STOP CONSONANT VOICING IN YOUNG CHILDREN'S SPEECH: EVIDENCE FROM A CROSS-SECTIONAL STUDY

A THESIS
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Emily Ganser

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF ARTS

Benjamin Munson, Ph.D.

May 2016

Acknowledgements

Thanks are due to the National Institute for Deafness and Other Communicative Disorders (grant NIDCD 02932) and the National Science Foundation for providing funding for the Learning to Talk project. Thanks are also due to the many members of the Learning to Talk project team, whose work made this study possible, and to the many families who participated in the research study. I also thank Dr. Ben Munson for his willingness to advise me.

I am grateful for my parents, who happily supported all of my academic endeavors and were willing to invest in the opportunity for me to write this thesis, and for good friends, who patiently listened to, supported, and encouraged me throughout the writing process.

Abstract

There are intuitive reasons to believe that speech-sound acquisition and language acquisition should be related in development. Surprisingly, only recently has research begun to parse just how the two might be related. This study investigated possible correlations between speech-sound acquisition and language acquisition, as part of a large-scale, longitudinal study of the relationship between different types of phonological development and vocabulary growth in the preschool years. Productions of voiced and voiceless stop-initial words were recorded from 96 children aged 28-39 months. Voice Onset Time (VOT, in ms) for each token context was calculated. A mixed-model logistic regression was calculated which predicted whether the sound was intended to be voiced or voiceless based on its VOT. This model estimated the slopes of the logistic function for each child. This slope was referred to as *Robustness of Contrast* (based on Holliday, Reidy, Beckman, and Edwards, 2015), defined as being the degree of categorical differentiation between the production of two speech sounds or classes of sounds, in this case, voiced and voiceless stops. Results showed a wide range of slopes for individual children, suggesting that slope-derived Robustness of Contrast could be a viable means of measuring a child's acquisition of the voicing contrast. Robustness of Contrast was then compared to traditional measures of speech and language skills to investigate whether there was any correlation between the production of stop voicing and broader measures of speech and language development. The Robustness of Contrast measure was found to correlate with all individual measures of speech and language, suggesting that it might indeed be predictive of later language skills.

Table of Contents

List of Tables

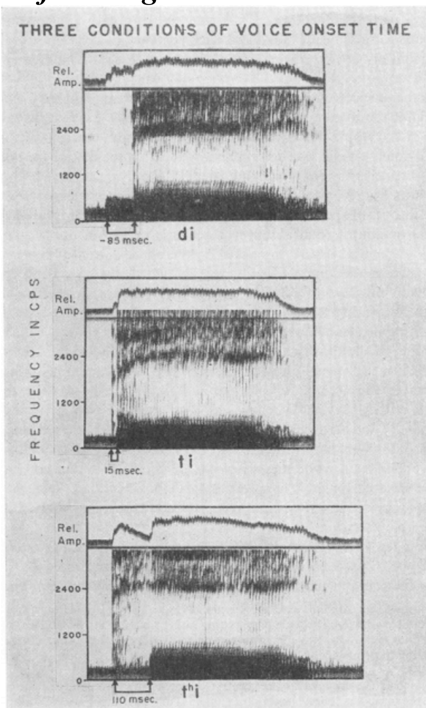
List of Figures

1 Introduction

It is without question that human language (i.e., the formal system used to share ideas and mental states among individuals) and the modalities used to convey language are interrelated. This thesis examines relationships between speech-sound acquisition and language acquisition. The study of these interrelationships is motivated in part by the fact that there are intuitive reasons to believe that the two should be related in development. After all, speech is arguably the most commonly used medium for expression of language, and it serves no function other than to convey language. However, it is a great undertaking to parse exactly how they are related, and how they might influence one another in development. Indeed, it seems reasonable enough to presume that speech and language *do* have an influence on one another. The study of relationships between speech and language is focused on numerous questions, including the direction an influence might go, or whether such an influence may be bidirectional. Some of the challenge lies in the very nature of the speech signal: it is produced as a continuous stream with no clear boundaries to designate the beginning or end of words; its components (phonemes) last only milliseconds, and the subtlest of variations in their productions can result in drastically different outcomes; the speech signal, unlike written language, is fleeting; and it is highly influenced by the perception of its recipient (i.e., the listener). It is understandable that the intricacies of the relationship between speech and language have yet to be fully understood. This study, therefore, will examine one small aspect of the speech-language relationship in the course of speech and language development. The goal of this thesis is to document and better understand the relationships between speech and language acquisition, with a broader goal of contributing to the understanding of how these topics are related more generally.

The specific topic that this thesis examines is the development of voicing in initial stop consonants in children acquiring English. Voice Onset Time (VOT, typically measured in ms) is the duration between two events: (1) the end of the stop consonant closure and the subsequent release of air that built up during the closure, and (2) the initiation of vocal fold vibration in the subsequent vowel. VOT is a continuous variable: a VOT of 0 ms indicates that the two events happen simultaneously; a negative VOT indicates that voicing begins before the release of the stop consonant closure; and a positive VOT indicates that voicing begins after the release of the closure. Though VOT is a continuous variable, it is generally described by experimental phoneticians as falling in three categories, depending on the length of time between the release of energy and the initiation of voicing: minus, or prevoicing, when the onset of vocal fold vibration begins before the stop closure is released; zero, or short-lag, when vocal fold vibration begins essentially simultaneously with the release of the stop; and long-lag, when there is a considerable amount of time (generally at least 40 ms) between the release of the stop closure and the onset of voicing (Figure 1). This distinction of VOT is just one example of how phonetic contrasts can be cued. While all three variations of VOT can co-exist in any one language (such as Thai), only short-lag and long-lag VOT exist in adult-like speech in English, where, in word-initial position, a short-lag VOT is associated with a phonologically [+voice]/voiced stop (e.g., /d/ or /g/) and a long-lag VOT is associated with a phonologically [-voice]/voiceless stop (e.g., /t/ or /k/).

Figure 1: ***Waveform representation of the three conditions of voice onset time: minus, short-lag, and long-lag, taken from Figure 1 Lisker and Abramson (1964:p. 390).***



The current study compared the VOT of voiced stops (/d/ and /g/) and voiceless stops (/t/ and /k/) in the production of children aged 28-39 months to determine whether individual children produced a distinct difference in VOT between the voiced and voiceless stop targets. The general principle that underlies the use of this measure is that phonological acquisition involves the gradual emergence of contrast. Classic studies of phonological development have used categorical measures of speech-production accuracy, like phonetic transcriptions. In these models (i.e., Jakobson, 1941), contrasts are thought to emerge in a stepwise, all-or-none fashion. Conversely, more recent work using a variety of experimental techniques has found that development involves the *gradual* differentiation between pairs of sounds or classes of sounds. Indeed, studies of VOT acquisition provide the foundation for much of this work. The acquisition and then refinement of VOT is something that takes years to master. Only milliseconds between the release of energy of a stop consonant and the onset of vocal fold vibrations in the subsequent vowel distinguish an unvoiced plosive (e.g., /t/ or /k/) from its voiced counterpart (i.e., /d/ and /g/). While this miniscule difference of VOT across stop consonants can be perceived by infants as young as one month (as found in Eimas, Siqueland, Jusczyk, and Vigorito's seminal 1971 study of infant speech perception), consistent, systematic production of VOT is not achieved until years later. Numerous studies have been conducted to determine the age of acquisition of the voicing contrast (e.g., Hammarström, Larsson, Wiman, & McAllister, 2012; Hitchcock, 2005; Hitchcock & Koenig, 2004; Hitchcock & Koenig, 2015; Lowenstein, & Nittrouer, 2008; Nittrouer, 1993; Nittrouer, Estee, Lowenstein, & Smith, 2005; Smit, Hand, Freilinger, Bernthal, & Bird, 1990), but the results have shown great variability in the development of these fine distinctions.

Young children are physiologically capable of producing voicing contrasts, but they do not always use these contrasts systematically as do adults. This lack of clear and consistent distinction of VOT between voiced and voiceless stops is the first stage in the development of voicing contrast. In the second stage of development, children begin to produce systematic distinctions between voiced and voiceless stops (i.e., *voicing contrasts)*, but not to the same extent of robustness as adults. This means that there are measurable, albeit unperceivable, differences in VOT production, which are referred to as *covert contrasts*.

Macken and Barton, perhaps the pioneers for research of covert contrasts, found some of the earliest evidence for this subject in their 1980 study. In this hallmark study, four children (aged 1;4-1;7 at the onset of the study) were recorded every two weeks over the course of eight months producing voiced and voiceless word-initial stop consonants. The children's productions were both transcribed phonetically (to determine listener perception) and analyzed acoustically, where VOT was determined. The results of the study showed that the children's productions fell into one of three categories: no difference in VOT between the voiced and voiceless stops; a difference in VOT between the voiced and voiceless stops that was considered to be adult-like both in its acoustic measures and by listener perception; or a difference in VOT between voiced and voiceless stops that was indicated by acoustic measures but was not perceivable to a listener.

