# ABSTRACT

| | |
|---|---|
| Title of Document: | VOICE ONSET TIME IN INFANT-DIRECTED SPEECH AT TWO AGES |
| | Anna Synnestvedt, Master of Arts, 2010 |
| Thesis Directed by: | Dr. Nan Bernstein Ratner<br>Department of Hearing and Speech Sciences |

Studies have reported differences between infant-directed speech (IDS) and adult-directed speech (ADS), suggesting that mothers adjust speech to their infants in ways that may help children process the incoming acoustical signal. One aspect of IDS that has been examined is clarification of voice onset time (VOT). Results have been inconsistent and many studies only report differences in VOT values rather than differences in amount of overlap between voiced and voiceless items. The present study examines 15 mothers' VOT in IDS at 7.5 months old and again at 11 months as compared to their VOT values in ADS. Words with initial stop consonants that occurred in IDS and ADS conditions were analyzed using PRAAT. Contrary to hypotheses, results show that VOT in IDS was less differentiated than VOT in ADS. Additionally, voiced items had significantly longer VOT in IDS than ADS, with no difference for voiceless items. Possible explanations are discussed.

# VOICE ONSET TIME IN INFANT-DIRECTED SPEECH AT TWO AGES

by

Anna Synnestvedt

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Arts
2010

Advisory Committee:
Dr. Nan Bernstein Ratner, Chair
Dr. Rochelle Newman
Dr. Monita Chatterjee

# Table of Contents

# List of Tables

# List of Figures

# *Introduction*

Infant-directed speech (IDS), the special speech register that adults in many cultures use when speaking to infants, is a fairly well-studied phenomenon. Prosodic, acoustic-phonetic, and lexical characteristics of IDS distinguish it from adult-directed speech (ADS), the speech register that adults generally use when speaking to each other. Furthermore, IDS is distinct from speech directed to pets and speech directed to listeners of a different language background, having more clearly articulated vowels than speech to animals (Burnham, Kitamura, & Vollmer-Conna, 2002), and higher pitch and rated affect than speech to foreigners (Uther, Knoll, & Burnham, 2007). Though very few causal studies have been conducted, several theories exist regarding the possible utility IDS may serve, and most of these theories link IDS with facilitating infant language development either directly or indirectly.

IDS is characterized by a higher pitch than ADS and greater pitch variation (Fernald et al., 1989; McRoberts & Best, 1997). IDS has longer vowels and pauses than ADS (Bernstein Ratner, 1986; Fernald et al., 1989), and a slower rate of speech overall (Fernald & Simon, 1984). It has been suggested that these prosodic characteristics of IDS help to engage and maintain infants' attention during mother-child interactions (Fernald & Simon, 1984). Indeed, infants from one month to at least seven months prefer to listen to this type of exaggerated prosody, with the highest preference at four months (Cooper & Aslin, 1990; Fernald, 1985; Werker & McLeod, 1989), though this preference decreases at nine months and older (Newman & Hussain, 2006). It has also been theorized that the prosody of IDS communicates exaggerated positive affect to infants

(Fernald & Simon, 1984). In fact, a cyclical effect has been hypothesized, such that infants smile and engage more when listening to IDS as compared to ADS, which causes mothers to enjoy the interaction more, and therefore the interaction lasts longer than it otherwise might have (Werker & McLeod, 1989). Maintaining attention and creating an affective bond between mother and infant could both indirectly affect infant language development by encouraging infants to focus on the language input.

However, some of the syntactic, lexical, and prosodic characteristics of IDS have led to theories linking IDS to infant language development more directly. For example, durational (Bernstein Ratner, 1986) and prosodic cues (Kemler Nelson, Hirsh-Pasek, Jusczyk, & Cassidy, 1989) in IDS may help to mark word or clause boundaries in the incoming signal. It appears that infants use these cues to group syntactically-related words together (Mandel, Kemler Nelson, & Jusczyk, 1996; Mandel, Jusczyk, & Kemler Nelson, 1994). In fact, the prosody in IDS may help infants organize and remember phonetic information (Mandel et al., 1994) and encode word order information (Mandel et al., 1996). IDS also appears to facilitate segmentation of the speech stream by giving infants better access to the statistical structure of speech, such that infants can more consistently segment whole words as compared to part-words or syllables than they can from ADS (Thiessen, Hill, & Saffran, 2005). Additionally, in English, object names tend to occur at the ends of utterances, and display highest amplitude as well as an exaggerated pitch contour (Fernald & Mazzie, 1991; Messer, 1981), which may help to facilitate lexical learning. It may also be the case that other aspects of IDS, such as its acoustic-phonetic characteristics, play a role in infant language development.

### *Acoustic-Phonetic Characteristics of IDS*

It is generally acknowledged that ADS is characterized by a “messy” acoustic signal. Vowels are often poorly differentiated and coarticulation effects blur boundaries both between phoneme classes and words. However, there is evidence that this acoustic-phonetic information is clarified in IDS as compared to ADS.

For example, vowel length before word-final voiced consonants is generally longer than vowel length before voiceless consonants in ADS. However, this lengthening was found to be nearly doubled in mothers’ speech to their infants who were learning their first words (Bernstein Ratner & Luberoff, 1984), making the distinction between syllable-final voiced and voiceless sounds even greater. This distinction decreased as children began to acquire more language (Bernstein Ratner & Luberoff, 1984), implying that mothers adjust their speech according to the language needs of their children. Additionally, phonological rules that distort the intended signal, such as dental deletion (e.g., [want ɪt] → [wanɪt]) and initial [ð] deletion (e.g., [bɛnd ðɛm] → [bɛndɛm]) have been shown to occur less often in IDS than ADS (Bernstein Ratner, 1984a). This theoretically helps infants to segment the speech stream by keeping forms of the same word more consistent. However, avoidance of phonological rules was not seen in speech directed to children between two and four years of age (Shockey & Bond, 1980), again implying that speech input changes as children’s language abilities develop.

An early study of vowel characteristics of IDS showed that mothers speaking to their infants who were at the one-word stage of language development produced vowels with significantly less overlap among phonemic categories as compared to the vowels they produced when speaking to an adult (Bernstein Ratner, 1984b). In other words, the mothers produced the vowels in such a way that they would be less likely to be confused for one another when speaking to their infants. More recent studies have supported this finding, and extended the age range for which this clarification occurs down to 6-8 months (Liu, Kuhl, & Tsao, 2003; Liu, Tsao, & Kuhl, 2009). It has also been shown that the clarification that occurs is language-specific, meaning that the changes mothers make are relevant to the specific distinctions in their language rather than a general enlargement of the vowel space (Werker et al., 2007). However, a study that examined speech to infants over the first 6 months of life found that the vowel space of mothers speaking to their infants was smaller than when the mothers spoke to an adult (Kjellrun Englund & Behne, 2006). These results imply that IDS is not stable over the course of infancy, but may be tailored to the language abilities of the infants to which it is addressed.

Some studies have shown that the quality of input infants receive is correlated with their speech segmentation or perception abilities. For example, the size of the vowel space in IDS has been shown to be strongly correlated with speech perception test performance both at 6-10 and 10-12 months of age (Liu et al., 2003). Other studies have shown that speech input can also affect how infants learn the distinction between difficult phonemic contrasts, such as vowels that differ only in the tense/lax dimension, or consonants that differ only by presence or absence of voicing. These studies have indicated that infants who are exposed to a bimodal distribution of these contrasts, i.e., a majority of instances being at either end of the continuum in question, have enhanced abilities to detect that contrast in tests of perception as compared to infants who are exposed to unimodal distributions, i.e., a majority of instances being in the middle of the continuum in question (Maye, Weiss, & Aslin, 2008; Maye, Werker, & Gerken, 2002). This supports the theory that IDS (particularly maternal speech input, as a majority of studies have been conducted using mother-child speech) can serve to help or hinder speech discrimination and therefore language learning depending on how well mothers differentiate difficult contrasts. In fact, a computer model given input from English and Japanese IDS successfully learned vowel contrasts specific to each language, implying that the input from IDS contained sufficient distributional properties for the contrasts to be differentiated (Vallabha, McClelland, Pons, Werker, & Amano, 2007).

### *Voice Onset Time*

Another important phonetic distinction is voicing differences in word-initial stop consonants. In English, stop consonants can be divided into three pairs that are distinguished by place of articulation: bilabial ([b] and [p]), alveolar ([d] and [t]), and velar ([g] and [k]). Within each pair, the phonemes are perceptually distinguished by the amount of time that elapses between the release of the air built up behind the articulators (burst) and the onset of the vibration of the vocal folds. The elapsed time between these two events is known as voice onset time (VOT). Acoustically, these events correspond to an abrupt change in amplitude at the release of the burst of air and a low frequency periodic signal at the onset of voicing.

The perception of VOT in English is categorical, such that stops with shorter VOTs are perceived as voiced ([b, d, g]) and stops with longer VOTs are perceived as voiceless ([p, t, k]). However, the production of VOT occurs along a continuum, such that the onset of voicing can begin anywhere from 0 to 80 or more milliseconds after release of the burst of air. In fact, in some instances, voicing can begin even before the release of the stop. This is known as prevoicing and results in a negative VOT. Studies have shown that English-speaking listeners perceive stops with VOTs of less than approximately 30 milliseconds as voiced and stops with VOTs of more than 30 milliseconds as voiceless, though this value varies with place of articulation (Abramson & Lisker, 1965; Zlatin, 1974). In controlled studies, when a stop is presented with a VOT of around 30 milliseconds, listeners perceive the stop as voiced about half the time and as voiceless the other half of the time; thus a VOT of 30 milliseconds is perceptually unstable (Abramson & Lisker, 1965; Zlatin, 1974). In conversational speech, when the production of VOT (along with many other acoustic cues) is likely to be variable, listeners resolve this type of perceptual instability by using semantic and syntactic cues available in the context to decide what makes the most sense (Abada, Baum, & Titone, 2008; Sivonen, Maess, & Friederici, 2006). In other words, listeners employ top-down processing in order to decode words that are acoustically unclear.

