Vowels and Consonants: The Relative Effect of Speech Sound Errors on Intelligibility

by

Kaitlin Michelle Mackie

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

In

Speech-Language Pathology

Department of Communication Sciences and Disorders
University of Alberta

ABSTRACT

Although speech-language pathologists (SLPs) have access to a wealth of information to guide the selection and prioritization of targets for intervention with children who have speech sound disorders (SSDs), empirical evidence needed to support such decisions is lacking. The role of vowels is often neglected in the literature (Davis & MacNeilage, 1990; Pollock & Keiser, 1990) and yet, an understanding of vowels may help to recognize the full picture of intelligibility for children with SSDs. The present study is a partial replication of an unpublished investigation by Vaughn and Pollock (1997). The present study aims to determine if: i) there is a significant difference between the effect of vowel and consonant error patterns on intelligibility; ii) there is a significant difference between individual error patterns on intelligibility regardless of vowel or consonant status. The present study differs from the previous unpublished study in that the target items were controlled for frequency and phonological density. Furthermore, speech production of words were recorded from a child's speech rather than using computer generated speech. Participants in the present study listened to the recordings of a child saying real English words with and without specific vowel and consonant errors. Adult listeners were asked to type out the real English word that they believed the child was trying to say. Percent accuracy for each of the error categories (i.e. correct, vowel errors, consonant errors, and combined errors) or individual error patterns (e.g., Tensing, Stopping) was used as a measure of intelligibility. Analysis showed no significant differences between vowel and consonant error categories. Only one of the individual error patterns, Prevocalic Voicing, was significantly different from the 5 other error patterns. Post-hoc analysis of the joint effect of word position and individual error pattern suggested that different error patterns may affect intelligibility uniquely as a function of distinctive word positions. These data provide the evidence-based support needed to encourage clinicians to investigate vowel errors more closely and consider selecting them as targets in the remediation of SSDs.

PREFACE

This thesis is an original work by Kaitlin Mackie. The research project, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board, Project Name “Predicting the Impact of Speech Sound Errors on Intelligibility”, No. Pro00042855, February 21, 2014.

DEDICATION

To the Department of Communication Sciences and Disorders and the Department of Linguistics at the University of Alberta. Together they have been my home and extended family for the past 5 years. At every step I have been challenged and encouraged to learn more than I ever thought possible. The people in each of these departments have shaped me into both the academic and clinician I am today.

This is especially dedicated to my supervisor, Karen Pollock PhD, who has guided my project with enthusiasm and careful consideration since our very first meeting.

Finally, this is dedicated to my fiancé who, along with supporting me for the past 10 years, volunteered his knowledge of programming to help me finish this project.

ACKNOWLEDGEMENTS

Figures from “The Relative Contribution of Vowel and Consonant Errors to Intelligibility” by A. Vaughn & K. E. Pollock, 1997, Unpublished manuscript, p. 3, were reprinted with permission.

Part of this research was completed at the Alberta Phonetics Laboratory

APhL
Alberta Phonetics Laboratory

This research was supported by the Social Sciences and Humanities Research Council of Canada.

Social Sciences and Humanities
Research Council of Canada

Conseil de recherches en
sciences humaines du Canada

Canada

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Chapter 1: Introduction

Although speech-language pathologists (SLPs) have access to a wealth of information to guide the selection of targets for intervention with children with speech sound disorders (SSDs), the information often lacks empirical evidence. This leaves SLPs in a position where they must decide which targets to select without evidence-based information. Targeting sound error patterns that negatively affect intelligibility is a common recommendation, but knowledge of the relative effects of different types of errors on intelligibility is lacking. In particular, one area which lacks information that could guide target selection is impact of vowel errors on intelligibility. Although vowel errors are less common than consonant errors, their presence often signals a relatively severe diagnosis of SSD (Pollock & Berni, 2003), and thus they may be considered high priority for intervention. The present study aims to understand how vowels and consonants affect intelligibility using a partial replication of an unpublished study which showed vowel and consonant errors had equal impact on intelligibility (Vaughn & Pollock, 1997). Greater empirical data could provide guidance for selecting targets that have the most functional impact on a child's speech and maximize clinical effectiveness.

Western Canadian Vowel System

In Western Canadian English, there are 13 non-rhotic vowel phonemes, 10 of which are phonemic monophthongs: /i/ /ɪ/ /e/ /ɛ/ /æ/ /ɑ/ /o/ /ʊ/ /u/ and /ʌ/. Phonemic diphthongs in Western Canadian English include /a͡ɪ/, /a͡ʊ/, and /ɔ͡ɪ/. Like many other forms of North American English, Western Canadian English also contains rhotic vowels: the monophthong /ɝ/ and diphthongs /ɪ͡ɚ/, /ɛ͡ɚ/, /ɔ͡ɚ/, and /ɑ͡ɚ/ (which are also sometimes described as postvocalic and transcribed with a final consonant r: /ɪɹ,/, /ɛɹ/, /ɔɹ/, and /ɑɹ/). Some Western Canadian English monophthongs behave phonetically like diphthongs (e.g., /o/ is produced as [o͡ʊ] and /e/ as [e͡ɪ]), especially in stressed and open syllables. Vowels in unstressed syllables are reduced in English, and typically produced as a [ə] (non-rhotic), [ɚ] (rhotic), or another non-peripheral vowel.