Lowenstein and Nittrouer (2008) investigated the acquisition of VOT by analyzing the speech samples of seven children who were recorded between approximately 14 and 31 months at two-month intervals. (The non-uniform age range was due to initiating the study when the children had at least ten recognizable words and ceasing when the children were speaking in three word sentences.) The study found variability of VOT acquisition amongst the seven children. Four of the seven followed the expected pattern of acquisition of VOT: their VOTs for initial voiced stops were considered to be adult-like at the first session, and their VOTs for initial voiceless stops increased past what is typical of adult productions over the course of the study. The other three children, however, produced VOTs for voiceless stops on the "low end" of what it considered to be normal for adult-like productions; two of the three increased the length of their VOTs for voiceless stops over the course of the study, but one child's VOT did not change.

Overall these findings suggest that the acquisition of VOT for the seven children followed a gradual process. At the beginning of the study, the children's VOT for voiceless stops was in the "short-lag" or "ambiguous" range (i.e., not perceived as voiceless by a listener); but by 23-24 months, VOT was within a range to be consistently perceived as voiceless by a listener. Additionally, this study found great variability for VOT for words with voiceless initial stops amongst the seven children for the entire duration of the study.

A study by Hammarström et al. (2012) looked at the acquisition of adult-like VOT in relatively older Swedish-speaking children. They used speech samples from 150 children, who were divided up into four age groups (7;9-8;8, 8;9-9;8, 9;9-10;8, 10;9-11;8) and compared the VOT of the children's productions of plosives across the age groups to that of 36 adults. All six Swedish plosives (voiceless /p/, /t/, /k/ and voiced /b/, /d/, /g/) were elicited in two contexts: sentence completion and picture naming. The results showed that all age groups of the children produced a voicing lag (i.e., positive VOT) for voiceless stops, and the mean VOT increased as the place of articulation moved posteriorly, with the difference across place being significant. There were statistically significant differences in VOT for each stop across the age groups. For voiced stops, incidence of prevoicing (typical of adult productions of Swedish voiced stops) increased with age. Additionally, compared to the adults, the two youngest groups of children had notably different VOTs for voiced stops. In sum, Hammarström et al. found that all subjects demonstrated a clear distinction between voiced stops and voiceless stops by producing very different VOTs for the two voicing categories. The findings of the study also showed a developmental trend for length of VOT in voiceless stops and incidence of prevoicing in voiced stops over the four age groups. Moreover, while there was variability in VOT times for both voiced and voiceless stops within age groups, this variability decreased with age. Overall, the study concluded that Swedish children seem to acquire adult-like VOT production between nine and ten years of age.

In Hitchcock and Koenig's (2013) study, voiced and voiceless word-initial stops (/b, p, d, t/) were elicited and recorded from ten children every other week for four months, and VOT was subsequently measured for each stop token context. Additionally, measures of "accuracy" (adult-like values), "discreteness" (the extent of overlap between contrastive VOT categories), and "overshoot" (exaggerated long-lag values) were calculated. The researchers then compared the children's mean VOT measures with a narrower token-by-token analysis. They found that the children showed changes in their development of VOT that were not always evident based on VOT means alone: the children's accuracy, discreteness, and overshoot changed statistically significantly even after VOT means were of a statistically significant difference. Moreover, the greater analysis revealed lower accuracy and greater category overlap in the children's productions than the VOT means had suggested, and the children were found to go through a phase of overshoot. The researchers also found that the children were not always consistent with their productions across sessions (i.e., distinct voicing contrasts observed one week were not always observed in the subsequent session), and the children demonstrated different speeds and means of development of the voicing contrast (some made abrupt, rapid gains; others made gains gradually). These findings suggest that exclusively using mean VOTs as the measure of the acquisition of voicing may not be enough to give a complete picture of a child's development of voicing. The study also found that even though the 2-year-olds produced statistically significant differences between categories of voicing, their productions still differed from the target, adult-like voicing contrast. This finding is consistent with other studies that have also found that children seem to gradually acquire an adult-like voicing contrast even after having developed a statistically significant contrast.

Hitchcock and Koenig extended this study to follow one of the child participants for an additional eight months (i.e., 12 months total), and concluded that the development of voicing includes a period of fluctuation, and it may take months or years for a child to develop a consistent, adult-like voicing contrast (Hitchcock & Koenig, 2015).

In a study that analyzed the production of voiced and voiceless word-final stops in eight children (four 5-year-olds and four 7-year-olds), Nittrouer et al. (2005) found that the development of articulatory gestures was not uniform across the individual speakers. The children were observed to produce distinct voiced and voiceless word-final stops, but their productions were not consistent. The researchers concluded that children as old as seven years still have not fully mastered the gestures required to produce voiced and voiceless word-final stops.

Macken and Barton's (1980) original finding that children could produce a third category of VOT productions, indicated by acoustic measures but not perceivable to a listener, led to many subsequent studies of covert contrast for other features in acquisition. Forrest, Elbert, Weismer, and Dinnsen (1994) used acoustic measures to show that listener perception of children's production of /t/ and /k/ did not always align with the corresponding acoustic outputs. Gibbon (1990) used electropalatography to show that children, whose production of /d/ and /g/ were perceived to be identical, were actually using distinctly different articulatory gestures to produce the two sounds.

Gierut and Dinnsen (1986) examined two children who were perceived to have similar voice-contrast error patterns but were shown to have markedly different voice contrasts when considering acoustic analyses. They emphasized the limitations of convention means of gathering phonological data (i.e., phonetic transcription). The acoustic analyses of their study revealed a greater productive knowledge of contrastive voicing in stops in one child, who used voicing distinctions systematically, than the other child, who used voicing in no such systematic way. This productive knowledge could not be captured by phonetic transcription alone.

In recent years, studies documenting covert contrasts have become ubiquitous. They have been shown for the acquisition of lingual sibilant contrasts in English and Japanese (Li, Beckman, & Edwards, 2009; Li, 2012), and for lingual stop contrasts in English (Edwards, Gibbon, & Fourakis, 1997; Forrest, Weismer, Hodge, Dinnsen, & Elbert, 1990), among others.

Studies of covert contrast are not merely important for giving a full picture of speech-sound development. They have also shown that covert contrast has prognostic utility when describing the speech of children with speech sound disorder. Tyler, Figurski, and Langsdale (1993) showed a clinical application for determining a child's productive knowledge (assessed via acoustic analysis) rather than relying solely on phonetic transcription. They found that children with phonological disorders who demonstrated a distinction in stop voicing contrast (indicated by acoustic rather than perceptive measures) required a shorter treatment period than their peers who did not demonstrate such distinction.

The studies cited in the preceding paragraphs suggest that a more comprehensive view of speech-sound acquisition can be gained when considering the extent to which children produce a contrast between speech sounds. This, in turn, calls for the development of a measure of the extent to which an individual produces a contrast between two categories, like the voiceless and voiced stops in this thesis. This degree of differentiation of VOT between voiced and voiceless stop targets is referred henceforth as *Robustness of Contrast.* The specific measure used in this thesis was presented by Holliday, Reidy, Beckman, and Edwards (2015). Holliday et al. used a mixed-model logistic regression to measure the degree of sound category overlap, which they termed Robustness of Contrast (ROC). Phonemes that are more separated in a particular acoustic dimension are considered to have greater ROC, and completely overlapping or undifferentiated categories are considered to have the weakest ROC. Holliday et al. examined children's productions of /s/ and /ʃ/, the difference between which is well characterized by measures of the spectral peak in the frication interval. Mixed-model logistic regression predicted whether the target was /s/ or /ʃ/ (arbitrarily coded as 1 or 0) from the peak frequency of the spectrum of the fricative. This model estimated the overall effect of peak frequency on whether a sound was classified as /s/ or /ʃ/ and the extent to which this was true for the individual children who contributed data to the model. The model generated individual-subjects' slopes for the effect of peak frequency on target-sound classification. The slope of the logistic function was very steep for children with clear distinction between production of /s/ and /ʃ/. For children with overlapping productions of /s/ and /ʃ/ (i.e., not clearly differentiated), the slope was found to be much shallower. The slopes varied continuously. Holliday et al. found these measures of slope to have a positive correlation with age and vocabulary size. That is, older children and children with larger-sized vocabularies had more robust contrasts between /s/ and /ʃ/ than did younger children and children with smaller-sized vocabularies.