However, infants do not have sufficient previous linguistic knowledge to employ top-down processing, and yet they perceive voicing distinctions categorically well before they produce their first words (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). The question then arises, how do infants make sense of this contrast? One possibility is that the input infants receive contains highly clarified VOT values such that the overlap between VOT for voiced and voiceless cognates is reduced.

### *Previous Voice Onset Time Studies*

Only a handful of studies have examined VOT in IDS. In one of the earliest of such studies, Moslin (1979) followed six American middle-class mothers over a six month period. At the onset of the study, two of the mothers had infants between 6 and 8 months of age who were preverbal, two mothers had toddlers between 15 and 16 months of age who were producing single words, and two mothers had children between 2;5 and 5;2 years of age who were producing multi-word utterances. Recordings of mother-child and adult-adult interactions were completed in the home of the subjects and all word-initial singleton stop consonants were included in the analysis. The mothers were given no specific instructions other than to play with the child, and though an experimenter was present for all of the tapings, she rarely spoke or interfered during the play sessions in order to maintain as natural an environment as possible. Moslin found that there was a significant reduction in overlap of target voiced and voiceless VOTs produced by the mothers of the middle age-range children. The reduction in overlap was fueled by the fact that VOTs for voiceless stops were significantly longer (13 milliseconds) in the mother-child interactions than in mother-adult interactions. VOTs for voiced stops were also significantly longer in IDS than ADS, though the difference was only two milliseconds, not a large enough increase to override the category separation that the increased voiceless VOTs created.  However, this reduction in overlap was not found in the productions of mothers of children in the other two age ranges. Moslin (later writing as Malsheen) proposed the independent phonetic clarification hypothesis (IPC) to explain this tendency for clarification during the one-word stage of language development. IPC states that mothers alter their speech to their children on the phonetic level according to the language abilities of their children (Malsheen, 1980; Moslin, 1979).

This hypothesis was supported by work by Sundberg and Lacerda (1999), who analyzed VOT in IDS of six Swedish mothers of 3-month-old infants. Languages like Swedish and Norwegian have similar VOT categories as English, so comparisons among these languages are appropriate (Rimol, Eichele, & Hugdahl, 2006; Stoel-Gammon, Williams, & Buder, 1994). The mothers in this study were recorded playing with their infants alone in a sound-treated studio. The mothers were given a set of preselected toys and were instructed to play with their infants as they would at home. Following the IDS session, an experimenter came into the room and spoke informally with the mother to obtain an ADS sample. All singleton, prevocalic initial and medial stop consonants were included in the analysis. At this young age, it was found that there was more variability of VOT in IDS than in ADS, which the researchers interpreted to mean that voiced and voiceless cognates were less differentiated in IDS than in ADS. Additionally, it was found that overall VOT duration was shorter in IDS than ADS for these mothers. This would be the expected result, as 3-month-old infants are certainly preverbal and therefore not linguistically sophisticated enough to warrant clarification in the mothers' speech. In a working paper, Sundberg (2001) reported preliminary data based on speech samples of three Swedish mothers of infants between 11 and 14 months of age using a similar protocol as Sundberg and Lucerda (1999, above), which showed an overall mean VOT that is significantly longer in IDS than ADS. This appears to support Moslin's IPC hypothesis, as this is the typical age range for children to acquire their first words. However, this study did not examine clarification of voiced and voiceless VOT values, and therefore no conclusions can be drawn about whether VOT was better differentiated in the sample.

In an earlier, smaller study, Baran and colleagues examined VOT in IDS of three American mothers of infants around one year of age (Baran, Zlatin Laufer, & Daniloff, 1977). Thirty minutes of naturalistic play sessions between mother and child were recorded in an infant laboratory. ADS samples were recorded during interviews between the experimenter and the participant mothers about the infants' overall development. All analyzable prevocalic stop consonants were included in the analysis. No difference was found in overall VOT length; however, when subsets of 20 exemplars of each stop were examined by place of articulation, a generally *shorter* VOT was found in IDS than ADS. This appears to contradict the IPC hypothesis, since these infants are presumably producing or about to produce their first words; however, this study also did not include an analysis on amount of overlap between voiced and voiceless VOT values, so no information is available with relation to differentiation of VOT.

In another VOT study, six Norwegian mothers were visited and recorded in their homes at 10 points during the first six months of their infants’ lives (Englund, 2005). Each mother was recorded speaking to her infant while she changed his or her diaper in an effort to keep the context of the conversation consistent across visits and participants. Following the diaper change, an experimenter spoke with the mother and recorded an ADS sample. In each sample, the mother was asked if she remembered any of the words she used when speaking to her infant in an attempt to elicit similar words in the ADS sample. However, the mothers only remembered a mean of one word between IDS and ADS samples. All word-initial stop consonants were included in the analysis. In accordance with the IPC hypothesis, VOT was stable across the first six months of life. However, VOT for each stop consonant except [p] was significantly *longer* in IDS than ADS. But again, this study provided no information on overlap and therefore it is difficult to say whether these results support the IPC hypothesis.

Prior studies have asked a good question -- is VOT clarified in IDS and if so, when. However, analyses have not really permitted an evaluation of clarification inasmuch as no overlap computations were done. Therefore, such an analysis appears warranted.

### *Present Study*

To further test the IPC hypothesis, the present study compares VOT in speech addressed to 7.5-month-old infants and 11-month-old infants to VOT in the ADS of the same speakers. This study examines the speech of fifteen mothers who have been recorded speaking to their infants at 7.5 months of age and again at 11 months of age and compare it to the speech of these same mothers when speaking with an unfamiliar adult. Typically, infants' cognition and language abilities mature quite a bit between these two ages. They develop the ability to engage in joint attention and begin to attain a means-end concept. Additionally, at 7.5 months, infants typically engage in canonical babbling, but by 11 months have begun to enter the jargon stage, incorporating adult-like intonational patterns, and some infants have begun to produce their first words by 11 months.

The sample in the current study includes more than twice as many mothers than have previously been examined in a single study. Additionally, only matched tokens, that is, identical words that the mother has spoken in both IDS and ADS conditions, are analyzed. This ensures that the two samples are as comparable as possible because vowel characteristics can affect VOT, and is a step that previous VOT in IDS studies have overlooked.

### *Hypotheses*

The first hypothesis of the present study is that overall duration of VOT for voiced and voiceless targets will be greater in speech directed to 11-month-old infants as compared to speech directed to 7.5-month-old infants. If mothers are adjusting their speech according to the language abilities of their infants, and if infants are making notable language gains between 7.5 and 11 months, then changes in VOT production should be evident between these two ages. This duration analysis will also allow the results of this study to be directly compared to previous studies.

Next, it is predicted that voicing of stop consonants in speech directed to 11-month-old infants will be better clarified than in speech in the other two conditions, as measured by $d_{(a)}$. $d_{(a)}$ is a calculation that determines how distinguishable the means of two distributions are, in this case the means of the distributions of VOTs of voiced and voiceless sounds. This would again show that mothers are changing their stop consonant productions to be more easily distinguishable based on the age and/or stage of language development of their infants, as 11-month-old infants are presumably more linguistically sophisticated than they were at 7.5 months.

A third hypothesis of the present study is that the clarification of voiced and voiceless VOTs will be directly correlated with a measure of the infants’ receptive vocabulary, based on parent report on the MacArthur-Bates Communicative

Development Inventories (MCDI). In other words, as the infants' receptive vocabularies increase, voiced and voiceless VOTs will become better clarified. This would support Moslin's IPC hypothesis as it would suggest that mothers are varying their VOT productions based on how much receptive language they believe their infants have.

## ***Methods***

### *Participants*

Participants were 15 mother-infant dyads (nine male infants and six female infants) who were part of a larger longitudinal study at the University of Maryland (NSF grant BCS 0745412, co-PIs Rochelle Newman and Nan Bernstein Ratner). All infants were born within three weeks of their due dates, are learning English as their first language, have native-American English-speaking mothers, and have no previous diagnosis of developmental problems. Data were collected at the first and third visits of the larger longitudinal study when the infants were 7.5 and 11 months old, respectively. Ages at 7.5 month visit ranged from 0;6.28 to 0;8.1 (*M* = 0;7.11), and ages at 11 month visit ranged from 0;10.8 to 0;11.8 (*M* = 0;10.28). Before the first visit, mothers signed consent forms, which are kept on file in the Infant Language Development and Perception Lab at the University of Maryland.

### *IDS and ADS Speech Samples*

Speech samples were collected from mother-child and mother-experimenter interactions in a sound-treated room at the University of Maryland. Mothers were provided with a pre-selected set of toys to facilitate similar topics of conversation across interactions, and were instructed to play with their infants as they would at home. Toys included small stuffed animals, such as a bee, a deer, and a kangaroo; play food items, including plates and utensils; a doll; and storybooks. Mother-child play sessions lasted from 10 to 15 minutes. Immediately following the play session, an experimenter spoke with the mother about their child's toy preferences in an attempt to elicit many of the same words used in the play session. Mothers were not told that their speech would be analyzed to ensure the most natural speech possible within the laboratory setting. Mothers wore an Audio-Technica ATR-35S lavalier microphone on their clothing throughout both interactions, and the speech samples were recorded as uncompressed WAV files using a Marantz PMD660 Professional Portable Digital Recorder at a sampling rate of 44.1 kHz.

### *Receptive Language Scores*

Before each play session, mothers were asked to fill out the MCDI form, indicating which words they believed their infants could understand at the time of testing. Raw scores were calculated by adding up the number of items checked off on the list.

### *Data Selection Procedure*

The WAV files of the mother-infant and mother-experimenter interactions were uploaded to a PC. Each raw file contained both the mother-child play session and the interview with the experimenter. Audacity, a free sound recording and editing program (http://audacity.sourceforge.net/), was used to split each file into two files: one containing only the mother-child play session and the other containing only the interview. Each sound file was orthographically transcribed into CHAT using the Computerized Language Analysis (CLAN) program developed by the CHILDES project (MacWhinney, 2009). CLAN allows portions of the sound file to be linked to each line of the transcript using sound "bullets." Each bullet can then be played back individually for more accurate transcription.