Speech Sound Development

In typical development, vowels are generally mastered at an early age, before consonants. Non-rhotic monophthongs and diphthongs are reportedly mastered between 2 to 3 years of age (Templin, 1957; Pollock, 2002; Pollock & Berni, 2001; Pollock, 2013; Irwin & Wong, 1983; Donegan, 2013). Vowels within the corner regions of the vowel quadrilateral (e.g., /i/ /ʊ/ /ɑ/) seem to have the highest accuracy in early development (Paschall, 1983; Stoel-Gammon & Herrington, 1990; Otomo & Stoel-Gammon, 1992). Early consonants (e.g., /m/, /p/, and /b/) are mastered around age 3 (Sander, 1972; Irwin & Wong, 1983; Smith, 1973), around the same time as mastery of all English non-rhoticized vowels. The last vowels to be mastered are rhotic vowels sometime after 4 years of age (Stoel-Gammon & Herrington, 1990; Pollock, 2002; Pollock & Berni, 2001; Pollock, 2013). Late developing consonants (i.e., /θ/ and /ð/) and consonant clusters are often acquired much later, sometime after age 7 (Irwin & Wong, 1983; Sander, 1972; Stoel-Gammon & Dunn, 1985).

In addition to the typical age of acquisition of individual speech sounds, common patterns of errors used by children are also well documented. Prevocalic Voicing, Word Final Devoicing, Final Consonant Deletion, Velar and Palatal Fronting, Consonant Harmony, and Stopping are common consonant error patterns used by typically developing children. These six patterns are largely phased out by age 4 (Bowen, 1998). In contrast, error patterns such as Cluster Reduction and Gliding continue after age 4 (Bowen, 1998) but generally resolve around age 6 (Shriberg, 1997). Children with SSDs typically exhibit these same patterns but may also use unique error patterns (Stoel-Gammon & Dunn, 1985; see also Bowen, 2009).

Common vowel error patterns in typical development have received relatively less attention. Otomo and Stoel-Gammon (1992) suggest that patterns for typical children may include Raising, Lowering, and Backing depending on which vowels are being acquired. In children with SSDs patterns such as Diphthong Reduction, and Lowering are common (Pollock & Keiser, 1990; Reynolds, 1990). Depending on the dialect, it appears as though Vowel Fronting (Reynolds, 1990) or Backing (Pollock & Keiser, 1990) can also be common. Pollock (2013) noted that although children with SSDs often made the same types of vowel errors as typically developing children, they were more systematic in their application of error patterns. When children who use vowel errors had larger vocabularies and complex words there was a chronological mismatch. Furthermore, vowel error patterns exhibited by children with SSDs can be highly idiosyncratic and still variable in their application (Reynolds, 1990).

Children who produce incorrect phonetic productions for target phonemes for longer than is developmentally appropriate are considered to have speech sound disorders (SSDs; Bowen, 2009). SSDs include Childhood Apraxia of Speech (CAS), dysarthria, phonological disorders, articulation disorders, and speech sound errors secondary to hearing loss or other aetiologies (Bowen, 2009). Children with SSDs differ from typically developing children in the order of acquisition of speech sounds or the timing of acquisition of the speech sounds. Intelligibility is a primary concern for all SSDs. While this paper focuses on children with functional phonological disorders (i.e., children with SSDs in the absence of neurological damage or anatomical/structural deformities and whose speech sound errors have identifiable patterns), there is a small body of literature on the impact of vowels and consonants on intelligibility for children with other SSDs (i.e., hearing loss, dysarthria, and CAS). For example, strong associations between intelligibility and accuracy of corner vowels exist for children with hearing loss and CAS (Monsen, 1976; Pollock & Keiser, 1990; Walton & Pollock, 1993; Davis, Jacks & Marquardt, 2005).

Unlike typically developing children, children with SSDs do not always master vowels prior to consonants. The correlations between percentage of consonants correct and percentage of vowels correct have been investigated and were found to be relatively strong for children with typical development but only moderate for children with phonological disorders (Pollock, 2013). Roughly 50% of children with severe consonant errors had not yet mastered vowel production although vowel accuracy scores were higher than consonant accuracy (Pollock & Berni, 2003). Some children with severe CAS have roughly equal measures of consonant accuracy and vowel accuracy even between the ages of 8 and 10 years (Pollock & Hall, 1991; Walton & Pollock, 1993; Davis, Jacks & Marquardt, 2005). Although the underlying deficit in CAS is motor planning, studies show a tendency towards patterns of Backing, Lowering, and Diphthong Reduction (Pollock & Berni, 2003), the same patterns seen in children with phonological disorders.

Target Selection

The tendency to emphasize consonants over vowels during assessment is an underlying bias that is infrequently discussed (Davis & MacNeilage, 1990; Pollock & Keiser, 1990). This preference is consistent across different populations such as CAS (Gibbon, 2009). Gibbon (2009) suggests that the acceptable variability in vowel production due to dialectal differences as well as the difficulty of perceptual categorization of vowels compared to consonants may be contributing factors. Vowel errors are anecdotally less frequently occurring for children with SSDs (Watts, 2004) and thus SLPs may not anticipate vowel errors. However, when identified, the presence of vowel errors does typically indicate more severe diagnoses (Pollock, 2013), as discussed previously. Furthermore, vowels are traditionally more difficult to transcribe due to their less discrete categorization (Pollock & Berni, 2001) and this may lead to greater difficulty identifying vowel errors for targeting. Together these possible reasons create an environment in which vowels are not investigated during typical assessment procedures which is further fuelled by tools that don't intend to investigate vowels closely.

Within the typical assessment, vowels are inadequately represented in standardized tests of articulation and phonology (Pollock, 1991). A review of the standardized tests of phonology and articulation for children with SSDs, showed that less than half of the tests were designed to evaluate vowel production (Eisenberg & Hitchcock, 2010). Only two of the eleven tests had a child produce at least a single word for each vowel in the English inventory including all diphthongs and at least one rhotacized vowel: *The Fisher-Logemann Test of Articulation* (FLTA; Fisher & Logemann, 1971) and the *Templin-Darley Test of Articulation* (TDTA; Templin & Darley, 1969). These two tests were included specifically because they considered vowels. However, they are rarely, if ever, used by clinicians in their everyday practice (Skahan, Watson & Lof, 2007). This suggests that to understand the complete phonological system of a child with vowel errors as part of their SSD, considerable additional information beyond that obtained from common standardized tests is required.