For this study, ROC was defined as the extent to which VOTs were differentiated by individual subjects for target voiced and voiceless tokens (i.e., /d/ and /g/ versus /t/ and /k/). For those children who were found to produce robust voicing contrasts, this study also investigated the degree to which this ROC in voicing might correlate with traditional measures of speech development and with language skills, as measured by standardized assessments. For many children of this age, voicing contrast is emerging systematically but does not yet mirror adult-like voicing contrast. This means that for this study, a large range of the maturation of voicing contrast was represented, allowing for maximal comparison with language ability.

The research has made it clear that the development of articulatory abilities, including the complex articulatory coordination required to produce a particular VOT, takes years to master. Additionally, it has been affirmed that listener perception is not sensitive enough to fully capture a child's phonological skills or productive knowledge, as is the case with the voicing contrast; rather, acoustic measures must also be considered. But even with complete consideration for a child's otherwise unperceived phonological skills or productive knowledge, the literature is lacking in its comparison of such "speech" skills with language skills. While many studies have looked at phonological skills mostly independently from language skills (and vice versa), and some studies have investigated a relationship between phonological skills and language skills (e.g., Beckman & Edwards, 2000; Edwards, Beckman & Munson, 2004; Munson, Edwards, & Beckman, 2005; Munson, Kurtz, & Windsor, 2005; Sosa & Stoel-Gammon, 2012) only a few studies have begun to attempt answering this much bigger question of whether phonological development could in fact be a *predictor* of later language skills.

In 2001, McCune and Vihman conducted a study to compare phonetic development with vocabulary acquisition and growth. Twenty children, nine months of age at the onset of the study, were recorded at their home once a month for eight months during unstructured mother-child interactions, and the recordings were transcribed. Each month a parent report regarding the child's word production and comprehension was also completed to supplement the recordings. The transcriptions of the children's speech sounds were examined for consistency, and the level of production consistency was then compared to the children's word production. This comparison showed that the number of specific speech sounds that were produced consistently over the course of the study predicted word production (specifically, "referential lexical use"), and the children who demonstrated referential lexical use earliest had actually demonstrated consistent use of certain speech sounds prior to the onset of the study. These findings are certainly consistent with the notion that phonological development is related to, and may well be a predictor of, language development.

A literature review by Storkel and Morrisette (2002) similarly looked at how phonology interacts with the lexicon during language acquisition. They examined numerous descriptive and experimental studies to explore the link between lexical and phonological development in children with 50 or more words and considered the how the lexicon might affect phonological development (i.e., learning sounds) and how phonological consideration might affect word learning. Their findings supported the notion that lexical and phonological development continue to influence each other even after a child has 50 or more words.

Finally, Zanobini, Viterbori, and Saraceno (2012) also looked into possible relationships between phonology and language. This study investigated which phonetic factors might affect lexical and morphosyntactic skills, and it examined correlations between phonological skills and language skills. Two standardized tests, one measuring phonological ability and one measuring "general linguistic ability," were administered to 30 Italian children aged 36 to 42 months and the results were analyzed. The findings from the phonemic and linguistic measures indicated a strong negative correlation between the intelligibility of the children's productions (i.e., their phonological accuracy) and their linguistic ability, which supports the hypothesis that there is a relationship between phonology and linguistic, or language, skill.

1.1 Aims of this study

The aims of this study were twofold. First, this study set to investigate the possible use of robustness of voicing contrast as a measure of the acquisition of voicing. Research has already determined that traditional means of measuring phonological skills (i.e., phonetic transcription) is neither complete nor fully accurate due to listener perception bias and the presence of covert contrast. But even while considering covert contrasts in the assessment of a child's phonological skills, the literature has yet to offer an objective measure of when exactly a child has fully acquired voicing contrast (Hitchcock, 2005; Hitchcock & Koenig, 2004; Lowenstein & Nittrouer, 2008; Smit, Hand, Freilinger, Bernthal, & Bird 1990). Being able to determine with certainty whether a child had developed a voicing contrast would both contribute to the ongoing research to more completely understand the development of speech and language and would be beneficial clinically to better assess and diagnose phonological disorders. For this study, it was predicted that ROC would prove to be a better, more objective predictor of the acquisition of VOT. Since research has already found that children produce covert voicing contrasts during their development of VOT (Forrest, Elbert, Weismer, & Dinnsen, 1994; Gibbon, 1990; Gierut & Dinnsen, 1986; Macken & Barton, 1980), it seems reasonable that the measure of the robustness of these covert contrasts could be used as a predictor for VOT acquisition.

The second aim of this study was to compare the ROC measure in the stop production of children to the children's corresponding speech and language assessment scores, to investigate whether a more robust voicing contrast could be predicted by current language skills, which might suggest that it would perhaps predict later language skills. Research has shown a correlation between vocabulary size and phonological skills, where a larger vocabulary can be predictive of certain aspects of phonological skills (e.g., Edwards, Beckman & Munson, 2004; Stoel-Gammon, 1991), but there has been little research investigating the effect that strong phonological skills have on language ability. Given the findings of previous research on this relationship between phonological skills and language, it was hypothesized that stronger phonological skill (as determined by production of a more robust voicing contrast during speech production) would correlate with stronger speech and language skill (as measured by standardized assessments).

2 Methods

Nota Bene

The current study used data collected from the participants of a larger longitudinal research project, *Learning to Talk* (see http://www.learningtotalk.org/). It is to be expected that there will be overlap between the Methods of this study and those of previous studies that also used data from this project, for example, Sara R. Bernstein's 2015 thesis, *"Individual differences in the acquisition of the /t/ - /k/ contrast: A study of* *adults' perception of children's speech"* (University of Minnesota) and 2016 University of Wisconsin-Madison theses by Hyuna Kim and Allison Johnson.

2.1 Children participants

The speech sounds analyzed by this study were produced by 96 children, aged 28-39 months, who were recruited to participate in the larger longitudinal *Learning to Talk* project, which is investigating the development of phonological knowledge and vocabulary. The children were recruited to participate via newspaper advertisements and fliers posted around the University of Minnesota and the surrounding community. All the children were from monolingual English-speaking households (as determined by caregiver report) and represented a range of maternal education. Both Mainstream American English (MAE) and African American English (AAE) speakers participated in the study. The participants also included *late talker* children, who were defined as such by having typically developing receptive language and prelinguistic skills but having expressive language skills that fell outside the normal limits for age-matched peers without any other speech, language, hearing, or developmental diagnoses.

The child participants were recorded at the University of Minnesota and the University of Wisconsin - Madison. Before beginning their initial session, the participants passed a hearing screening at 1000, 2000, and 4000 Hz at 25 dB HL. The children completed their testing over two or three visits for one to two hours per visit. A variety of measures were used to determine child-level differences in speech, language, and related skills in order to identify potential predictors of speech production abilities. This project focused on measures of a child's individual performance. The measures consisted of a series of standardized and nonstandardized assessments, including experimenter-administered tasks and parent report determined via questionnaires. Variables measured included speech perception, vocabulary, executive function, and articulation (Table 1). Tests were administered by trained undergraduate and graduate students in accordance with any standardized protocols.

2.2 Individual performance assessments

Speech perception was measured because of its close relationship to speech production. Many errors with speech sound production can be a result of a difficulty with speech perception (Rvachew & Grawburg, 2006). Speech perception was also determined to be an important skill to consider because it provides insight into a child's phonological knowledge. It was hypothesized that the children who demonstrated greater phonological knowledge in the speech perception task would also produce the target sounds of the speech production task with greater adult-like accuracy. In this study, speech perception was measured through a minimal pair picture discrimination task. For this task, a target word was presented to the child via speakers, and two corresponding pictures were displayed on a touch screen, one of the target word, the other of a word that differed from the target word by one speech sound (e.g., the word "goat" presented over speakers and a picture of a goat and a picture of a boat displayed on the screen). To indicate which word they had perceived, the children responded by directly selecting one of the images.

Vocabulary size was measured through multiple assessments, most notably because a large vocabulary size has been shown to be correlated with a variety of types of phonological knowledge, including phonotactic knowledge (Edwards, Beckman & Munson, 2004; Munson, Edwards, & Beckman, 2005; Munson, Kurtz, & Windsor, 2005; Sosa & Stoel-Gammon, 2012), speech perception (Edwards, Fox, & Rogers, 2002), and articulatory ability (Holliday, Reidy, Beckman, & Edwards, 2015). Administered assessments included the Expressive Vocabulary Test – 2nd Edition (EVT-2, Williams, 2007), to measure vocabulary production, and the Peabody Picture Vocabulary Test – 4th Edition (PPVT-4, Dunn & Dunn, 2007), to measure vocabulary comprehension. The MacArthur Bates Communication Development Inventory, a parent-completed questionnaire, was also used to determine the total number of words a child produces across environments (Fenson, Marchman, Thal, Dale, Reznick, & Bates, 2007). It was hypothesized that the children with the highest language scores across all measures would produce the target sounds of the speech production task with more robust contrasts.

