hearing Services in Schools, (33),* 237-252.

Nittrouer, S. (2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *Journal of the Acoustical Society of America, 115 (4),* 1777-1790.

Nittrouer, S. (2005). Age-related differences in weighting and masking of two cues to word-final stop voicing in noise. *Journal of the Acoustical Society of America, 118 (2),* 1072-1088.

Nittrouer, S. (2006). Children hear the forest. *Journal of the Acoustical Society of America, 120,* 1799-1802.

Nittrouer, S., & Miller, M. E. (1997). Developmental weighting shifts for noise components of fricative-vowel syllables. *Journal of the Acoustical Society of America, 102,* 572-580.

Nittrouer, S., & Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *The Journal of Speech and Hearing Research, 30,* 319-329.

Stephens, J. D. W., & Holt, L. L. (2003). Preceding phonetic context affects perception of nonspeech. *Journal of the Acoustical Society of America, 114 (6),* 3036-3040.

Sussman, J. E. (2001). Vowel perception by adults and children with normal language and specific language impairment: Based on steady states or transitions? *Journal of the Acoustical Society of America, 109 (3),* 1173-1180.

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