Conversely, all of the tests investigated by Eisenberg and Hitchcock (2010) are designed to evaluate consonant production. With stringent phonetic and phonemic criteria (e.g., nonharmonic singleton consonants), four of the tests included at least a single word representing word initial consonant production. The stringent criteria showed that overall the eleven tests had all of the initial consonants of English 95% of the time and all of the final consonants 71% of the time. Notably, the *Goldman Fristoe Test of Articulation 2nd ed*. (GFTA – 2; Goldman & Fristoe, 2000) does investigate all of the initial word position consonants and all final consonant sounds. The GFTA-2 is reported by 50% of clinicians to be used every time they investigate speech sound usage in children (Skahan, Watson & Lof, 2007). Overall, this suggests that information on consonant errors is more readily available with these commonly used tools and consonants are more closely investigated by clinicians in the field.

When developing an intervention plan for a child with a SSD, clinicians must often prioritize from among multiple potential targets. Factors to consider include how many word positions an error is observed in, whether the error sound is found in the child's name, and sounds the parent specifically request (McLeod & Baker, 2014). Fey (1986) highlights the baseline accuracy of a target sound as an important factor for target selection. Powell (1991) identified about 20 factors that could be considered when selecting appropriate sound targets. They included factors relating to:

- Developmental norms (e.g., *child's age, age-appropriateness of error(s), normative order of acquisition*)
- Language theories (e.g., *feature specifications, homophony, linguistic markedness and implicational relationships, morphological status of error, phonotactic constraints*)

- Articulatory and phonological considerations (e.g., *phonetic inventory, stimulability, productive phonological knowledge, ease of production, frequency of sound occurrence, type of error, severity of disorder, phonological process type, number of phonological processes*)
- Functional impact (e.g., *relevance of the sound to the child, effect of errors on intelligibility, perceptual saliency of the error*)

Clinical theories incorporate a number of these factors in unique ways. For example, traditional approaches select targets that are stimulable, early developing sounds, inconsistently produced sounds, or sounds that the child shows underlying knowledge of (Williams, 2005). Comparatively newer theories encourage treatment of sounds that may result in the largest overall change in the child's sound system (i.e., greatest broad generalization; Gierut, 2005; Gierut & Hulse, 2010) including the selection of non-stimulable sounds, later developing sounds, consistent error patterns, and sounds which the child shows limited underlying knowledge (Williams, 2005). Unfortunately, reviews of these theoretical approaches to treatment of SSDs such as Kamhi (2006) and Williams (2005) consider specific error patterns only briefly, with no discussion of vowels as part of the target selection process.

Powell (1991) did not create a hierarchy indicating which of the factors are "more" or "less" important than others. Practicing clinicians seem to place emphasis on early developing sounds and stimulable sounds over non-stimulable and later developing sounds (McLeod & Baker, 2014). The International Classification of Functioning (ICF; World Health Organisation, 2001; Howe, 2008) encourages clinicians to consider a person within their environment and what impact a given disorder may have on their everyday lives. In a survey, approximately 75% of

SLPs reported selecting speech sounds target and target phonological patterns with the largest impact on intelligibility (Brumbaugh & Smit, 2013). However, how they determine which sounds or error patterns have the greatest impact on intelligibility was not discussed.