The "Fruit Stroop" test was administered to measure one aspect of executive function skills - inhibitory control - since attending to relevant information while ignoring irrelevant information is an important skill for speech perception and production and for completing the complex assessments in this protocol. For this test, a child was showed a picture of a small fruit overlaid on a different, larger fruit, and he was asked to attend to the small fruit while ignoring the larger fruit. Additionally, the Behavior Rating Inventory of Executive Function (BRIEF) questionnaire was completed by the children's parents as a parent-report measure of the children's behavior regulation and metacognition (Gioia, Espy, & Isquith, 2003).

Finally, the Goldman-Fristoe Test of Articulation -2nd Edition was also administered to the participants (GFTA-2, Goldman & Fristoe, 2000). This traditional means of assessing articulation using phonetic transcription was selected to be a direct comparison to the non-standardized means of assessing speech sounds that the current study investigated. It was hypothesized that higher scores of articulation would correlate with more robust voicing contrasts during the speech production tasks, but it was also hypothesized that some lower GFTA scores could be correlated with robust voicing contrasts as a result of covert contrasts. Numerous studies have concluded that using phonetic transcription as a measure of articulation does not fully represent a child's phonological knowledge (Forrest et al., 1990; Forrest et al., 1994; Gierut & Dinnsen, 1986; Li, 2012), thus finding a correlation between the traditional means of assessing phonology and using acoustic signals to assess phonology would be an important foundation for eventual shifting towards the use of objective data as a superior way of measuring a child's phonological knowledge.

Table 1: *Individual performance assessments*

Task Name	Reference	Construct Measured	Description
GFTA-2	Goldman & Fristoe (2000)	Articulation	Standardized, norm-referenced assessment of articulation using picture book to prompt naming response.
EVT-II	Williams (2007)	Expressive Vocabulary	Standardized, norm-referenced assessment of expressive vocabulary using a picture book and prompting questions to produce the desired vocabulary.
PPVT-IV	Dunn & Dunn (2007)	Receptive Vocabulary	Standardized, norm-referenced assessment of receptive vocabulary using a picture book displaying four pictures on each page. Examiner prompted pointing response using standardized "show me" statement or equivalent.
Fruit Stroop	Carlson (2005)	Executive Function	A measure of cognitive flexibility. Examiner used cards depicting three different fruits that were both large and small sizes. After labeling the fruit and the size, the examiner displayed cards of smaller fruits inside larger fruits and asked the child to point to a particular small fruit. Trials were scored correct (i.e. correct small fruit) or incorrect (i.e. large fruit)
Minimal Pair Discrimination	Baylis, Munson, & Moller (2008)	Speech Perception	Two picturable, early-acquired minimal pair words were presented to a child one at a time. A recording of one of those two words was then presented with both pictures on the screen. Child participants chose which picture the recording produced.
Real Word Repetition	Edwards & Beckman (2008)	Articulation	Using a recorded voice, children repeated a list of early-acquired, picturable words, balanced for vowel context.

Table used with permission from Kramer, 2016.

2.3 Speech production data collection

The speech productions used for this study were recorded during a picture-based auditory word repetition task. The task was administered via a computer running E-Prime software. Klipsch BT77 speakers, which had been normalized to 70 dB in a sound-treated booth, were used to present the auditory prompts. An Audito Technica (AT 4040) cardioid capacitor microphone and a Marantz Professional solid state recorder (PMD671) were used to record speech productions. Speech production data were collected by trained undergraduate and graduate students.

For the word repetition task, 99 test trials of target words, which were selected to be highly familiar to children, were presented over the speakers (with an accompanying picture on the computer screen) to the child participants, who verbally repeated the stimulus. Each target word was presented at least twice during the 99 test trials, and all the stimuli were presented in a random order. Children were reinforced to participate during the task with a visual reinforcer (an image of an animal climbing a ladder as progress was made), verbal praise/encouragement, and stickers. If a child did not respond to the presented stimulus or produced an incorrect response, test administers were instructed to give a general verbal prompt rather than a direct model.

The stimuli consisted of 17 target words with an initial voiceless stop. The targets were selected to include high front, high back, and low back vowel contexts. Nine of the 17 voiceless stop words were /t/ (alveolar) initial (tummy, table, toast, tooth, tongue, tape, teddy bear, tickle), and eight were /k/ (velar) initial (kitty, kitchen, candy, coat, car, cake, cup, cat, cookie).

The stimuli also consisted of 15 target words with an initial voiced stop with various vowel contexts. Seven of the voiced stop words were /d/ (alveolar) initial (daddy, dance, dinner, dish, dog, door, duck) and the remaining seven were /ɡ/ (velar) initial (garbage, get, girl, give, go, good, gum).

The remaining stimuli consisted of words with other initial speech sounds to be used for other studies, such as the /s/ and /ʃ/ productions examined in Kramer's 2016 summa cum laude thesis, "*Predictors of early sibilant fricative production as evidenced by naive listener perception ratings"* (University of Minnesota).

2.4 Recording segmentation

After speech productions were elicited, target words were extracted from the recordings in a process referred to as *segmentation*. Trained students used scripts written by members of the *Learning to Talk* project on Praat software to segment the recordings. For each child's recording, a text grid was created that included the target stimulus, boundaries of the child's production, and the production number. Notes were included to provide information about the nature of the child's production (e.g., whether it immediately followed the stimulus or whether it was elicited by a verbal prompt) and any issues with the recording (e.g., background noise, production too quiet or loud). All segmented recordings were checked by an additional trained student before being used for tagging acoustic events.

2.5 Acoustic event tagging

Since great detail of the process of tagging acoustic events can be found in previous papers (e.g., Bernstein, 2015), a broader overview of the process will follow in order to avoid redundancies.

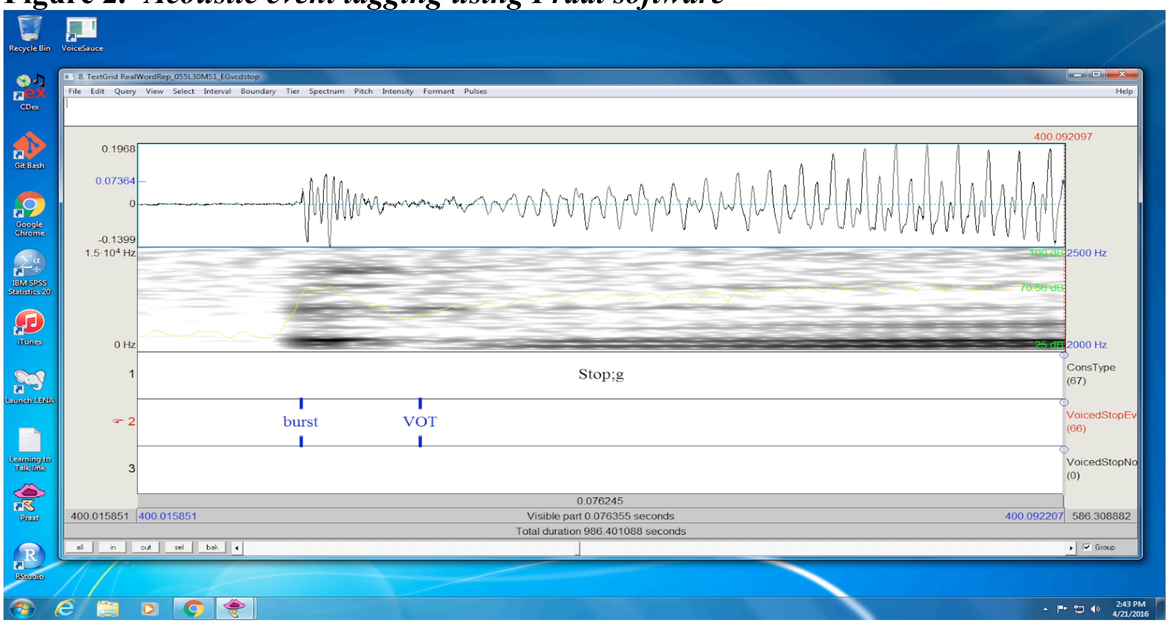
Acoustic events were tagged using Praat software with custom-made scripts. Four trained graduate students tagged voiceless stops for all recordings and one trained graduate student tagged all voiced stops for all recordings. All graduate students, aka *burst-taggers*, followed a specific pre-determined protocol for tagging acoustic events. (This protocol can be found in the Appendix of Bernstein, 2015.) Burst-taggers first opened the text grids that were extracted during the segmentation process using Praat software. One trial at a time, the burst-taggers listened to the initial consonant and vowel of the child's production of the target word and determined if the production would be usable for tagging acoustic events. If the first production was deemed unusable, alternative productions (if any) were also listened to for usability. If no production was considered to be usable for tagging, the trial was omitted. Reasons why a production would have been considered unusable included background noise, clipping of the waveform, or inaudible or deleted burst.