A frequency count of all words spoken by each mother in each transcript was run using the CLAN command FREQ (freq +t*MOT). This generates a list of all of the words spoken by the mother in each transcript and how many times they were used. The experimenter first identified all possible tokens from each of the child play sessions for the IDS targets. Possible tokens included any content word with an initial singleton (i.e., no consonant blends) stop consonant (/g, k, d, t, b, p/). Function words were excluded from analysis, as they are more likely to be affected by phonological rules, such as geminate reduction. The experimenter then compared these tokens with the FREQ outputs from the interview condition to obtain a list of tokens that occurred both in IDS and ADS. The CLAN command KWAL (kwal +t*MOT +s[token]) was then used to identify in which lines of the transcripts the tokens occurred. The experimenter confirmed the intended addressee upon listening to the sound bullets for each line containing a target. If a target word was addressed to the infant within an interview session, that token was considered IDS and was analyzed with the other IDS tokens from the mother-child play sessions and vice versa.

Matched tokens for each mother were recorded in a separate Excel workbook, each with four spreadsheets: tokens addressed to infants at 7.5 months of age, tokens addressed to an adult that match the tokens addressed to infants at 7.5 months, tokens addressed to the infants at 11 months of age, and tokens addressed to an adult that match the tokens addressed to the infants at 11 months. Each token was tabulated for the following variables: word, file from which the token was taken, line on which the token appears in the file, target phoneme for analysis, whether the target is voiced or voiceless, and resulting VOT for the initial stop consonant in seconds.

### *Acoustic Analysis*

Each target was acoustically analyzed using PRAAT (Boersma & Weenink, 2009). CLAN includes a utility that exports a chosen sound bullet directly to PRAAT. Using the spectrogram and acoustic signal, the target word was identified. VOT measurements were taken from the release of the stop consonant to the onset of low frequency periodic signal corresponding to the voicing of the following vowel. Judgments were made by visual inspection of the waveform in conjunction with the spectrogram. On rare occasions, other cues were used to help determine the end point of the measurement (i.e., voicing onset). These cues were the formant and pitch displays automatically generated by PRAAT. These parameters can theoretically only be calculated in the presence of a periodic signal and thus were used to help guide the experimenter in any case where voicing onset could not be determined by visual inspection of the waveform or spectrogram alone. Figure 1 shows how VOT was measured using the PRAAT display. VOTs were recorded in the spreadsheets to the nearest millisecond. Tokens that occurred in the presence of noise (e.g., rustling of toys or overlap with another speaker), were sung, or had no observable release were excluded from analysis. An example of a token with no observable release is shown in Figure 2. Additionally, prevoiced tokens were not included in the present analysis because it is very difficult to reliably distinguish prevoicing from voicing relating to prior words in running speech.

### *Reliability*

To estimate interrater reliability, two mothers were randomly chosen and VOT for all of the IDS and ADS tokens for those mothers were re-measured by another researcher, which represents just over 10 percent of the total sample. The correlation (Pearson's *r*) between pairs of experimenter measurements was 0.85, and the mean difference between measurements was 0.0016, or 1.6 milliseconds. Figure 3, below, shows a Tukey Mean-Difference Plot, where the *x*-axis is the mean of the two measurements and the *y*-axis is the difference between the measurements. The plot shows that the difference between the two measurements did not differ systematically with the size of the measurements.