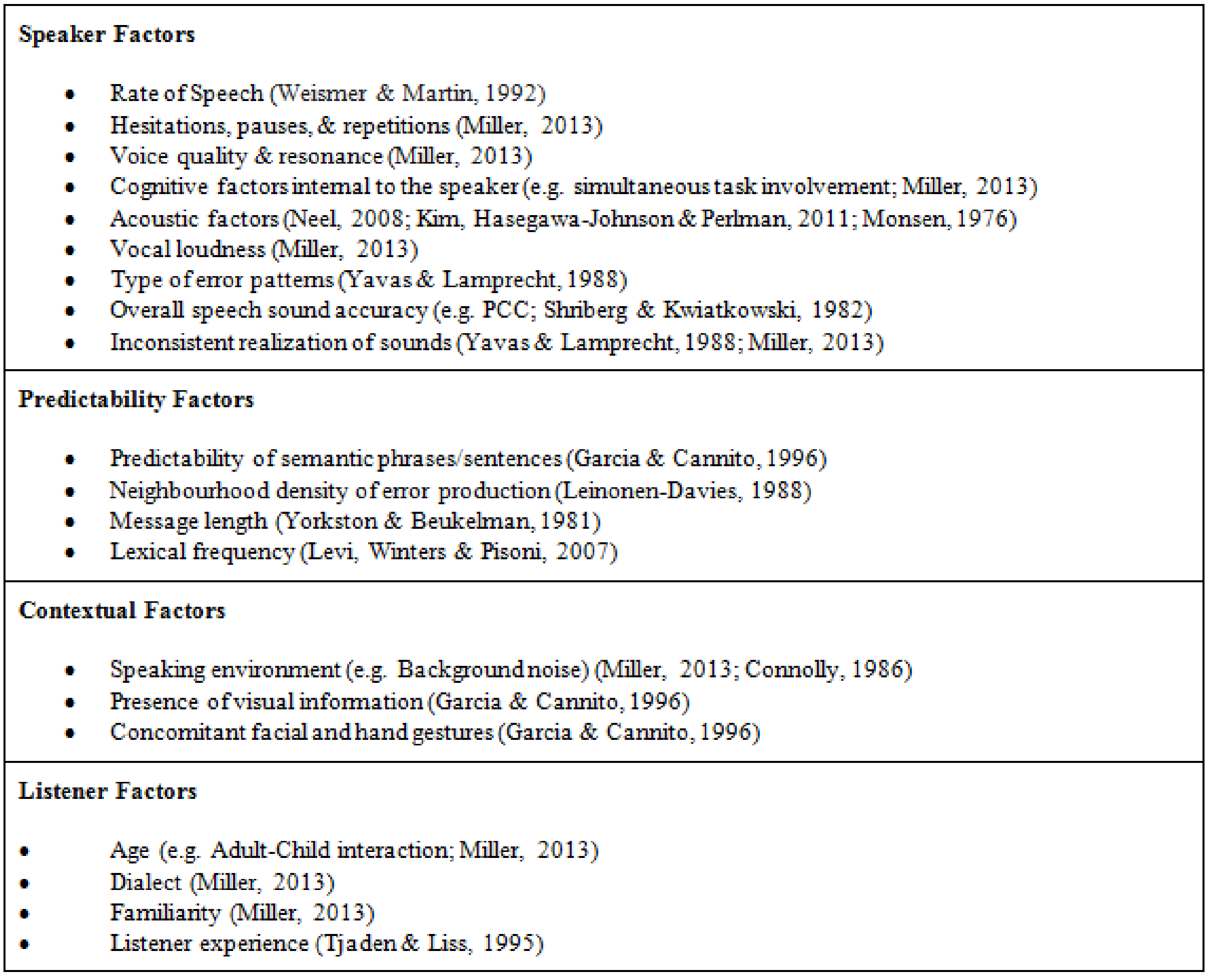
Intelligibility Measures and Factors

It is difficult to find consistent definitions across the literature for intelligibility (Dagenais, Adlington & Evans, 2011). Intelligibility has been defined as how well a listener is able to decode the intended message provided by the speaker, not necessarily the overall comprehension, although the two are related (Kent, Weismer, Kent & Rosenbek, 1989; Hustad, 2008). Miller (2013) outlines the difference between two different components of intelligibility: *Signal-dependent intelligibility,* which includes information within the communicative output of the speaker such as the acoustic signal, verbal cues, and non-verbal cues; *signal-independent intelligibility* includes information outside of communicative output such as strategies that help or hinder the speaker-listener dynamic (e.g., clear speech strategies). Intelligibility then is a variable component of communication even when considering the same speaker and listener in different contexts. Intelligible speech is generated from a complex mixture of the dependent and independent aspects ultimately resulting in the successful identification of meaningful words from the contextualized auditory signal. The present study considers speech presented in a largely decontextualized manner. Therefore, for the purposes of the present study intelligibility will be defined as how well a listener is able to decode the intended verbal message of the speaker.

Factors influencing intelligibility. The effect of speech sound errors on intelligibility is an important consideration for SLPs. However, multiple factors impact intelligibility. For the purposes of this paper, the many different factors have been grouped into broad categories (see Table 1.1). The broad categories identified here are speaker-specific speech factors (including acoustic factors), predictability factors, contextual factors, and listener factors.

Table 1.1

Factors Influencing Intelligibility



Speaker Factors

- Rate of Speech (Weismer & Martin, 1992)
- Hesitations, pauses, & repetitions (Miller, 2013)
- Voice quality & resonance (Miller, 2013)
- Cognitive factors internal to the speaker (e.g. simultaneous task involvement; Miller, 2013)
- Acoustic factors (Neel, 2008; Kim, Hasegawa-Johnson & Perlman, 2011; Monsen, 1976)
- Vocal loudness (Miller, 2013)
- Type of error patterns (Yavas & Lamprecht, 1988)
- Overall speech sound accuracy (e.g. PCC; Shriberg & Kwiatkowski, 1982)
- Inconsistent realization of sounds (Yavas & Lamprecht, 1988; Miller, 2013)

Predictability Factors

- Predictability of semantic phrases/sentences (Garcia & Cannito, 1996)
- Neighbourhood density of error production (Leinonen-Davies, 1988)
- Message length (Yorkston & Beukelman, 1981)
- Lexical frequency (Levi, Winters & Pisoni, 2007)

Contextual Factors

- Speaking environment (e.g. Background noise) (Miller, 2013; Connolly, 1986)
- Presence of visual information (Garcia & Cannito, 1996)
- Concomitant facial and hand gestures (Garcia & Cannito, 1996)

Listener Factors

- Age (e.g. Adult-Child interaction; Miller, 2013)
- Dialect (Miller, 2013)
- Familiarity (Miller, 2013)
- Listener experience (Tjaden & Liss, 1995)

Speaker factors are those which are unique to the speaker (e.g., rate of speech). Acoustic factors such as mean fundamental frequency (Neel, 2008) and vowel space area (Neel, 2008; Kim, Hasegawa-Johnson & Perlman, 2010) have been found to correlate with measures of intelligibility. Of specific interest to child development is consistency of speech sound production at a given stage during development. More consistent speech sound productions/error patterns are tied to higher levels of intelligibility (Yavas & Lamprecht, 1988).

A speaker factor, which is highly relevant to the current discussion, is error type, specifically, vowel errors compared to consonant errors. Measures of consonant production accuracy are consistently obtained during initial assessment (Skahan, Watson & Lof, 2007). One such measure which is strongly linked to intelligibility is the percentage of consonants produced correctly (PCC; Shriberg & Kwiatkowski, 1982). PCC is a measurement in which all consonant omissions, substitutions, and distortions are considered errors in child's speech. Shriberg and Kwiatkowski recommend PCC be calculated from spontaneous speech samples. PCC is a strong diagnostic tool to determine severity for children with SSDs across age groups (Shriberg, Lewis, McSweeny & Wilson, 1997a; Shriberg, Lewis, McSweeny & Wilson, 1997b). Percentage of vowels produced correctly (PVC) is also described in literature that specifically investigates vowel accuracy (e.g., Zarifian, Tehrani, Salavati, Modaresi & Kazemi, 2014). Calculation of PVC follows the same guidelines as the calculation of PCC. However, the link between PVC and intelligibility has not been investigated. There are unfortunately very few assessment tools that use PCC or PVC in spontaneous speech as an indicator of intelligibility in part due to the intensity of analysis required to analyze spontaneous speech. Yet, many standardized assessment tests (e.g., Structured Photographic Articulation Test - II; [Dawson & Tattersall, 2001]) will encourage clinicians to report a percentage of elicited consonants correct. As mentioned previously, very few standardized assessment tools investigate vowels closely and thus calculation of a percentage of elicited vowels is only encouraged in a few less popular tests. One example of a test which does encourage a percentage of elicited vowels is the Diagnostic Evaluation of Articulation and Phonology (DEAP; Dodd, Hua, Crosbie, Holm & Ozanne, 2009).