Once a useable production was determined, burst-taggers transcribed the perceived manner (i.e., stop, affricate, or other) and place of articulation (e.g., alveolar [t], velar [k], intermediate [t] sounding a bit like [k], intermediate [k] sounding a bit like [t], or other). Any productions whose manner was perceived to be *affricate* or *other* were not included in the dataset analyzed.

After transcribing perception, burst taggers noted any anomalies with the production or sound of the trial (e.g., background noise, clipping of the waveform, deleted burst). Next, the burst taggers looked at the spectrogram to determine where the burst of the initial consonant and the onset of voicing were (the two acoustic events tagged) (Figure 2). The burst was considered to be the first peak of the waveform of the child's production, clearly deviant from the baseline waveform and was tagged as such.

Voice onset was defined to be the beginning of the voice cycle, noted by an upswing of the waveform followed by a clear downswing below the zero line, with a continuation of the waveform pattern proceeding subsequently. Voice onset was always tagged at a zero crossing. VOT was then calculated by measuring the time between the burst tag and the voice onset tag.

Figure 2: ***Acoustic event tagging using Praat software***



3 Results

3.1 Individual differences measures

In all of the individual differences measures (i.e., GFTA-2, EVT-2, PPVT-4, Fruit Stroop, and Minimal Pair Identification) a wide range of scores were represented. The ranges of performance for the individual differences measures can be found in Table 2. VOTs for voiced and voiceless token followed expected patterns: target voiced stops were produced with shorter VOTs than were target voiceless stops. A wide range of measures of VOT for both voiced and voiceless stop targets was also observed. This affirms that the participants did not all produce uniform VOT for either voicing target. The range of measures of VOT can be seen in Figures 3 and 4 below. Overall, these findings indicate that there was no restriction of range in any of the individual differences measures.

Figure 3: ***Histogram of [-voice] stop targets***

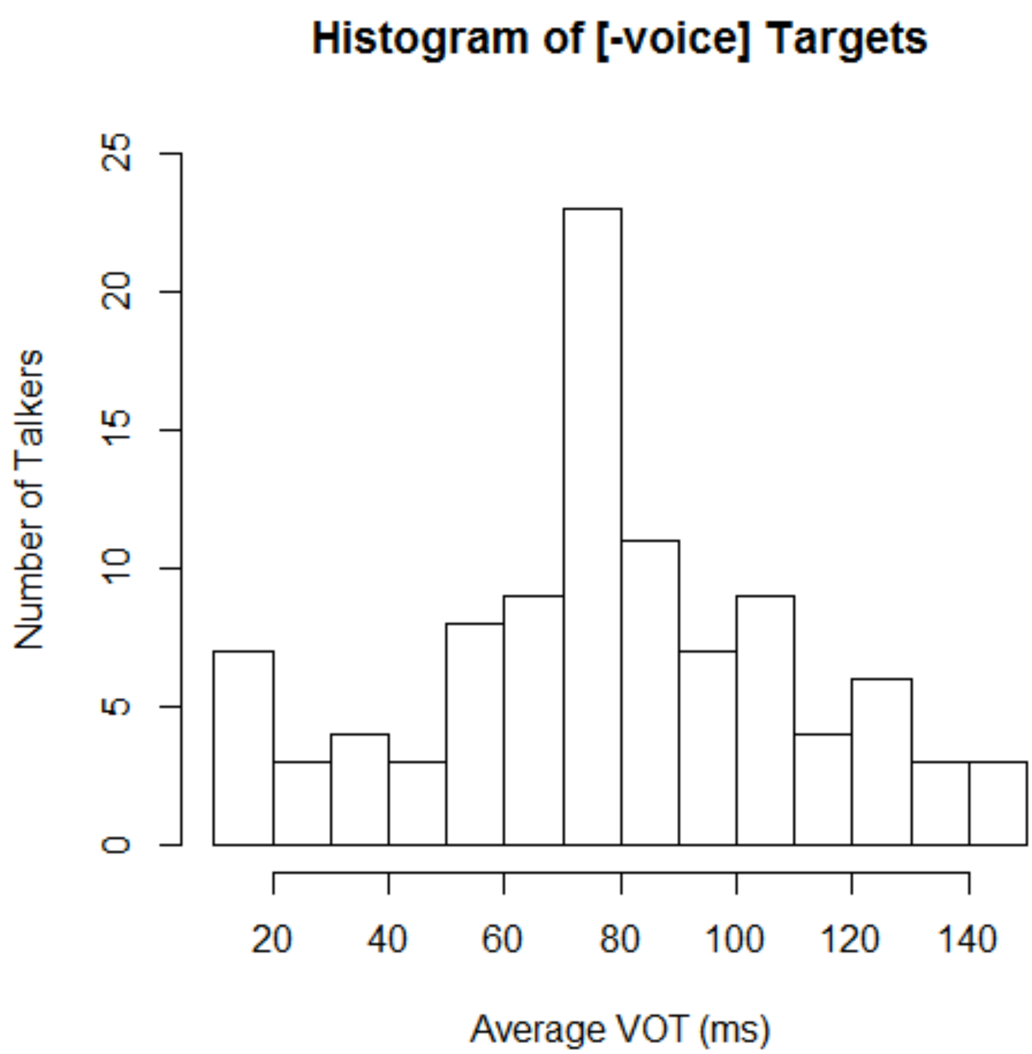


Figure 4: ***Histogram of [+voice] stop targets***

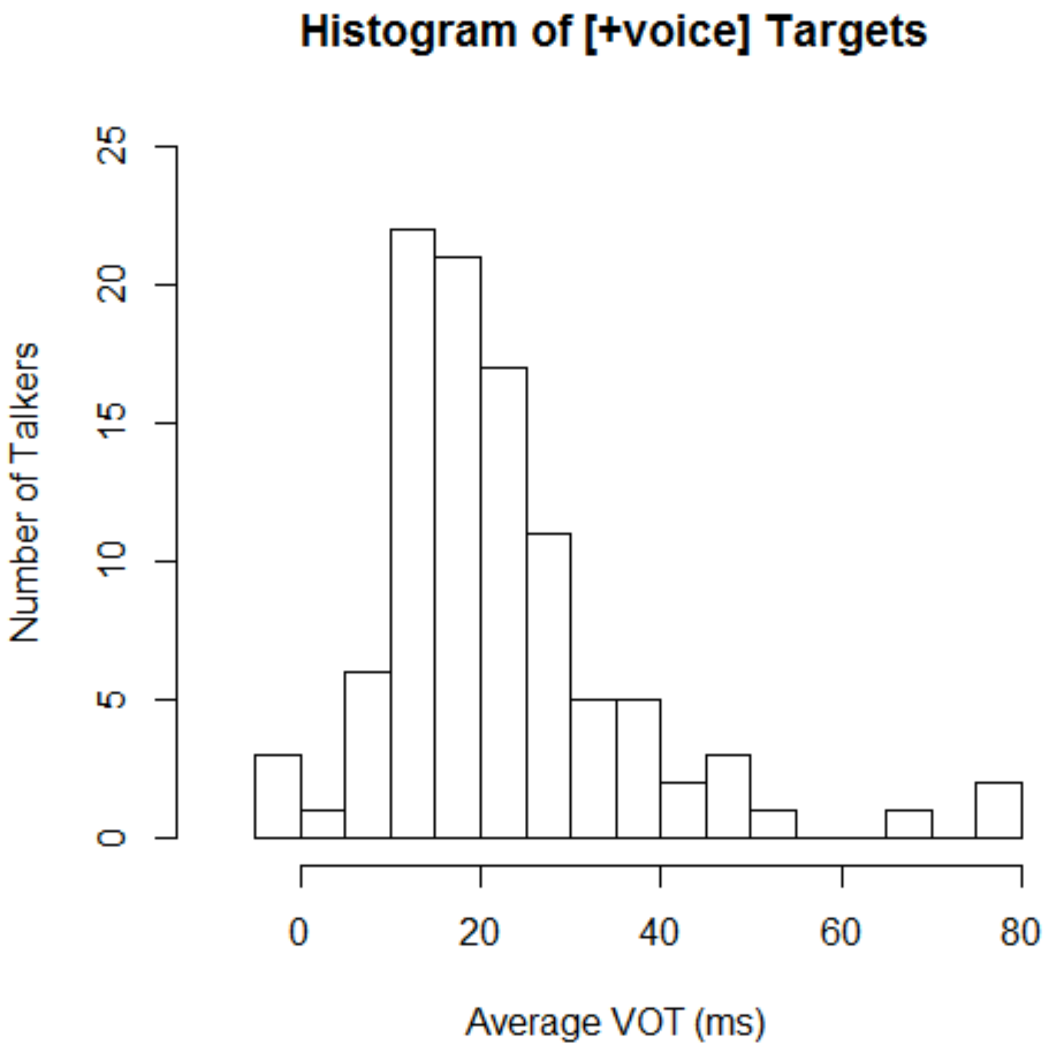


Table 2: ***Range of individual differences measures***

Measure	**Mean**	**Standard Deviation**	**Range**
Age (Months)	32.5	3.5	28-39
Sex (Proportion Female)	0.53	*NA*	*NA*
Goldman-Fristoe Test of Articulation-2 Standard Score	91	15	61-119
Expressive Vocabulary Test-2 Growth Value Score	116	14	81-148
Peabody Picture Vocabulary Test-4 Growth Value Score	103	18	70-151
Inhibitory Control (Fruit Stroop Task, possible range 0-3)	2.07	0.74	0.89-3
Minimal Pair Identification, proportion correct	0.68	0.17	0.07-0.98
Voice Onset Time, voiced targets (ms)	22	13	-3-76
Voice Onset Time, voiceless targets (ms)	78	31	10-144
Robustness of voicing contrast (logistic regression slopes)	0.07	0.04	-0.14-0.01