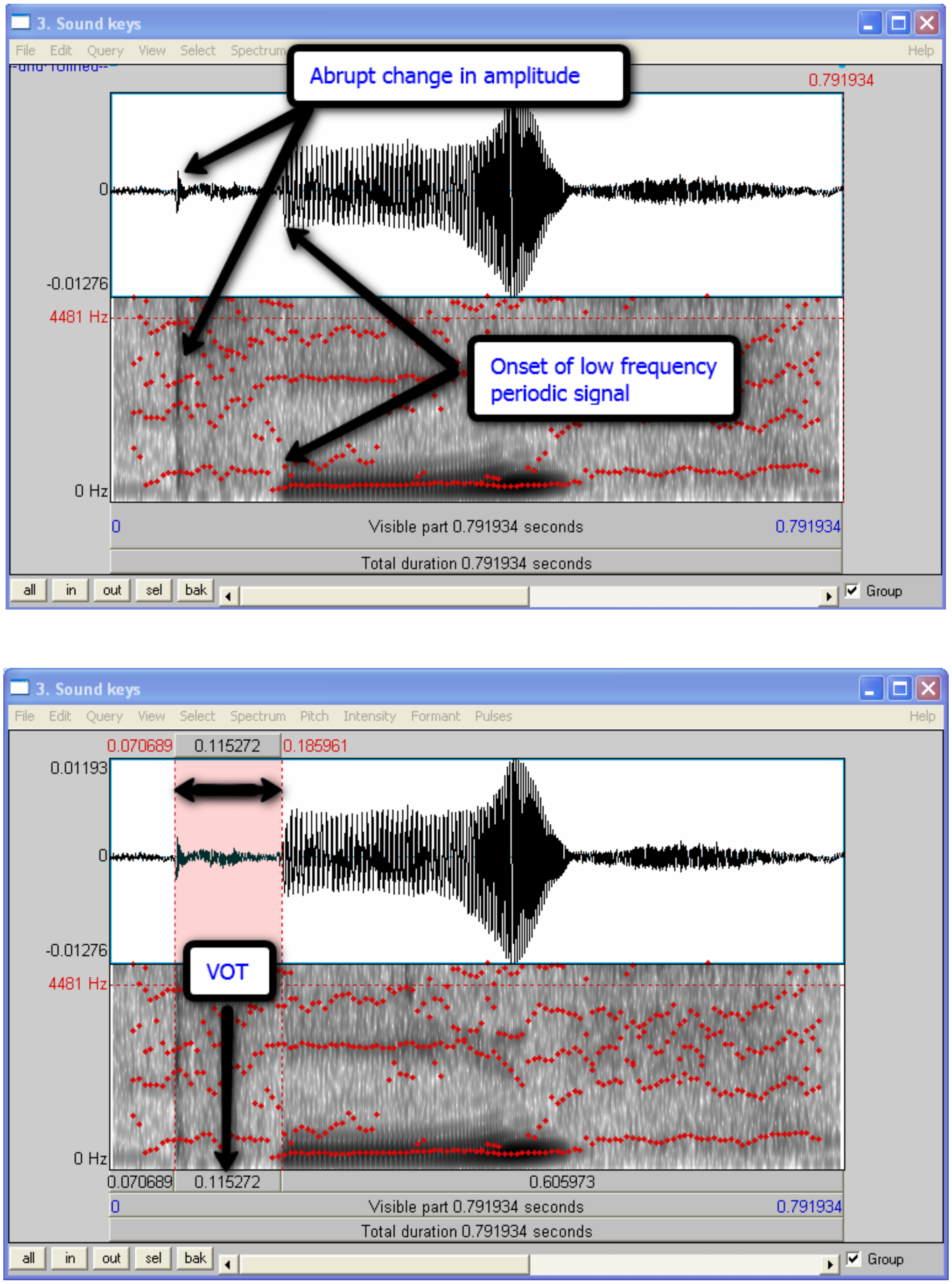

*Figure 1.* Measuring VOT for the word "keys" using the PRAAT display

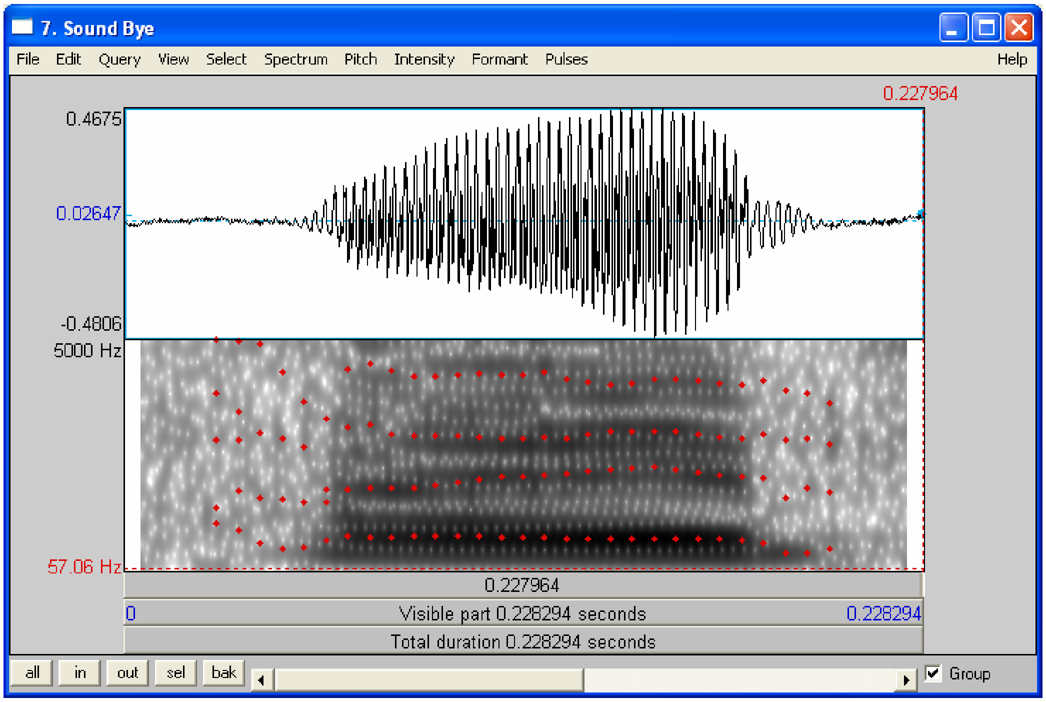

*Figure 2.* A token ("bye") excluded from analysis due to no observable release

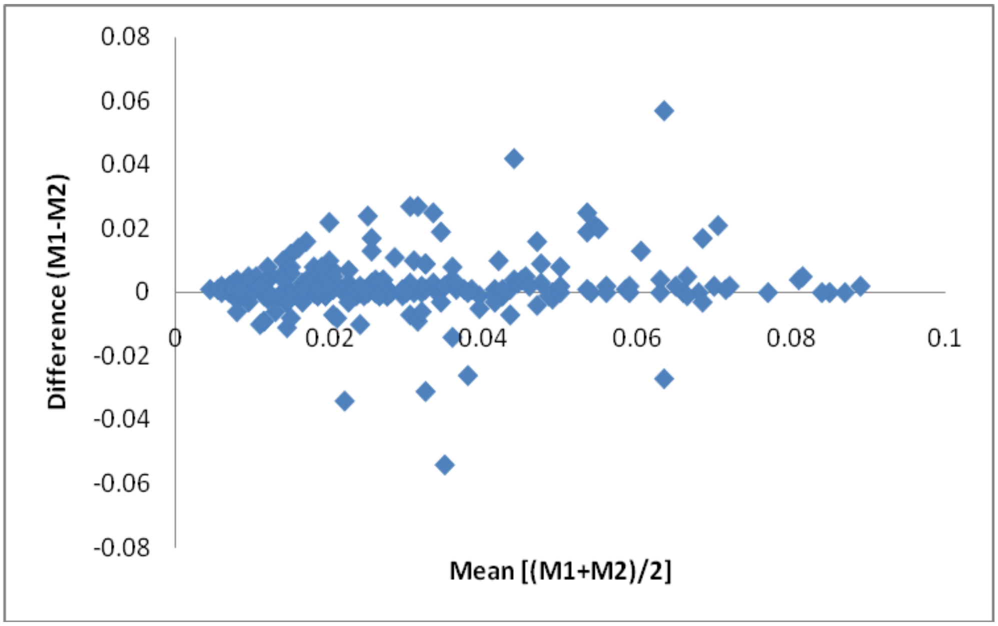

*Figure 3.* Tukey Mean-Difference Plot

### *Data Analysis*

A nonparametric test (Wilcoxon Signed-Rank Test, corrected for tied ranks, and converted to yield a Z score) was used to compare IDS and ADS because the assumption of equal variances was violated for some of the comparisons. All comparisons were computed using NCSS 2000 (Hintze, 1999), were two-tailed, and had an alpha level set at .05.

To determine whether overall VOT duration differed between ADS and IDS at 7.5 and 11 months, a mean for voiced and voiceless items was calculated for each mother in each condition. Comparisons were made between the ADS and IDS means for voiced and voiceless stops separately at each age (7.5 and 11 months) as well as with combined data (all IDS data compared to all ADS data).

Amount of clarification of VOT was determined by calculating $d_{(a)}$ for each mother at each age. Borrowed from signal detection theory, $d_{(a)}$ is a variation of $d'$, a calculation used to determine how easily the means of two distributions can be distinguished, with a higher value representing distributions that are more easily discriminated (Macmillan & Creelman, 2005). $d'$ is a measure of perceptual distance along a continuum in units of standard deviations, the continuum in this case being VOT. $d_{(a)}$ is used in place of $d'$ when variances of the two distributions in question are unequal. Because it is likely that the variance of the voiceless VOT distribution will be greater than the variance of the voiced VOT distribution, $d_{(a)}$ was chosen for the present study. When variances are equal, $d_{(a)}$ reduces to $d'$ (Swets & Pickett, 1982). $d_{(a)}$ is calculated by taking the difference in the means of the two categories times the square root of 2, divided by the square root of the sum of the variances. A $d_{(a)}$ value was calculated from VOT values for each mother in each condition and then ADS and IDS values were compared at each age as well as with combined data.

Items were then broken down to be examined according to place of articulation (bilabial, alveolar, or velar). A $d_{(a)}$ value was calculated for voiced-voiceless pairs ([b,p], [d,t], or [g,k]) for any mother who had at least six examples of each at both 7.5 and 11 months to determine whether results varied based on place of articulation. ADS and IDS values were compared at each age as well as with combined data. However, six pairs constitutes a relatively small sample size to estimate the mean and standard deviation of a population; therefore, these results should be reviewed cautiously.

Finally, to determine whether amount of overlap between voiced and voiceless VOTs varies with mothers' perceptions of infants' receptive vocabularies, Pearson's $r$ was calculated between each mother's $d_{(a)}$ value and MCDI score at each age as well as with combined data.

## ***Results***

### *VOT Duration*

At 7.5 months, there was no difference in VOT duration between IDS and ADS voiceless consonants, with an IDS mean of 0.0672 and ADS mean of 0.0658 ($Z$-value = 0.057, $p > 0.05$). The same is true for voiceless measures at 11 months (IDS mean = 0.0620, ADS mean = 0.0672; $Z$-value = 1.65, $p > 0.05$) and all data combined (IDS mean = 0.0646, ADS mean = 0.0665; $Z$-value = 1.08, $p > 0.05$).

However, with voiced measures, there was a significant difference between IDS and ADS conditions at 7.5 months (IDS mean = 0.0211, ADS mean = 0.0166; *Z*-value = 3.35, $p < 0.05$), 11 months (IDS mean = 0.0208, ADS mean = 0.0168; *Z*-value = 1.99, $p < 0.05$), and with all data combined (IDS mean = 0.0209, ADS mean = 0.0167; *Z*-value = 3.01, $p < 0.05$). In all cases, the IDS mean was significantly longer than the ADS mean. Means, standard deviations, Z-values, and exact *p* values for VOT duration comparisons are shown in Table 1.

| | IDS | | ADS | | | |
|---|---|---|---|---|---|---|
| Age | *M* | (*SD*) | *M* | (*SD*) | Wilcoxon *Z* | *p* Value |
| 7.5 mos | | | | | | |
| Voiced | 0.0211 | (0.004) | 0.0166 | (0.003) | 3.35 | 0.0008* |
| Voiceless | 0.0672 | (0.014) | 0.0658 | (0.016) | 0.057 | 0.95 |
| 11 mos | | | | | | |
| Voiced | 0.0208 | (0.005) | 0.0168 | (0.004) | 1.99 | 0.047* |
| Voiceless | 0.0620 | (0.016) | 0.0672 | (0.016) | 1.65 | 0.099 |
| Combined data | | | | | | |
| Voiced | 0.0209 | (0.004) | 0.0167 | (0.003) | 3.01 | 0.003* |
| Voiceless | 0.0646 | (0.013) | 0.0665 | (0.012) | 1.08 | 0.28 |

*Table 1.* VOT duration results, $^*p < 0.05$

### $d_{(a)}$ *Results*

The comparison of $d_{(a)}$ values between IDS and ADS conditions at 7.5 months were significant, with an IDS mean of 2.03 and an ADS mean of 3.09 (*Z*-value = 2.61, $p < 0.05$). Results were significant at 11 months (IDS mean = 2.09, ADS mean = 2.87; *Z*-value = 2.56, $p < 0.05$) as well as with all data combined (IDS mean = 2.06, ADS mean = 2.98; *Z*-value = 3.63, $p < 0.05$). The ADS $d_{(a)}$ mean was consistently significantly higher than the IDS $d_{(a)}$ mean, as shown in Figure 4. Means, standard deviations, *Z*-values, and exact *p* values for $d_{(a)}$ comparisons are shown in Table 2. The distribution of IDS and ADS VOT values at 7.5 months and 11 months are displayed in Figures 5 and 6, respectively, to illustrate amount of overlap between voiced and voiceless VOT.

| | IDS | | ADS | | | |
|---|---|---|---|---|---|---|
| Age | *M* | (*SD*) | *M* | (*SD*) | Wilcoxon *Z* | *p* Value |
| 7.5 mos | 2.03 | (0.30) | 3.09 | (1.76) | 2.61 | 0.009* |
| 11 mos | 2.09 | (0.74) | 2.87 | (0.55) | 2.56 | 0.011* |
| Combined data | 2.06 | (0.56) | 2.98 | (1.28) | 3.63 | 0.0003* |