Predictability is the increased or decreased likelihood of correctly identifying linguistic identity (e.g. word identity) based on other information within the prior or co-occurring linguistic environment. For example, given a cloze sentence (e.g., It's raining cats and ______) how likely it is that a specific word (e.g., dogs) is used to complete the sentence. Another predictability factor important for phonological disorders is neighbourhood density. Phonological neighbourhood density is a measure of the number of words that differ by a single phoneme from a base word, creating a grouping of like words (Marian, Bartolotti, Chabal & Shook, 2012). Leinonen-Davies (1988) explored this area by developing Functional Loss (FLOSS) values. FLOSS quantifies the loss of phonological contrast in children's productions by "performing" common error pattern analyses on commonly produced lexical items to determine what the verbal output would be. Many of the consonant error patterns create words that would be homophonous productions for other common lexical items in a child's repertoire. Multiple patterns tended to create an even denser phonological neighbourhood and fewer distinctive words being produced.

Contextual factors include anything separate from the listener, speaker, or linguistic output, which may affect intelligibility such as background noise or visual information. Visual information could include a physical representation of the topic of discussion (e.g., discussing cooking in the kitchen with many lexical items in view).

Listener factors include anything that is unique to the listener, such as the dialect and age (Miller, 2013). If factors are congruent with the speaker, then this would likely aid intelligibility whereas incongruent social factors will likely lead to a breakdown in intelligibility. For example, studies of adults' speech generally consist of judgements from other adults of a similar age group.

Intelligibility measures. The importance of measuring intelligibility is identified consistently by practicing clinicians. According to a survey, 75% of American SLPs working with children with SSDs reported always using an estimate of intelligibility in their assessment (Skahan, Watson & Lof, 2007). In a similar survey, 55% of Australian SLPs reported always estimating intelligibility and 31% of SLPs reported sometimes estimating intelligibility (McLeod & Baker, 2014).

Measures of intelligibility require a listener's interpretation of that output. Subjective measures of intelligibility within research literature include the ranking of multiple speakers against each other (Yavas & Lamprecht, 1988) or rating scales for an individual speaker compared to a theoretical ideal of intelligibility. Rating scales efficiently estimate intelligibility and are based on the listener's subjective perception of how much of the speaker's speech they understood or how much they had to interpret the speaker's intent. Rating scales include highly subjective measures such as "ease of listening" (Landa, Pennington, Miller, Robson, Thompson & Steen, 2014). In everyday social interaction, intelligibility is evaluated subjectively by the listener. Quality of social interaction interactions is often a goal of speech therapy (Howe, 2008). Therefore rating scales could represent meaningful difficulties a speaker with a SSD has in social interactions. Rating scales or subjective impressions are likely used frequently to determine intelligibility of children's speech (Kent, Miolo & Blodel, 1994) although confirmation of which measure is used most commonly by practicing SLPs is currently unknown (Miller, 2013). Rating scales are, unfortunately, susceptible to poor inter-rater and intra-rater reliability (Miller, 2013). This often makes subjective rating scales undesirable for research purposes. However, there is evidence showing correlation between speech intelligibility rating scales and more objective measures (Landa et al., 2014), suggesting they are measuring similar constructs.

More objective measures of intelligibility generally consist of closed set and open set intelligibility tests. There are also measures incorporating rate of speech such as the number of intelligible words per minute (Yorkston, Strand & Kennedy, 1996). During closed set tests the listener is asked to identify the word or sentence produced by selecting the word they believed was produced from a set of closely related phonetic words. These types of closed set tests quickly highlight the phonetic distinctions which influence intelligibility and are also known as diagnostic intelligibility testing (Kent, Weismer, Kent & Rosenbek, 1989). Many closed set intelligibility standardized tests exist (e.g., *The Test of Children's Speech,* Hodge & Gotkze, 2011; *Children's Speech Intelligibility Measure,* Wilcox & Morris, 1999) although many of these also have open set sections.

In open set tests, a listener will typically hear the speaker's message and be required to write down the orthographic form of what they believe the speaker intended to say. The listener is often encouraged to guess any words they do not understand. Open set transcription lends itself easily to conversational and full sentence productions. To obtain an open set sample from a speaker, usually verification of oral productions is required or the speaker must reproduce a set of previously generated sentence stimuli such as those used during the sentence production of the *Speech Intelligibility Test* (SIT) (Yorkston & Beukelman, 1981). The percentage of correctly identified words is then used to quantify intelligibility. For young children, who may be unable to read, single word productions elicited from picture naming tasks are sometimes used. Open set tests generally have lower overall scores of intelligibility when compared to closed set tests (Vigouroux & Miller, 2007).

Regardless of how the factors are measured within the literature, there is a clear understanding that many factors influence intelligibility. The list of factors influencing intelligibility could become a very useful tool for SLPs if we can rank the factors according to their impact on intelligibility. That is, all other things being equal, which factors should be prioritized to maximally benefit outcomes of a client's therapy?