3.2 Pearson correlations

Pearson correlations among these individual differences measures, or, *predictor variables*, (all indexing some component of language ability) were strongly significant, with p-values well below the uncorrected a-level of 0.05 and the Bonferroni-corrected a-level of 0.001. This was true both with age included as a predictor (Table 3) and when considering partial correlations, where age had been controlled statistically (Table 4). There were two notable exceptions: non-significant correlations were observed between age and GFTA scores (Pearson's r = 0.093, p = 0.370), and age and Fruit Stroop scores (Pearson's r = 0.158, p = 0.123).

Table 3: ***Correlations among predictor variables***

	ROC[a]	Age[b]	GFTA-2[c]	EVT-2[d]	PPVT-4[e]	Minimal Pair[f]	Inhibitory Control[g]
ROC		0.304**	0.447**	0.348**	0.283**	0.289**	0.141
Age	0.304**		0.093	0.322**	0.387**	0.240*	0.158
GFTA	0.447**	0.093		0.382**	0.428**	0.361**	0.286**
EVT-2	0.348**	0.322**	0.382**		0.682**	0.361**	0.475
PPVT-4	0.283**	0.387**	0.428**	0.682**		0.471**	0.341**
Minimal Pair ID	0.289**	0.240*	0.361**	0.361**	0.341**		0.184
Inhibitory Control	0.141	0.158	0.286**	0.475**	0.471**	0.184	

[a]Robustness of Contrast in Voicing (individual-subjects' slopes), [b]Age (in months), [c]Goldman-Fristoe Test of Articulation – 2, [d]Expressive Vocabulary Test – 2, [e]Peabody Picture Vocabulary Test – 2, [f]Minimal Pair Discrimination Task, [g]Fruit Stroop Task

**p<0.01, *0.01<p<0.05

Table 4: ***Partial Correlations - controlling for age***

	ROC[a]	GFTA-2[b]	EVT-2[c]	PPVT-4[d]	Minimal Pair[e]
ROC		0.442**	0.278**	0.189	0.231*
GFTA	0.442**		0.374**	0.427**	0.351**
EVT-2	0.278**	0.374**		0.638**	0.310**
PPVT-4	0.189	0.427**	0.638**		0.280**
Minimal Pair ID	0.231*	0.351**	0.310**	0.280**	

[a]Robustness of Contrast in Voicing (individual-subjects' slopes), [b]Goldman-Fristoe Test of Articulation – 2, [c]Expressive Vocabulary Test – 2, [d]Peabody Picture Vocabulary Test – 2, [e]Minimal Pair Discrimination Task

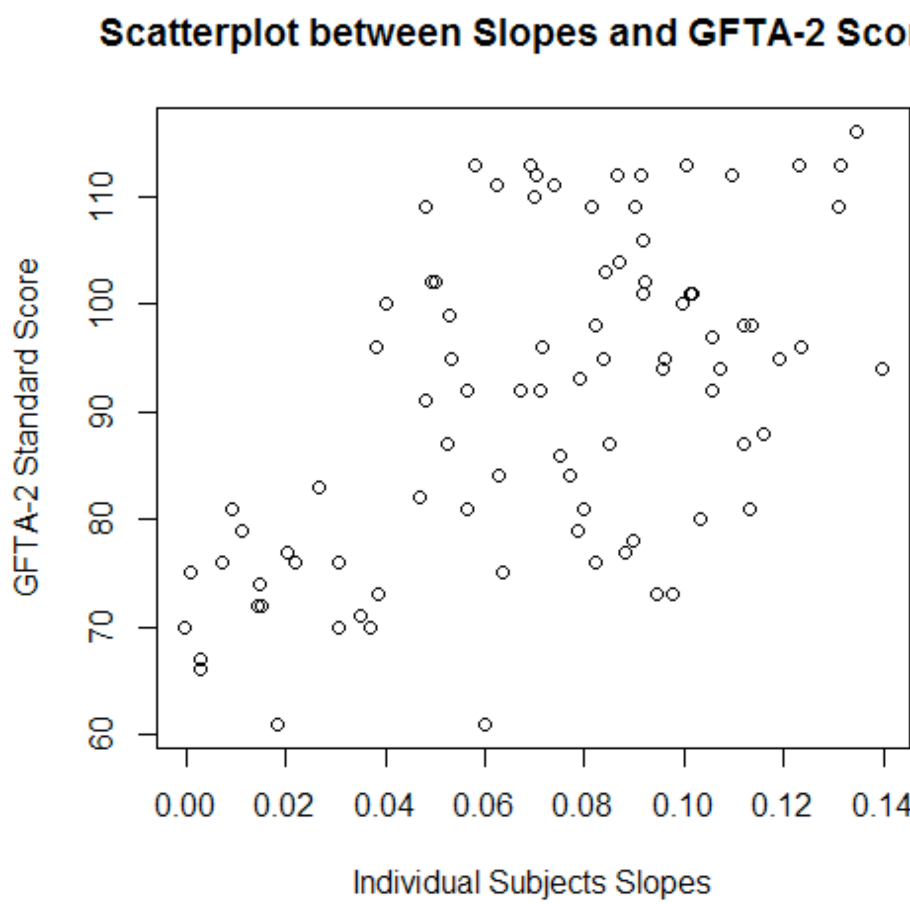
**p<0.01, *0.01<p<0.05

3.3 Robustness of contrast

As was described in the *Introduction* of this paper, the term *Robustness of Contrast* (ROC) is used to refer to the individual-subjects' slopes determined via a mixed-model logistic regression, which corresponds to the extent to which VOTs were differentiated by individual subjects for voiced and voiceless tokens. This is the summary measure of ROC used in this thesis, as in previous research by Bernstein (2015) and Holliday et al. (2015), among others. As Holliday et al. found these measures of slope to have a positive correlation with age and vocabulary size, so did this study find significant correlations between ROC and the measured predictor variables.

ROC (aka "Individual-Subjects' Slopes") was found to correlate positively with all measures of individual differences, most significantly with GFTA-2 scores (Pearson's r = 0.447, p = 0.000), EVT-2 scores (Pearson's r = 0.348, p = 0.001), and Minimum Pair Identification task scores (Pearson's r = 0.289, p = 0.004). (See Figure 5 for a scatterplot between individual-subjects' slopes and GFTA-2 scores.) ROC did not correlate significantly with Fruit Stroop scores (Pearson's r = 0.141, p = 0.171).

Figure 5: ***Scatterplot between Individual-Subjects' Slopes and GFTA-2 scores***



Additionally, a wide range in measured slopes was observed, suggesting a similarly large range in ROC, which indicates no restriction in range (Figure 6).