*Table 2.* $d_{(a)}$ results, * $p < 0.05$

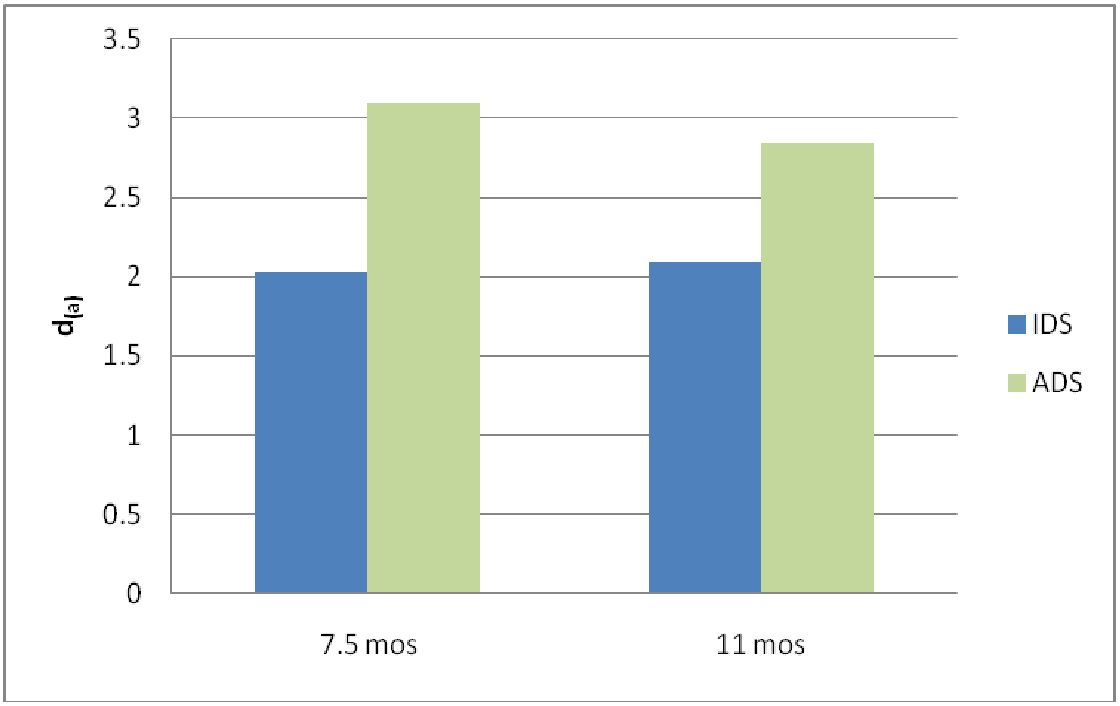

*Figure 4.* Average $d_{(a)}$ values for IDS and ADS at 7.5 and 11 months

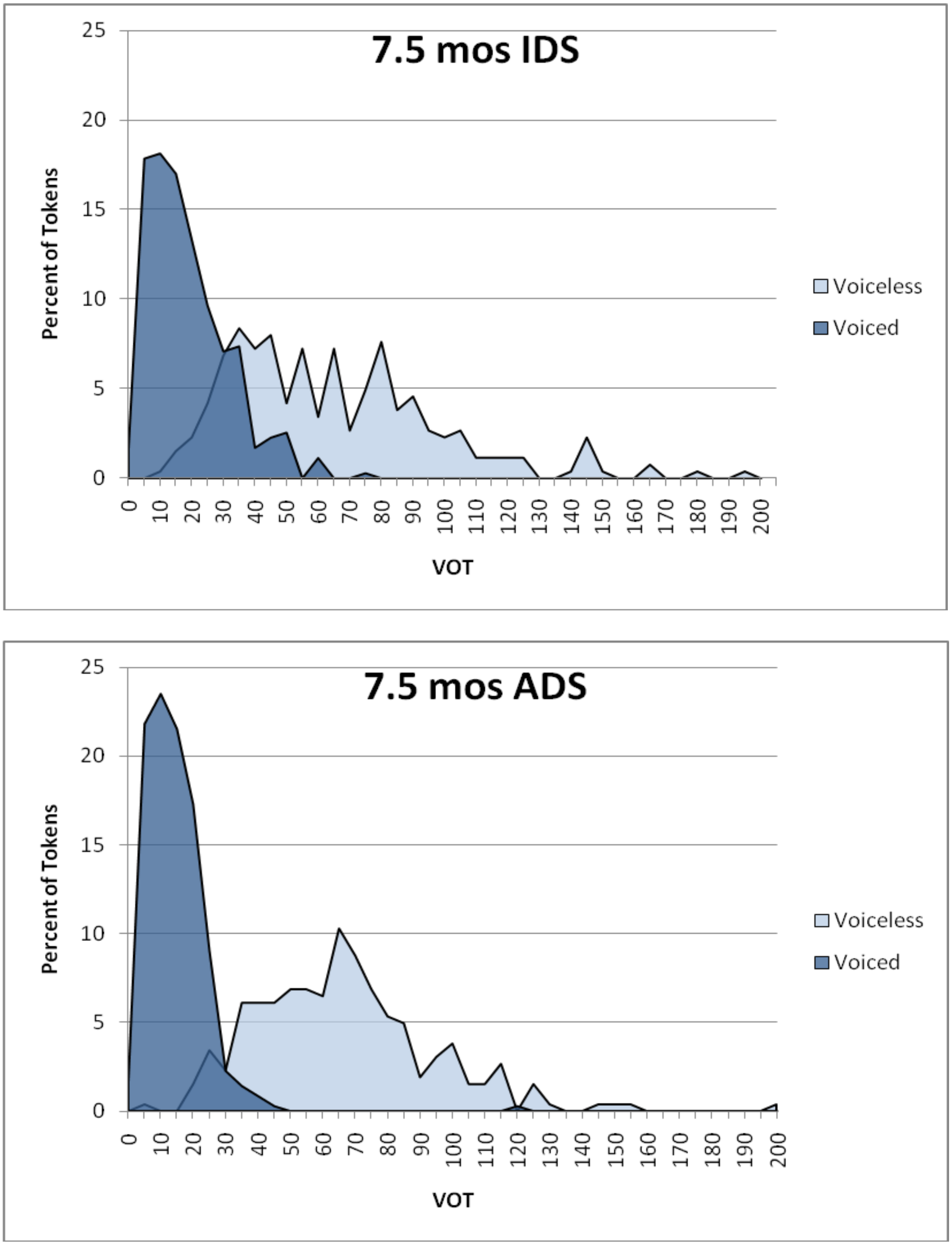

*Figure 5.* Distribution of VOT values (in milliseconds) at 7.5 months

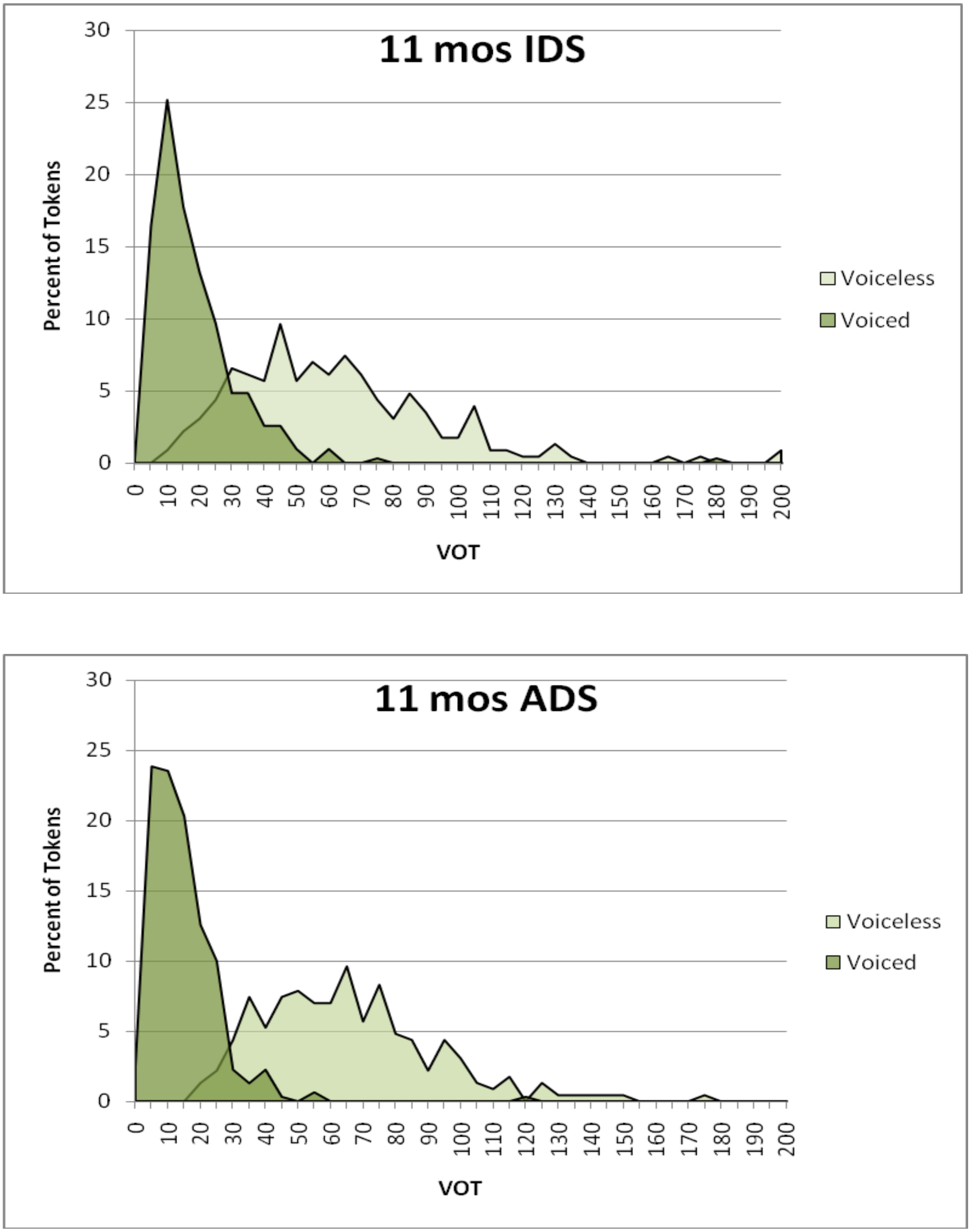

*Figure 6.* Distribution of VOT values (in milliseconds) at 11 months

### $d_{(a)}$ *Results by Cognate Pairs*

Comparisons of $d_{(a)}$ values for [p, b] phonemes were not significant at 7.5 months, 11 months, or with all data combined. Nor were they significant for [k, g] phonemes at 7.5 months, 11 months, or with all data combined. There was only one mother with enough examples of [t, d] pairs to compute $d_{(a)}$; therefore, no analysis of these phonemes could be carried out. Average $d_{(a)}$ values for bilabial and velar phoneme pairs are depicted in Figures 7 and 8, respectively. Again, it is important to review these results cautiously, as each $d_{(a)}$ value was computed with as few as six pairs of items and sample sizes within each comparison are small. Sample sizes, means, standard deviations, *Z*-values, and *p* values are listed in Table 3.

| | IDS | | | ADS | | | | |
|---|---|---|---|---|---|---|---|---|
| Age | *n* | *M* | (*SD*) | *n* | *M* | (*SD*) | Wilcoxon *Z* | *p* Value |
| [p, b] | | | | | | | | |
| 7.5 mos | 6 | 2.56 | (0.77) | 6 | 2.67 | (0.56) | 0.10 | 0.92 |
| 11 mos | 6 | 2.12 | (0.44) | 6 | 2.87 | (0.59) | 1.78 | 0.07 |
| Combined | 12 | 2.35 | (0.65) | 12 | 2.77 | (0.56) | 1.73 | 0.08 |
| [k, g] | | | | | | | | |
| 7.5 mos | 3 | 1.99 | (1.11) | 3 | 2.76 | (0.27) | 1.07 | 0.29 |
| 11 mos | 3 | 2.45 | (0.91) | 3 | 2.64 | (0.42) | 0.00 | 1.00 |
| Combined | 6 | 2.22 | (0.94) | 6 | 2.70 | (0.32) | 0.94 | 0.35 |