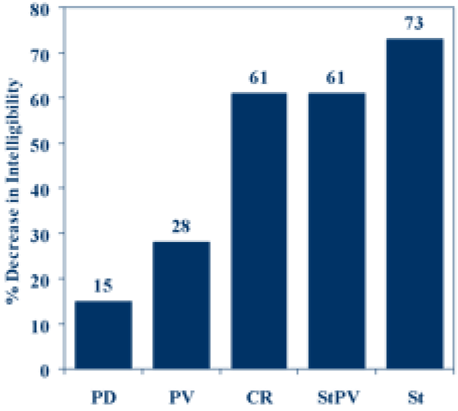
Prioritizing Speech Sound Error Patterns for Target Selection

Few known studies have attempted to predict how intelligibility is affected by specific speech sound errors. Leinonen-Davies (1988) suggested that error patterns could be rank ordered from high to low, with higher FLOSS values predicting a greater negative impact on intelligibility. Leinonen-Davies calculated the FLOSS values for 8 common error patterns and ranked their impact on intelligibility: Fronting > Gliding > Prevocalic Voicing > Stopping > Cluster Reduction in initial position and Final Consonant Deletion > Fronting > Devoicing > Stopping > and Gliding for final position. This could provide guidance for intervention when multiple consonant error patterns are present in a child's speech. However, FLOSS values are not available for vowel error patterns. For children with both consonant and vowel errors, additional information is needed to determining priority intervention targets

An unpublished study investigated the relative effect of vowel errors and consonant errors (both alone and in combination) on intelligibility (Vaughn & Pollock, 1997). In this study, participants listened to a sentence produced by a synthesized child voice with a single error in the final word in the sentence, using an open set test design (Yorkston & Beukelman 1980; Vigouroux & Miller, 2007). The final word was not predictable from the sentence context (e.g., *We discussed the house*; *He knew about the fork*). Errors included vowel error patterns, consonant error patterns, and combined error patterns. Vowel error patterns included: Laxing, (e.g., *cake* → [kɛk]), Tensing (e.g., *dish* → [diʃ], Diphthong Reduction (e.g., *clown* → [klɑn]), Backing (e.g., *bat* → [bɑt]), and Backing plus Lowering (e.g., *pen* → [pɑn]). The consonant patterns were Prevocalic Voicing (e.g., *cup* → [gʌp]), Postvocalic Devoicing (e.g., *egg* → [ɛk]), Cluster Reduction (e.g., *plate* → [pe͡ɪt]), Stopping (e.g., *seed* → [tid]), and Stopping plus Prevocalic Voicing (e.g., *fork* → [bɔ͡ɚk]). Attempts were made to include a range of error types, from those that involved relatively minor shifts (e.g., tense/lax changes and voicing changes) to those that involved multiple feature changes (e.g., vowel height and backness, manner and voicing). Target words were carefully controlled to result in phonotactically plausible English nonwords in the error condition (e.g., [kɛk] and [tid]).

Naïve undergraduate participants listened to one of two lists of 75 sentences and wrote out what they believed the word at the end of the sentence was. The number of words identified correctly was calculated as a percentage correct for the correct condition (with no error in the final word), for each error category (i.e., consonant error, vowel error, or consonant and vowel errors), and for each individual error pattern (e.g., Tensing, Stopping). To account for variability associated with synthesized speech, data were presented in terms of the percent decrease in intelligibility for each error condition, which was calculated by subtracting the percent correct in the error condition from the percent correct in the correct condition. As predicted, all error patterns resulted in a decrease in intelligibility. The combined vowel and consonant error patterns accounted for a greater loss in intelligibility (79%) than vowel or consonant errors on their own (~47%). Importantly, because there was no difference between the effect of vowel and consonant errors on intelligibility, both categories should be considered and evaluated when planning intervention.

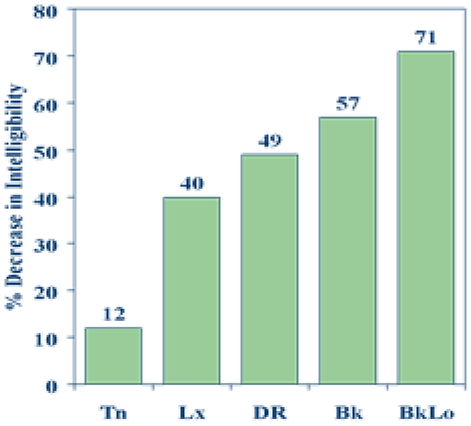
Figure 1.1. Mean Percent Decrease in Intelligibility for Consonant Error Patterns



Note. The results of Vaughn & Pollock (1997) study showing the percentage decrease in intelligibility as a result of the 5 consonant phonological processes. Reprinted with permission from "The Relative Contribution of Vowel and Consonant Errors to Intelligibility" by A. Vaughn, & K. E. Pollock, 1997, Unpublished manuscript, p. 3.

Note. PD = Postvocalic Devoicing; PV= Prevocalic Voicing; CR = Consonant Cluster Reduction; StPV = Stopping plus Prevocalic Voicing; St = Stopping

Figure 1.2. Mean Percent Decrease in Intelligibility for Vowel Error Patterns



Note. The results of Vaughn & Pollock (1997) study showing the percentage decrease in intelligibility as a result of the 5 vowel phonological processes. Reprinted with permission from "The Relative Contribution of Vowel and Consonant Errors to Intelligibility" by A. Vaughn, & K. E. Pollock, 1997, Unpublished manuscript, p.3.

Note. TN = Tensing; LX = Laxing; DR = Diphthong Reduction; BK = Backing; BkLo = Backing plus Lowering

Within each error category, however, individual error patterns accounted for varying amounts of intelligibility decrease. The amount of decrease in intelligibility compared to errorless words ranged between 12% (e.g., Tensing) and 73% (e.g., Stopping) (Figure 1.1 and 1.2). The single substitution patterns (i.e., Stopping, Consonant Cluster Reduction, Prevocalic Voicing, and Postvocalic Devoicing) do not seem to sequence themselves in the order predicted by FLOSS values (Leinonen-Davies, 1988).