Figure 6: ***Histogram of children's range of regression slopes***

Slopes were determined by using a mixed-model logistic regression model where the target voicing was associated with 0 for voicing and 1 for voiceless (extending on the y-axis) and was plotted against VOT (in ms) on the x-axis. The large range of measured slopes is exemplified by the following three participants (Figures 7, 8, & 9):

Figure 7: ***Participant s612 - highly overlapping voicing categories leads to a shallow slope, which is associated with weak ROC***

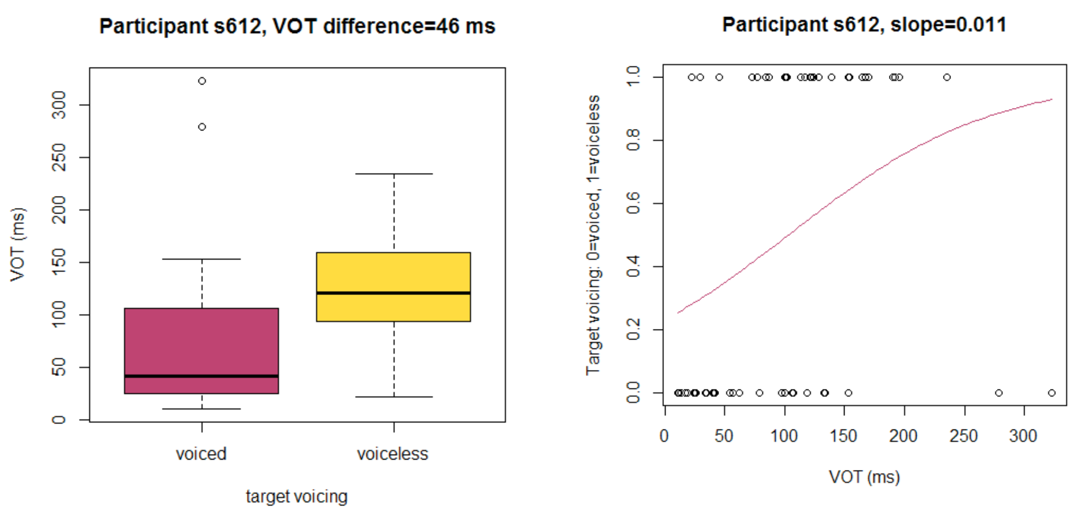


Figure 8: ***Participant s036 - moderately differentiated voicing categories leads to a moderately steep slope, which is associated with moderate ROC***

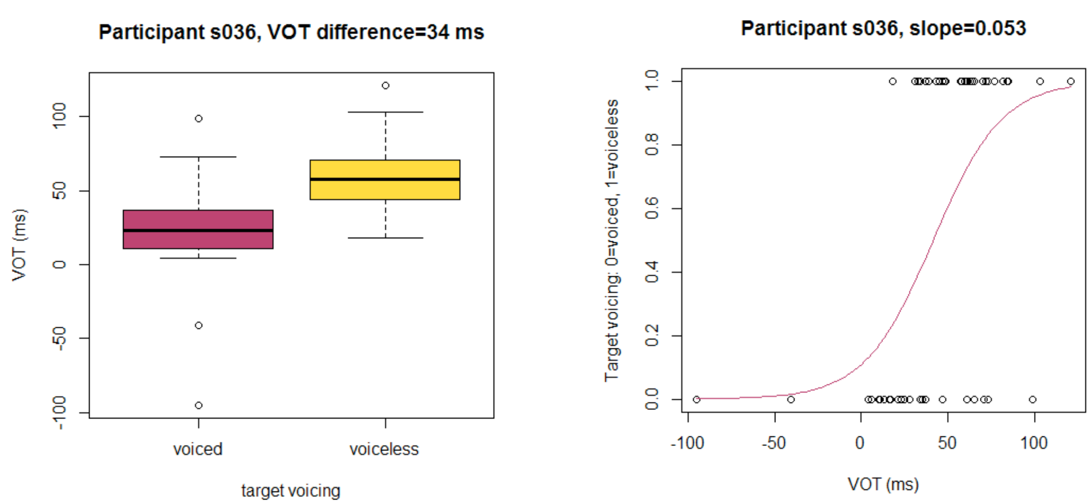
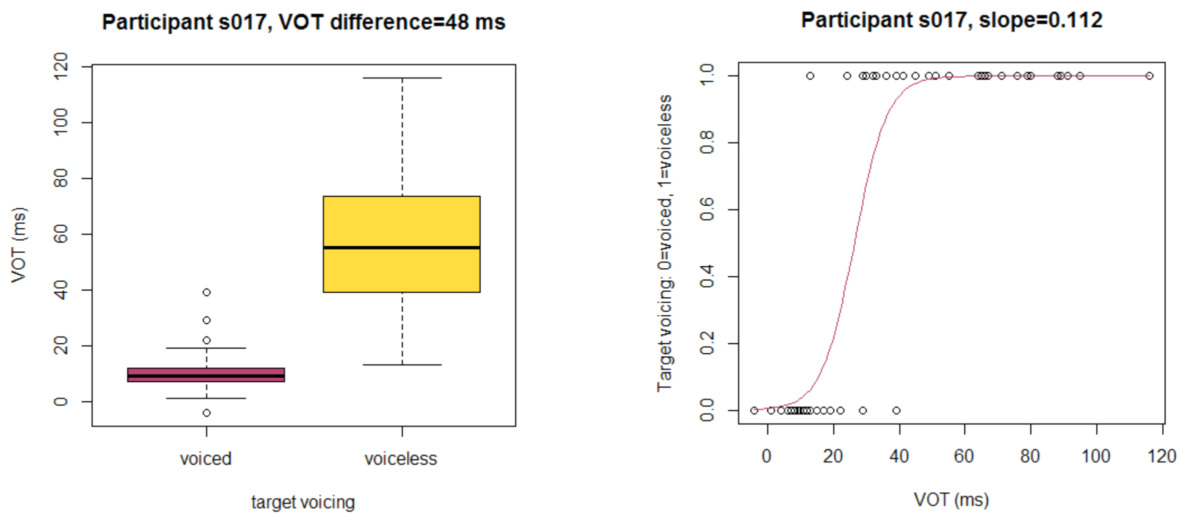


Figure 9: ***Participant s017 - clearly differentiated voicing categories leads to a very steep slope, which is associated with great ROC***



4 Discussion

The first aim of this study was to investigate the potential use of robustness of voicing contrast as an objective measure of the acquisition of voicing. This, in turn, could lead to better protocols for assessing normal phonological development and to better diagnosis of speech sound disorders in children. Previous studies attempting to determine when children fully acquire voicing contrast found great variability in their results (e.g., Hitchcock, 2005; Hitchcock & Koenig, 2004; Lowenstein & Nittrouer, 2008; Smit, Hand, Freilinger, Bernthal, & Bird 1990). In this study, acquisition of voicing was measured by using a mixed-model logistic regression (based on Holliday et al., 2015) predicting target consonant voicing from VOT. This was used to determine how robustly the participants contrasted target voiced and voiceless contexts, which we termed the robustness of contrast (ROC) of voicing. The results of this logistic regression were a wide range of individual-subjects' slopes for the functions predicting target-consonant voicing from VOT. These were used as measures of the ROC for voicing and showed a wide range of mastery of the voicing contrast. While previous studies have categorized children as either producing a voicing contrast or not (even when the contrast is *covert* and observed only by acoustic analysis), this is the first study to examine individual differences in the acquisition of voicing in a large cohort of young children with a variety of language levels. Since the results of this study found a range of ROC in the productions of voiced and voiceless stops, it can be concluded that the range represented a spectrum of the development of the voicing contrast. The participants with very robust contrasts, as indicated by steep slopes (e.g., participant s017 with slope = 0.112) demonstrated having a very advanced production of voicing contrast. This differs from the participants who produced very weak contrasts (e.g., participant s612 with slope = 0.011) whose productions show that they have not yet acquired a fully adult-like voicing contrast. Those participants with intermediately robust contrasts (e.g., participant s036 with slope = 0.053) demonstrated being in the process of acquiring the voicing contrast. Thus, the findings of the current study indicate that robustness of voicing contrast is a viable way of determining whether or not a child has fully acquired the voicing contrast. Because this study examined only children, we cannot conclude to what extent the most-advanced children in this study produced true adult-like voicing contrasts.

The second aim of this study was to compare the ROC measure in the stop production of children to the children's corresponding speech and language assessment scores, in order to investigate whether a more robust voicing contrast could be indicative of greater later language skills. Since previous research has found a correlation between vocabulary size and phonological skills (e.g., Edwards, Beckman & Munson, 2004; Munson, Edwards, & Beckman, 2005; Munson, Kurtz, & Windsor, 2005; Sosa & Stoel-Gammon, 2012; Stoel-Gammon, 1991), it was hypothesized that a more robust voicing contrast (indicative of stronger phonological skills) would correlate with stronger speech and language skills. This hypothesis was supported by the findings of the current study.

ROC was found to correlate with all measures of individual differences, both measures of speech (i.e., GFTA-2) and language (i.e., EVT-2, PPVT-2, and Minimal Pair Discrimination task). Additionally, since Pearson correlations among the individual differences were so strongly significant, it seems reasonable to conclude that the individual differences measures were all reflective of different components of the same overall communication skill set. Thus, for there to be a correlation between these individual differences measures and ROC suggests that ROC, too, is a component of a child's overall communication skill set. The presence of a strong correlation between ROC and the individual differences measures also suggests that ROC could indeed be indicative of future language skills.