*Table 3.* $d_{(a)}$ results by cognate pairs

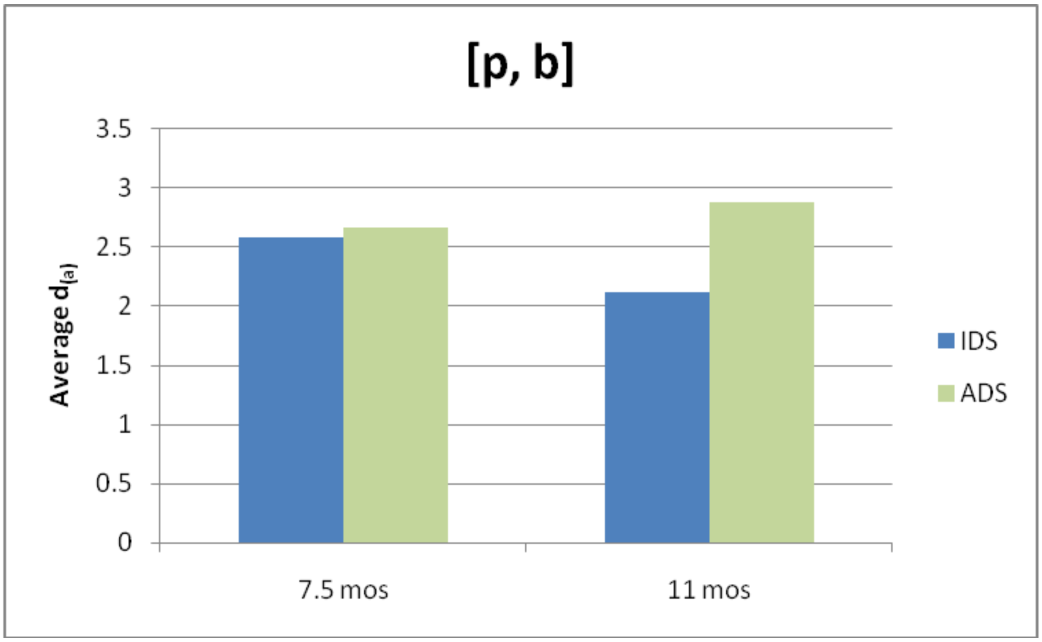

*Figure 7.* Average $d_{(a)}$ values for [p, b] in IDS and ADS at 7.5 and 11 months

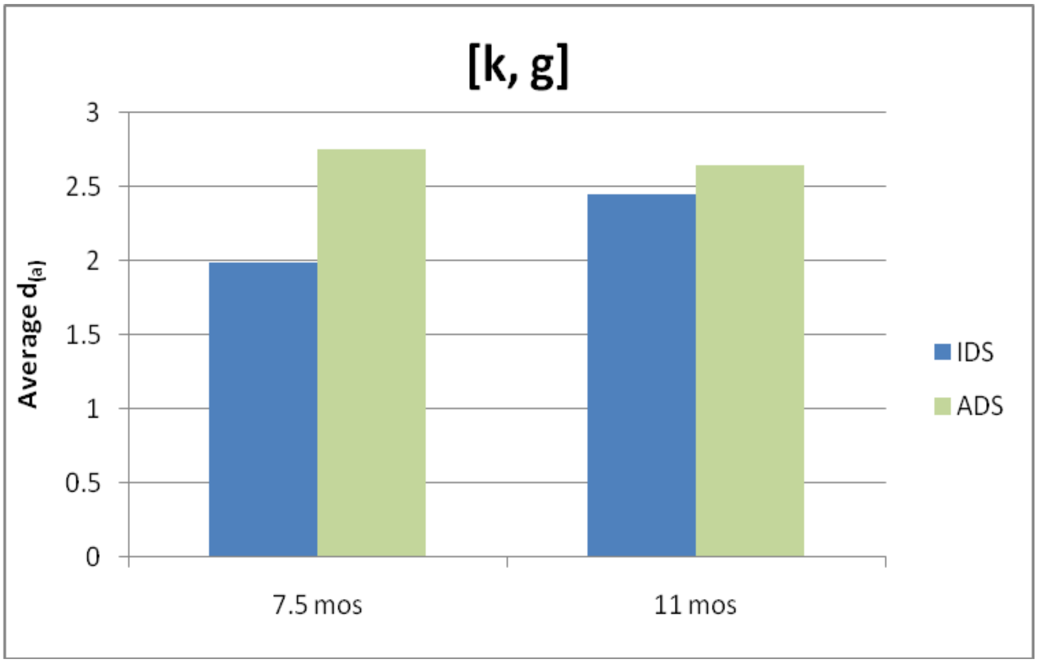

*Figure 8.* Average $d_{(a)}$ values for [k, g] in IDS and ADS at 7.5 and 11 months

*VOT Clarification and Infant Receptive Vocabulary*

MCDI scores for each participant at both ages are listed in Table 4. There was no correlation between IDS $d_{(a)}$ values and MCDI scores at 7.5 months ($r = 0.023$), 11 months ($r = 0.088$), or with all data combined ($r = 0.054$). Figures 9 and 10 show scatter plots and regression lines at 7.5 months and 11 months, respectively.

| | MCDI Scores | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Particpant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 7.5 mos | 0 | 2 | 18 | 0 | 16 | 1 | 6 | 0 | 9 | 2 | 8 | - | 4 | 12 | 0 |
| 11 mos | 27 | 41 | 36 | 10 | 55 | 2 | 15 | 32 | 51 | 4 | 44 | 72 | 13 | 76 | 2 |

*Table 4.* MCDI scores at 7.5 and 11 months

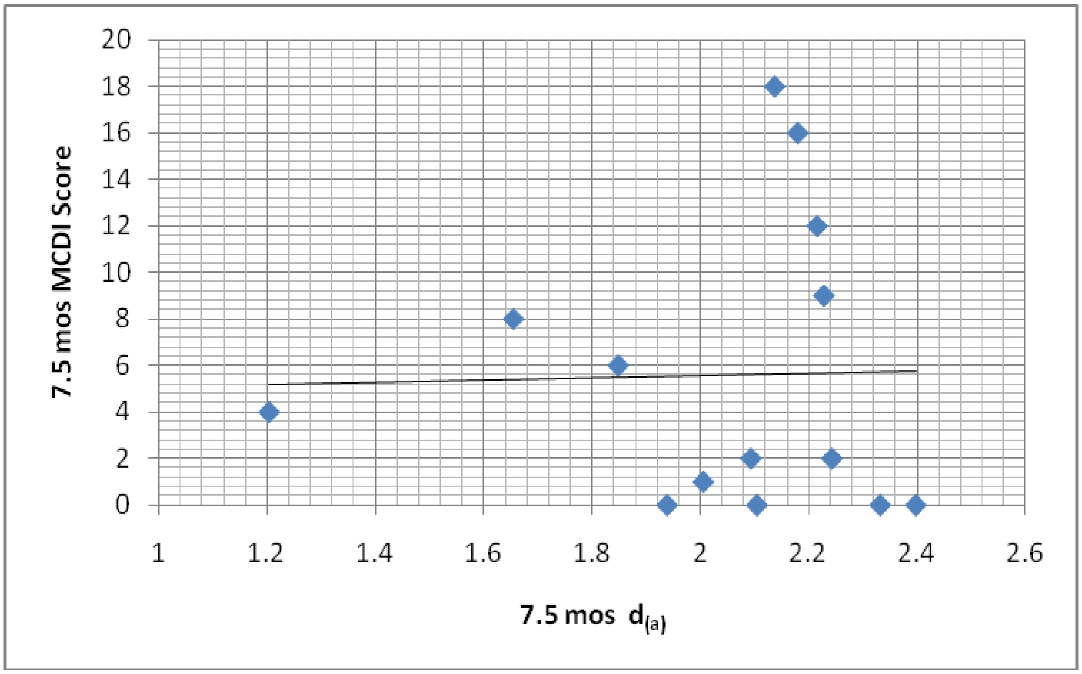

*Figure 9.* Scatter plot of MCDI vs. $d_{(a)}$, 7.5 months

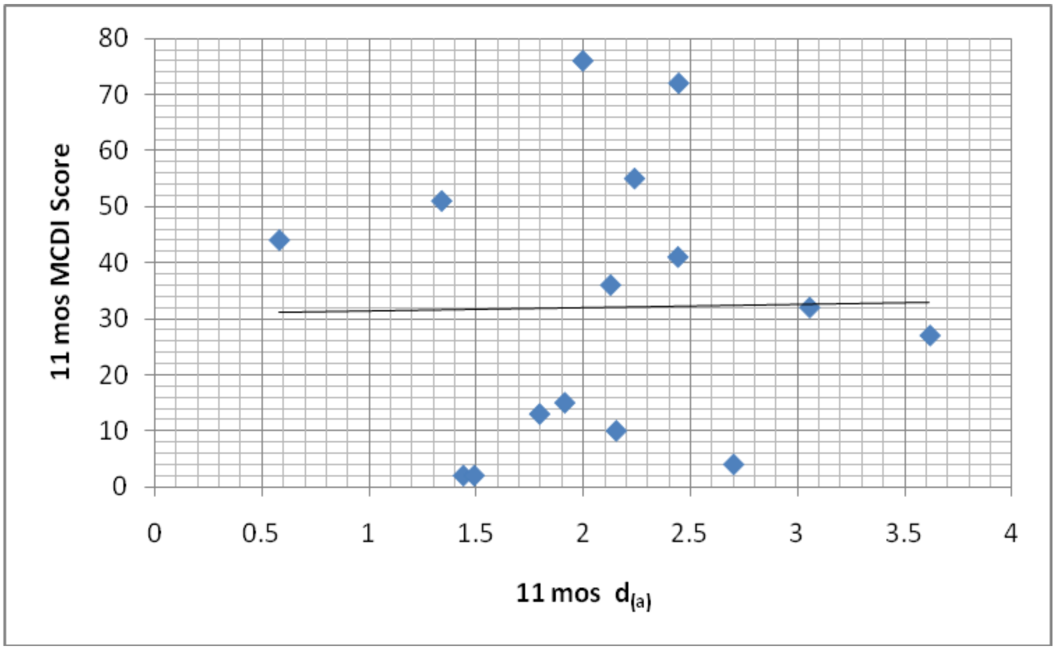

*Figure 10.* Scatter plot of MCDI vs. $d_{(a)}$, 11 months

## *Discussion*

The aim of this study was to examine whether mothers clarify their VOT productions when speaking to their infants as compared to when they speak to adults. This question was examined both by looking at VOT duration as well as discriminability, or how distinguishable the means of the two distributions are. Additionally, to further test Moslin's IPC hypothesis, a correlation analysis was computed to determine whether discriminability between voiced and voiceless VOTs varies with mothers' perceptions of infants' receptive vocabularies.