Upon further inspection of the original word stimuli, it became evident that some target words existed within very sparse phonological neighbourhoods (e.g., *hotdog*) and were subjectively easier to interpret whereas other words existed in very dense phonological neighbourhoods and were subjectively harder to interpret (e.g., *cake*). Furthermore, frequency of the words was not accounted for. Both lexical frequency and neighbourhood density can affect intelligibility. If we are to measure intelligibility change as accurately as possible, it would be prudent to attempt to control these factors as much as possible. Synthesized stimuli helped to control for speaker factors such as rate of speech, and voice quality other acoustic factors, although it reduced the naturalness of the stimuli.

Further investigation and refined analyses are warranted to determine which intervention targets might maximize their effect on intelligibility. Therefore, the present study recreated the original open-set test design used by Vaughn and Pollock (1997) but also accounted for phonological density and frequency of target items and used natural speech as produced by a child. Furthermore, as words in sentence frames create variability that can influence intelligibility (Garcia & Cannito, 1996), single word stimuli were used for this stage of the investigation. The following questions were addressed:

1) Is there a significant difference in the impact of vowel error patterns compared to consonant error patterns on intelligibility?
2) Is there a significant difference in the impact of the individual error patterns on intelligibility, regardless of consonant or vowel status?

Chapter 2: Methods

Participants

Adult listeners. Thirty-one listeners participated in the present experiment. They were recruited through mandatory participation in research credit as part of their course work through the University of Alberta Linguistics Department. Participants received 1% of their overall grade by participating in the experiment. Demographic information including native language, gender, age, and years of experience working with children was obtained.

Participants were limited to native speakers of Western Canadian English. A native Western Canadian English speaker was defined as a person who was raised in the provinces of British Columbia, Alberta, or Saskatchewan at least since the age of 3 and had not lived outside of Western Canada for more than 2 years. Only native Western Canadian English speakers were used in order to ensure that any influence from a first language that is not English or a dialect quite different from Western Canadian English would not interfere with the recognition of the words. For this reason, three participants were omitted from the final analysis.

A multiple regression analysis was run to determine if the two numerical between-subjects factors (i.e., age, and experience working with children) were significant predictors for overall percent accuracy (see Appendices A and B). The participants included in the final analysis (N= 28) were between the ages of 18 and 27 (M = 20.4 years). Self-reported experience working with children ranged between 0 years to 11 years with an average of 2.92 years of experience. Both values were found to be non-significant predictor variables for the model predicting overall participant accuracy (i.e., $F(2, 25) = .549$, $p = .584$).

The other two between-subjects variables (i.e. reported history of SLP services and gender) were categorical and were thus compared using independent samples t-tests with equal variances not assumed and were adjusted for multiple comparisons with the Bonferroni procedure (Haynes & Johnson, 2009). A significant difference would therefore have to report a *p*-value less than .025. Three participants reported a history of receiving speech services, however, their overall accuracy was not significantly different from the other participants (t = -.438, df = 2.583, p =.679; see Appendix C). They were therefore included in the overall analysis. Lastly, there were 5 males and 21 females in the final participant pool. Again, gender was not a significant factor for this group (t = .990, df = 5.005, p = .368; See Appendix D).

Between subject factors were not evenly represented in the data and some had very small participant pools (e.g., only 5 males). Subsequent investigation into interaction effects found that most could not be analyzed due to the limited number of groups being compared. However, those that were reported were not significant. Therefore, the between subjects variables were disregarded for further statistical analyses.

Child speaker. A girl aged 6 years and 7 months was recruited to record the stimulus words. She was born in Arizona and lived there for 2 years, but her mother was from Western Canada. The child moved to Western Canada at two years of age and has not lived outside of Canada since. Her dialect was agreed upon as Western Canadian as judged by the experimenters.

Over the course of a one hour session, the child produced all of the stimulus words, including elicited productions of real words (i.e., correct forms) and imitated productions of nonwords (i.e., derived words resulting from the application of error patterns). All of the 78 words (18 real, 54 nonwords, and 6 practice words) were recorded using a head-mounted microphone (Shure WH20) and Marantz Professional PMD661 audio recorder. Seven of the nonwords were re-recorded on a second day during the same week to ensure clear recording of error productions. The child was asked to produce each word/nonword approximately 3 times, at the discretion of the experimenter, in order to obtain recordings of sufficient quality and consistency. She received small sticker incentives to encourage her to continue with the task and a small toy at the end of the task.

For the correct productions, the child named images of real words or was asked to complete cloze sentences with pictures (e.g., A circle is ______). For the nonword error productions, she imitated the researcher's nonword productions. Two trained Western Canadian speakers blind to the experimental design broadly transcribed each of the error productions after the recordings were completed. When the sound-by-sound broad transcriptions were compared, the transcribers had 94.9% inter-transcriber consistency and an average of 94.4% consistency with the underlying target.

The recordings were selected using Praat (Boersma, 2001). The clearest production of each word and nonword that had a consistent quality, loudness, and rate was selected. A Praat script was used to standardize the individual wav files at 60 dB to account for variation in loudness (Sims, 2010).

Stimulus Items

Eighteen words common in the speech of children were selected as stimulus items. Eight of the words were selected from the *MacArthur Bates Communicative Development Inventories* (MCDI; Fenson, Marchman, Thal, Dale, Reznick & Bates, 2007) to ensure that they are high frequency words for children. The remaining 10 words were selected based on their applicability to the desired error patterns. All 18 words were confirmed as being high frequency words based on ClearPOND Density Database (Marian, Bartolotti, Chabal & Shook, 2012). The 10 words selected by the experimenter were within the same frequency range as the 8 words selected from the MCDI (i.e., between 7.6 words to 137.5 words per million) or had higher frequency than the MCDI words (See Appendix E for a complete list of stimulus words). Clinical judgments from two separate clinicians were used to confirm likelihood of the use of these words in child speech.