4.1 Contributions to the literature

This current study has added to the existing literature investigating the age of acquisition of the voicing contrast. Unlike previous studies, which only separated children who demonstrated a voicing contrast from those who did not, this study attempted to quantify the degree of voicing contrast (via the *Robustness of Contrast* measure) to better describe the *development* of the voicing contrast. The fact that the

ROC measure is a continuous measure means that subtle developmental changes can be better tracked as a child acquires the voicing contrast. Additionally, this study contributed to the growing body of evidence that phonological and language skills are interrelated. While previous research has shown a correlation between vocabulary size and phonological skills, minimal research has been done to investigate the reverse relationship. This study, however, did investigate how phonological skills might correlate with later language skills.

4.2 Limitations

One limitation of this study is that it represents just one time point in a longitudinal study. Another limitation was a lack of item-by-item transcriptions from the GFTA-2 assessment, which would have revealed which children had frank voicing errors and which did not. Additionally, this study did not attempt to clarify whether there were any other parameters that the children used to contrast voicing besides VOT. If a perception study (where adult listeners rated children's productions) were conducted, it could provide more information on how exactly children contrast voicing. For example, if adult listeners could discern voiced from voiceless tokens produced by children with a weak ROC for voicing, it would indicate that the children were using cues other than ROC for voicing.

4.3 Future studies

It is important to consider that this study only investigated the development of the English voicing contrast. As Kong, Beckman, and Edwards (2012) note, the age of development of VOT can vary depending on the language, so VOT should not be the only component of voicing acquisition that is considered. Rather, other language-specific acoustic measurements should supplement VOT. Future studies could consider the effect different languages have on the acquisition of VOT and determine what other components of the voicing acquisition should supplement VOT when investigating its development.

Additionally, while the current study examined the relationship between phonological skills and language skills, this research is in its infancy and could benefit from further investigation in future studies. Since the question at hand is prospective in nature (i.e., investigating how a measure of phonological skills at Timepoint X will relate to a measure of language skills at Timepoint Y), there is much research to be done to see how later language skills actually do (if at all) correlate with early phonological skills. This future research will be essential in developing a clinical application. If early phonological skills (e.g., ROC) are found to be strongly correlated with later language skills, assessing phonology could be used as a means of determining children with high-risk for later language disorders. Those children could then be provided with early intervention to pro-actively address their risk of future language disorder.

5 Bibliography

Beckman, M. E., & Edwards, J. (2000). The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. *Child development*, *71*(1), 240-249.

Bernstein, S. (2015). Individual differences in the acquisition of the /t/ - /k/ contrast: A study of adult's perception of children's speech. University of Minnesota department of Speech-Language-Hearing Sciences. Master of Arts, 73.

Dunn, L. M., & Dunn, D. M. (2007). Peabody Picture Vocabulary Test (4th ed.). Minnesota: Pearson

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47*(2), 421-436.

Edwards, J., Fox, R. A., & Rogers, C. L. (2002). Final Consonant Discrimination in Children: Effects of Phonological Disorder, Vocabulary Size, and Articulatory Accuracy. *Journal of Speech, Language, and Hearing Research*, *45*(2), 231-242.

Edwards, J., Gibbon, F., & Fourakis, M. (1997). On Discrete Changes in the Acquisition of the Alveolar/Velar Stop Consonant Contrast. *Language and Speech*, *40*(2), 203.

Eimas, P., Siqueland, E., Jusczyk, P., & Vigorito, J. (1971). Speech Perception in Infants. *Science, 171*(3968), 303-306.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. H., Reznick, J. S., & Bates, E. (2007). MacArthur-Bates Communicative Development Inventories: User's guide and technical manual (2nd ed.). Maryland: Paul H. Brookes Publishing Co.

Forrest, K., Weismer, G., Elbert, M., & Dinnsen, D. A. (1994). Spectral analysis of 58 target-appropriate/t/and/k/produced by phonologically disordered and normally articulating children. *Clinical linguistics & phonetics, 8*(4), 267-281.

Forrest, K., Weismer, G., Hodge, M., Dinnsen, D. A., & Elbert, M. (1990). Statistical analysis of word-initial/k/and/t/produced by normal and phonologically disordered children. *Clinical Linguistics & Phonetics*, *4*(4), 327-340.

Gibbon, F. (1990). Lingual activity in two speech-disordered children's attempts to produce velar and alveolar stop consonants: evidence from electropalatographic

(EPG) data. International Journal of Language & Communication Disorders, 25(3), 329-340.

Gierut, J. A., & Dinnsen, D. A. (1986). On word-initial voicing: converging sources of evidence in phonologically disordered speech. *Language & Speech*, *29*(2), 97-114.

Gioia, G.A., Espy, K.A., & Isquith, P. K. (2003). Behavior Rating Inventory of Executive Function–Preschool Version: Professional manual. Florida: PAR.

Goldman, R. & Fristoe, M. (2000). Goldman-Fristoe Test of Articulation (2nd ed.). Minnesota: American Guidance Service, Inc.

Hammarström, I. L., Larsson, M., Wiman, S., & McAllister, A. (2012). Voice onset time in Swedish children and adults. *Logopedics, Phoniatrics, Vocology*, *37*(3), 117-122.

Hitchcock, E. R. (2005). *Acquisition of contrastive voicing in typically developing American English-speaking children*.

Hitchcock, E. R. & Koenig, L. L. (2004). Contrastive voicing acquisition in 2-year-old children: Preliminary data. *The Journal of the Acoustical Society of America, 115*, 2466.

Hitchcock, E. R., & Koenig, L. L. (2013). The effects of data reduction in determining the schedule of voicing acquisition in young children. *Journal of Speech, Language, and Hearing Research*, *56*(2), 441-457.

Hitchcock, E. R., & Koenig, L. L. (2015). Longitudinal observations of typical English voicing acquisition in a 2-year-old child: Stability of the contrast and considerations for clinical assessment. *Clinical linguistics & phonetics*, *29*(12), 955-976.

Holliday, J. J., Reidy P. F., Beckman M. E., & Edwards J. (2015). Quantifying the robustness of the English sibilant fricative contrast in children. *Journal of Speech, Language, and Hearing Research. 58*(3): 622–637.

Jakobson, R. (1941). 1968. *Child language, aphasia and phonological universals*.

Kong, E. J., Beckman, M. E., & Edwards, J. (2012). Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *Journal of phonetics*, *40*(6), 725-744.

Kramer, C. (2016). Predictors of early sibilant fricative production as evidenced by naive listener perception ratings. University of Minnesota department of Speech-Language-Hearing Sciences. Bachelor of Arts, submitted.

Li, F. (2012). Language-specific developmental differences in speech production: A cross-language acoustic study. Child Development. 83(4)

Li, F., Edwards, J., & Beckman, M. E. (2009). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics*, *37*(1), 111-124.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384-422.

Lowenstein, J. H., & Nittrouer, S. (2008). Patterns of acquisition of native voice onset time in English-learning children. *The Journal of the Acoustical Society of America*, *124*(2), 1180–1191.

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language*, *7*(01), 41-74.

McCune, L., & Vihman, M. (2001). Early Phonetic and Lexical Development: A Productivity Approach. *Journal of Speech, Language, and Hearing Research,44*(3), 670-684.

Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, and Hearing Research*,*44*(4), 778-792.

Munson, B., Edwards, J., & Beckman, M. E. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language, and Hearing Research*, *48*(1), 61-78.

Munson, B., Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, *48*(5), 1033-1047.

Nittrouer, S. (1993). The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. *Journal Of Speech & Hearing Research*, *36*(5), 959.

Nittrouer, S., Estee, S., Lowenstein, J. H., & Smith, J. (2005). The emergence of mature gestural patterns in the production of voiceless and voiced word-final stops. *The Journal of the Acoustical Society of America, 117*, 351.

Rvachew, S., & Grawburg, M. (2006). Correlates of phonological awareness in preschoolers with speech sound disorders. *Journal of Speech, Language, and Hearing Research*, *49*(1), 74-87.

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders, 55*(4), 779-798.

Stoel-Gammon, C. (1991). Normal and disordered phonology in two-year-olds. *Topics in language disorders, 11*(4), 21-32.

Sosa, A. V., & Stoel-Gammon, C. (2012). Lexical and phonological effects in early word production. *Journal of Speech, Language, and Hearing Research*,*55*(2), 596-608.

Storkel, H., & Morrisette, M. (2002). The Lexicon and Phonology: Interactions in Language Acquisition. *Language, Speech, and Hearing Services in Schools,33*(1), 24-37.

Tyler, A., Figurski, G., & Langsdale, T. (1993). Relationships Between Acoustically Determined Knowledge of Stop Place and Voicing Contrasts and Phonological Treatment Progress. *Journal of Speech Language and Hearing Research, 36*(4), 746.

Williams, K.T. (2007). Expressive Vocabulary Test (2nd ed.). Minnesota: Pearson.

Zanobini, M., Viterbori, P., & Saraceno, F.. (2012). Phonology and Language Development in Italian Children: An Analysis of Production and Accuracy. *Journal of Speech, Language, and Hearing Research, 55*(1), 16-31.