Interestingly, results showed that mothers' VOTs were more easily distinguishable in ADS than IDS, the opposite of the hypothesized result. By examining the results of the VOT duration analysis, it is clear that VOT in IDS is less distinguishable than in ADS due to significantly longer VOT for voiced items and no difference in VOT duration for voiceless items. This means that in IDS, voiced items begin to encroach upon the territory of voiceless items, causing there to be increased overlap between categories. However, it is also interesting to note that the average VOT duration for voiceless items in all conditions was around 65 milliseconds, but the range of VOT for voiceless items was quite large, with some values less than 10 milliseconds, well within the range expected for voiced stops.

Patterns of VOT clarity are the same at 7.5 and 11 months, suggesting that mothers don't adjust their speech to the infants between these two ages. Results also showed that there is no correlation between degree of VOT differentiation and mothers' perceptions of their infants' receptive vocabularies either at 7.5 months or 11 months. The $d_{(a)}$ analysis of individual phoneme pairs revealed no significant difference between clarification of [p, b] or [k, g], though these results are limited by both a small number of items with which to calculate $d_{(a)}$, as well as a small sample size of mothers included in the comparisons.

These results do not align perfectly with any of the previous VOT in IDS studies, in that no previous study found only a difference in duration of VOT for voiced stops. Two of the previous studies found consistently longer VOTs for both voiced and voiceless stops in speech addressed to infants from 0-6 months of age (Englund, 2005), and to infants producing their first words, around 15 months of age (Moslin, 1979). Two other studies found consistently shorter VOTs for both voiced and voiceless stops in speech addressed to infants at 3 months of age (Sundberg & Lacerda, 1999), and in a subset of data from speech addressed to infants at 12 months of age (Baran et al., 1977).

The $d_{(a)}$ results from the present study are the opposite of the clarification results from Moslin (1979). In her study, the overlap of the voiced and voiceless distributions, defined as the percentage of tokens from one category in the range of the other category (mean +/- 2 standard deviations), was significantly reduced in the IDS of the two mothers whose children were producing their first words. For the mothers of older and younger children, there was no difference in amount of category overlap between IDS and ADS. In none of the cases was there less overlap in ADS as compared to IDS, which would be comparable to the higher $d_{(a)}$ values found in ADS in the present study. Because no other previous studies looked at VOT clarification in any systematic way, it is difficult to compare the $d_{(a)}$ results to the results of any of the other studies.

The outcome of this study, that mothers are less clear in their VOT distinction when speaking to their infants at 7.5 and 11 months as compared to adults, is unexpected; however, there are several possible explanations for these seemingly counterintuitive results. First, it is possible these infants are too young or not linguistically sophisticated enough at these ages for mothers to begin clarifying their VOT values. Though many infants begin to produce their first words at around 12 months of age, none of the infants in this sample were at the one-word stage at the time of the 11 month samples. Bernstein Ratner (1984b) has suggested that examinations of changes in IDS are best conducted based on the language age of the infants rather than chronological age because mothers are likely to adjust their speech according to the infants' language abilities rather than purely based on age. In this study, MCDI scores were used to estimate language ability of the infants; however, these scores only reflect mothers' perceptions of receptive language. Perhaps mothers begin to clarify their consonants only after infants begin developing expressive language. Early maternal speech studies have shown that mothers tend to clarify their speech most when their infants are in the one-word stage of language development (Bernstein Ratner, 1984b; Bernstein Ratner & Luberoff, 1984). Similarly, clarification of VOT might only occur for specific words that mothers feel their infants are on the verge of producing or of which they are producing approximations. In this case, VOT clarification would not be the general rule, but would emerge only when mothers use those specific words. This has been shown with regard to prosody of IDS, where infants acquire specific words earlier when those words have been emphasized in IDS with greater intensity and duration (Vosoughi, Roy, Frank, & Roy, 2010). However, if this were the only VOT adjustment mothers make, VOT would be expected not to differ significantly between IDS and ADS, which does not explain the results of the current study, which showed a pattern opposite to that predicted.

Second, mothers occasionally whispered to their infants, but never whispered to the unfamiliar adults. Whispering can cause VOT of voiced stops to be lengthened. This could explain why VOT in IDS is longer than in ADS, which in turn can explain why VOT productions are less differentiated in IDS as compared to ADS. However, it is unclear if there were enough whispered tokens included in this study to explain such significant results, with whispered items making up 7.6 percent of the voiced tokens at 7.5 months and 4.2 percent of the voiced tokens at 11 months.

Finally, it may be that parents are trying to emphasize voiced stops by lengthening the VOT slightly, because they have an otherwise very short and perhaps easily missed acoustic signal. Additionally, the properties that distinguish stop consonants from other types of consonants (the temporary blockage of airflow due to complete closure at the point of articulation) can be obscured in running speech due to incomplete closure or articulatory approximations, particularly with voiced stops. Moslin (1979), who found that VOT of voiced stops were significantly longer in IDS by around 2 milliseconds (a very similar result to those of the current study), suggests that mothers may be enhancing the complete closure of the voiced stops or avoiding approximations to the point of articulatory contact (Moslin, 1979, p. 136). In other words, mothers may be adjusting their speech to enhance the manner of the stop consonant production. This would be a strategy for clarifying voiced stop consonants from other types of voiced consonants, such as fricatives, rather than a strategy meant to clarify voiced from voiceless stops.

To determine whether increasing the VOT of voiced stops is a general strategy for speech clarification, it is appropriate to examine the literature on clear speech. Clear speech is the type of speech produced when addressing listeners with a hearing loss or from a different language background, or speech produced to clarify a previously misperceived message. However, studies in this area suggest that the VOT of voiceless stops is lengthened during clear speech while the VOT of voiced stops remains stable (Krause & Braida, 2004; Smiljanic & Bradlow, 2008), which would help to decrease the amount of overlap between VOT of voiced and voiceless categories. This pattern was not observed in the current study.

Therefore, the pattern of initial stop consonant clarification in maternal speech appears to be uniquely different from other forms of clear speech, and it may be that it is different for a reason. It could be that emphasizing the presence of voiced stops or the production manner of the voiced stops is perceived as more beneficial to infants' language development at these ages than clarifying the productions of voiced and voiceless stops. Such an emphasis might be more beneficial for infants before they begin producing language, and increased differentiation between voicing categories might be more beneficial after infants develop an expressive vocabulary. Early on, mothers may be simply alerting their infants to the acoustic signal that marks voiced stop consonants; once they begin producing approximations of these stops, mothers may change their strategy and begin reducing the overlap between voicing categories to help fine-tune their infants' productions.

## *Limitations of the Current Study*

While the current study attempted to improve upon the reliability and validity of past VOT in IDS studies by using natural speech samples of mothers unaware of the aims of the study, matching words between IDS and ADS samples to ensure comparable phonetic environments, and examining degree of clarification of voicing categories as well as VOT duration, there were a number of limitations to the study.

First, because the participants in this study are part of a larger longitudinal study, the ages of the infants at each visit were predetermined. Therefore, it was not possible to group the infants based on language ability, which might have yielded different results. Additionally, for this project, it was not possible to obtain speech samples at ages above 11 months to determine whether VOT clarity in IDS emerges at a later age or as expressive language develops.

Additionally, speech samples were obtained in a laboratory setting with an experimenter present. It is unclear how much, if any, these results would differ if recordings could have been made in each mother's home or another natural setting.

The fact that for this study tokens were only analyzed in matched pairs between IDS and ADS conditions to keep phonetic environments balanced means that many of the tokens available in the sample were simply not included because there was no match in the corresponding transcript. While it is unlikely that the tokens that were included were biased in any systematic way, it is possible that analyzing all possible tokens could change the results. In the present study 2,310 tokens were analyzed, 1,155 IDS and 1,155 ADS, while the total number of stop consonants present in the transcripts were 3,543 IDS tokens and 2,942 ADS tokens, for a total of 6,485. If all possible tokens had been included, results may have come out differently.

Similarly, some tokens had to be excluded due to no observable burst because it is difficult to reliably measure VOT from these items. Words with initial [b] were most likely to be excluded for this reason, although [b] tokens were also the most plentiful across transcripts, even after exclusions. However, tokens with weak bursts were still present in the transcripts, which means they made up some of the input that the infants (and adults) heard. Not including them in the analysis could potentially skew the results, particularly if more of this type of token were present in one condition as compared to another.

There were also not enough examples of each phoneme pair per mother in the present study to conduct a robust $d_{(a)}$ analysis of the pairs individually. If there were, it may have been possible to see if VOT differences between IDS and ADS are driven by one phoneme pair more than another or if there is a general difference across places of articulation. It is also possible that the amount of overlap evident in the current analysis is actually caused by overlap of the measurements of voiced stops with the longest VOT (i.e., [g]) and voiceless stops with the shortest VOT (i.e., [p]). If this were the case, $d_{(a)}$ values for phoneme pairs in IDS would be expected to be quite high, indicating very little overlap. This was not the case in the current study, as the IDS $d_{(a)}$ values for the individual pairs were between 2.0 and 2.6, very similar to the average overall IDS $d_{(a)}$ values (see Figures 4, 7, and 8 for illustration). However, without a robust analysis of differentiation for each pair it is not possible to look at this question in any real depth.

## *Future Directions*

Future studies examining VOT in IDS should analyze degree of overlap between voiced and voiceless VOTs in addition to VOT duration differences, as well as match phonetic environments of items used in the analysis between IDS and ADS conditions, as were done in the current study.

To determine whether whispered tokens in IDS have an effect on VOT, specifically if they significantly lengthen the VOT of voiced stops only, future analyses can compare results with whispered tokens included to results after whispered tokens have been removed. Some might argue that whispered tokens should not be included in a VOT analysis at all; however, it cannot be denied that infants are receiving a fair number of whispered utterances as input, and therefore excluding these items would ignore an entire segment of input that infants receive.

Another dimension that would be interesting to include in future studies is an examination of other acoustic measures related to stop consonant production. For example, rapid spectral changes associated with first-formant (F1) transitions and the low frequency onset of F1 have been identified as markers of voiced stops (Lisker, 1975;

Stevens & Klatt, 1974). Therefore, to determine whether mothers are indeed emphasizing the acoustic signal associated with voiced stops, it would be interesting to see if they are extending F1 transitions in addition to increasing the VOT of voiced stops. An examination of burst amplitude (the release of the stop) in relation to vowel amplitude is another way to determine whether mothers are attempting to emphasize voiced stops generally. If the burst amplitude of voiced stops is relatively larger than the burst amplitude of voiceless stops, then the hypothesis that mothers are emphasizing voiced stops in particular would be upheld. However, this would be difficult to measure in natural conversation with a lapel microphone like the one used in the current study because participants are free to turn their heads while speaking, causing the distance from the mouth to the microphone to change. Additionally, an examination of prevoiced items may be beneficial because prevoicing would certainly add emphasis to voiced stops.

It would also be beneficial for future studies to obtain speech samples from mothers beginning when their infants are around 6 months and following them at regular intervals of perhaps 2 months until they reach about 2 years of age. Having so many samples covering a longer period of time would allow for the analysis of any possible changes in clarification strategies over time. Alternatively, participants could be recruited based on the language age of their infants, as suggested by Bernstein Ratner (1984b). Samples could be obtained when infants reached a receptive vocabulary of a certain number of words, again immediately after the infants produce their first words, and again when their expressive vocabularies reach a certain number. This type of analysis would be more to the point in determining whether mothers adjust their stop consonant productions based on their infants' stage of development.

# Acknowledgements

This work was supported by NSF grant BCS 0745412 to the University of Maryland, co-PIs Rochelle Newman and Nan Bernstein Ratner.

Special thanks to my advisory committee, Nan Bernstein Ratner, Rochelle Newman, and Monita Chatterjee for endless amounts of guidance, reassurance, and support; and to Kerry McColgan for a whole host of things, not the least of which has been moral support

Thank you to Mara Steinberg for taking the time to be trained and completing the reliability coding in such a timely manner.

Additional thanks to the hard workers of the Language Development Lab and Dr. Ratner’s lab for scheduling visits, collecting data, and helping with some transcription: Saher Ali, Megan Askew, Kathleen Calaro, Justine Dombroski, Lauren Evans, Andrea Farina, Lauren Fischer, Richard Garcia, Laura Horowitz, Megan Janssen, Erica Levinson, Rachel Lieberman, Jenesia McCammon, Eileen McLaughlin, Sarah Michael, Giovanna Morini, Vidda Moussavi, Courtenay O'Connor, Sabrina Panza, Amanda Pasquarella, Lauren Polovoy, Rachel Ports, Allie Rodriguez, Maria Rodriguez, Judith Segal, Mara Steinberg, Justine Taweel, Allison Temple, Dena Tran, Michelle Zobel, Julie Sampson, Michelle Cass, Michelle Keenan, Audry Singh, Sarah Steele, Jaclyn Woodyat, Jennifer Hammer, Debbie Martinez, and Lauren Berger.

# References

Abada, S., Baum, S., & Titone, D. (2008). The effects of contextual strength on phonetic identification in younger and older listeners. *Experimental Aging Research*, *34*(3), 232-250.

Abramson, A. S., & Lisker, L. (1965). Voice onset time in stop consonants: Acoustic analysis and synthesis. In *Proceedings of the 5th International Congress of Acoustics*. Liege, Belgium.

Baran, J. A., Zlatin Laufer, M., & Daniloff, R. (1977). Phonological contrastivity in conversation: A comparative study of voice onset time. *Journal of Phonetics*, *5*, 339-350.

Bernstein Ratner, N. (1984a). Phonological rule usage in mother-child speech. *Journal of Phonetics*, *12*(3), 245-254.

Bernstein Ratner, N. (1984b). Patterns of vowel modification in mother-child speech. *Journal of Child Language, 11*(3), 557-578.

Bernstein Ratner, N. (1986). Durational cues which mark clause boundaries in mother-child speech. *Journal of Phonetics*, *14*, 303-309.

Bernstein Ratner, N., & Luberoff, A. (1984). Cues to post-vocalic voicing in mother–child speech. *Journal of Phonetics*, *12*, 285-289.

Boersma, P., & Weenink, D. (2009). *Praat*. Amsterdam, The Netherlands: University of Amsterdam. Retrieved from www.praat.org.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science (New York, N.Y.)*, *296*(5572), 1435.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*(5), 1584-1595.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science (New York, N.Y.)*, *171*(968), 303-306.

Englund, K. (2005). Voice onset time in infant directed speech over the first six months. *First Language*, *25*(2), 219-234.

Englund, K., & Behne, D. (2006). Changes in infant directed speech in the first six months. *Infant and Child Development*, *15*(2), 139-160.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*(2), 181-195.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, *27*(2), 209-221.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104-113.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*(03), 477.

Hintze, J. (1999). *NCSS 2000*. Kaysville, UT.

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, *16*(1), 55-68.

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, *115*(1), 362. doi:10.1121/1.1635842

Lisker, L. (1975). Is it VOT or a first-formant transition detector? *The Journal of the Acoustical Society of America*, *57*(6 Pt 2), 1547-1551.

Liu, H., Kuhl, P. K., & Tsao, F. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, *6*(3), F1-F10.

Liu, H., Tsao, F., & Kuhl, P. K. (2009). Age-related changes in acoustic modifications of Mandarin maternal speech to preverbal infants and five-year-old children: a longitudinal study. *Journal of Child Language*, *36*(4), 909-922.

Macmillan, N., & Creelman, C. D. (2005). *Detection Theory : A User's Guide* (2nd ed.). Mahwah N.J.: Lawrence Erlbaum Associates.

MacWhinney, B. (2009). *CLAN*. Pittsburgh, Pennsylvania: Carnegie Mellon University. Retrieved from childes.psy.cmu.edu/clan/

Malsheen, B. J. (1980). Two hypotheses for phonetic clarification in the speech of mothers to children. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child Phonology*, Perspectives in Neurolinguistics, Neurophsychology, and Psycholinguistics (Vol. 2, pp. 173-184). New York London: Academic Press.

Mandel, D., Kemler Nelson, D. G., & Jusczyk, P. W. (1996). Infants remember the order of words in a spoken sentence. *Cognitive Development*, *11*(2), 181-196.

Mandel, D. R., Jusczyk, P. W., & Kemler Nelson, D. G. (1994). Does sentential prosody help infants organize and remember speech information? *Cognition*, *53*(2), 155-180.

Maye, J., Weiss, D. J., & Aslin, R. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science*, *11*(1), 122-134.

Maye, J., Werker, J., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-111.

McRoberts, G. W., & Best, C. T. (1997). Accommodation in mean f0 during mother–infant and father–infant vocal interactions: a longitudinal case study. *Journal of Child Language*, *24*(3), 719-736.

Messer, D. J. (1981). The identification of names in maternal speech to infants. *Journal of Psycholinguistic Research*, *10*(1), 69-77.

Moslin, B. J. (1979). *The Role of Phonetic Input in the Child's Acquisition of the Voiced-Voiceless Contrast in English Stops: A Voice Onset Time Analysis* (Unpublished Doctoral Dissertation). Brown University, Providence, RI.

Newman, R. S., & Hussain, I. (2006). Changes in Preference for Infant-Directed Speech in Low and Moderate Noise by 4.5- to 13-Month-Olds. *Infancy, 10*(1), 61-76.

Rimol, L., Eichele, T., & Hugdahl, K. (2006). The effect of voice-onset-time on dichotic listening with consonant–vowel syllables. *Neuropsychologia*, *44*(2), 191-196.

Shockey, L., & Bond, Z. (1980). Phonological processes in speech addressed to children. *Phonetica*, *37*(4), 267-274.

Sivonen, P., Maess, B., & Friederici, A. D. (2006). Semantic retrieval of spoken words with an obliterated initial phoneme in a sentence context. *Neuroscience Letters, 408*(3), 220-225.

Smiljanic, R., & Bradlow, A. R. (2008). Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics*, *36*(1), 91-113.

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *The Journal of the Acoustical Society of America*, *55*(3), 653-659.

Stoel-Gammon, C., Williams, K., & Buder, E. (1994). Cross-Language Differences in Phonological Acquisition: Swedish and American /t/. *Phonetica, 51*(1-3), 146-158.

Sundberg, U. (2001). Consonant specification in infant-directed speech: some preliminary results from a study of voice onset time in speech to one-year-olds. *Working Papers*, *49*, 148-151.

Sundberg, U., & Lacerda, F. (1999). Voice onset time in speech to infants and adults. *Phonetica*, *56*(3-4), 186-199.

Swets, J. A., & Pickett, R. M. (1982). *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. New York: Academic Press.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53-71.

Uther, M., Knoll, M., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication*, *49*(1), 2-7.

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(33), 13273-13278.

Vosoughi, S., Roy, B., Frank, M. C., & Roy, D. (2010). *Effects of caregiver prosody on child language acquisition*. Presented at the Speech Prosody International Conference, Chicago.

Werker, J., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, *103*(1), 147-162.

Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: a developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology*, *43*(2), 230-246.

Zlatin, M. A. (1974). Voicing contrast: Perceptual and productive voice onset time characteristics of adults. *Journal of the Acoustical Society of America*, *56*(3), 981-994.