Phonological neighbourhood density was controlled for the target words. Each word had between 6 and 16 neighbours based on ClearPOND Density Database (Marian, Bartolotti, Chabal & Shook, 2012). A speaker of Western Canadian English confirmed the phonological neighbours produced by the database were acceptable words in Western Canadian English and words which would not be produced in the dialect were removed. Phonological density was investigated in post-hoc statistical analysis to confirm that the factor was controlled effectively (see: Lexical Item Analysis).

Lastly, efforts were made to use primarily single syllable words to reduce possible effects of word length. Only 2 of the 18 words were two syllables. This was investigated in a post-hoc analysis to determine any possible word length effects. Table 2.1 summarizes the experimental control planned for the factors known to affect intelligibility that were discussed earlier.

Table 2.1

Experimental Control for Factors Known to Affect Intelligibility

Factors	**Planned Control of Influencing Factor**
Speaker Factors	
• Rate of Speech	• Using a single speaker recorded within the same week eliminated much of the possible speaker-specific variability that could be attributable to other factors affecting intelligibility
• Hesitations, pauses, & repetitions	• Single word productions
• Voice quality & resonance	• Single speaker
• Cognitive factors internal to the speaker (e.g. simultaneous task involvement)	• Single speaker
• Acoustic factors	• Single speaker
• Vocal loudness	• Amplitude/loudness electronically controlled
• Type of error patterns (e.g. vowels vs. consonants)	• Manipulated variable
• Overall speech sound accuracy	• Phonetic identity confirmed by trained transcribers
• Inconsistent realization of sounds	• N/A due to only single opportunity for speech sound error to be present per stimuli
Predictability Factors	
• Predictability of semantic phrases/sentences	• N/A due to single word productions
• Neighbourhood density of error production	• Consistent range selected & Post-hoc analysis
• Message length	• N/A due to single word productions
• Lexical frequency	• Consistent range selected & Post-hoc analysis
Contextual Factors	
• Speaking environment (e.g. Background noise)	• Standardized listening environment
• Presence of visual information	• Auditory information only
• Concomitant facial and hand gestures	• Auditory information only
Listener Factors	
• Age (e.g. Adult-Child interaction)	• Post-hoc analysis
• Dialect	• Listeners with matching dialect only
• Familiarity	• Each listener heard the same number of items
• Listener experience	• Reported experience with children investigated in post-hoc analysis

Each target word was selected because specific single phoneme substitutions (or combined substitutions) resulted in English nonwords. For each target, four forms were generated: the correct real word form, and three nonword derived forms, including 1) a form with a single consonant substitution error, 2) a form with a vowel substitution error, and 3) a form in which both the consonant and vowel substitutions were combined. For example, for the target word *fast*, the four manipulations were: 1) the correct form [fæst], 2) consonant error [fæs], 3) vowel error [fɑst], and 4) combined consonant and vowel error form [fɑs]. Each of these forms was an experimental condition, which every listener heard. Across the 18 words, each consonant substitution error pattern and each vowel substitution error pattern was applied 6 times, and each combination of consonant and vowel error patterns was applied twice (see Appendix F for details).

The consonant and vowel error patterns selected were all common error patterns displayed by children with SSDs (Reynolds, 1990; Pollock & Keiser, 1990; Bowen, 1998), and included those that had the strongest impact on intelligibility in previous work by Vaughn and Pollock (1997). Diphthong Reduction, Backing, and Laxing were selected as vowel error patterns. The consonant error patterns consisted of Prevocalic Voicing, Consonant Cluster Reduction, and Stopping of Fricatives/Affricates.

Having nonword experimental stimuli ensured that the consonant and vowel error patterns were not confounded by word recognition factors such as frequency. When substitutions were applied, some words such as *candy* became names of people (i.e., /gɑndi/) or low frequency words not commonly spoken by children (e.g., *child* underwent Stopping to become *tiled*). To ensure that the listeners would treat them as errors, the adults were instructed that none of the words were names of people and all of the words were common words spoken by children.

Experimental Procedure

The experiment began with a screening of the listener's hearing following the *Hearing Screening Guidelines* as outlined by Alberta College of Speech-Language Pathologist and Audiologists (ACSLPA; 2008). The participants listened to pure tones at 1000, 2000, and 4000 Hz at 20 dB through headphones and were required to signal that they had heard the sound. All participants passed the hearing screening. The participants were then placed in a sound attenuated booth with a computer keyboard and headphones for the experimental task.

Prior to the start of the experiment, the participants were told that the words were real words of English, not names of people, and were common words that children might say. The participants typed out their answers and could review their answers on the screen before moving onto the next item. After responding, they also recorded their confidence in their answer as "high" or "low".

The participants listened to the recordings of the child through headphones and were asked to type out the real English word that they believed the child was trying to say, an open-set intelligibility design. As outlined in the intelligibility section, an open-set intelligibility test often yields lower overall intelligibility scores (Vigouroux & Miller, 2007) than closed-set intelligibility test. Because ceiling effects were possible due to the high frequency words selected for this study, an open-set design would more likely highlight difficulties in understanding. This is also the design that was used in the original Vaughn and Pollock (1997) study.

Each participant heard six practice items consisting of real and non-real words. Then the participants had an opportunity to ask the researcher any questions about the procedure during a brief break. There were no breaks during the administration of the experimental stimuli. The experiment was created using e-Prime software to ensure consistent instructions. The words were randomized for each listener to avoid any biasing based on list order or listener adaptation to a specific error type. The experiment took approximately 15 minutes to complete. The percentage of words correctly identified was used as the measure of intelligibility for the experiment.

Chapter 3: Results

Scoring Procedure

The participants' typed answers were run through a computer script which matched the answers to the correct orthographic form (Massie, 2014a) and scored either a "1" if the participant wrote the target word or a "0" if the participant typed out any word that was not the correct target word or failed to respond. Spelling errors (e.g., *tniy* when the correct answer was *tiny*) were manually identified and flagged by the researcher and given to a secondary researcher as a reliability check. Confirmed misspellings of the correct target word were re-coded as a "1."

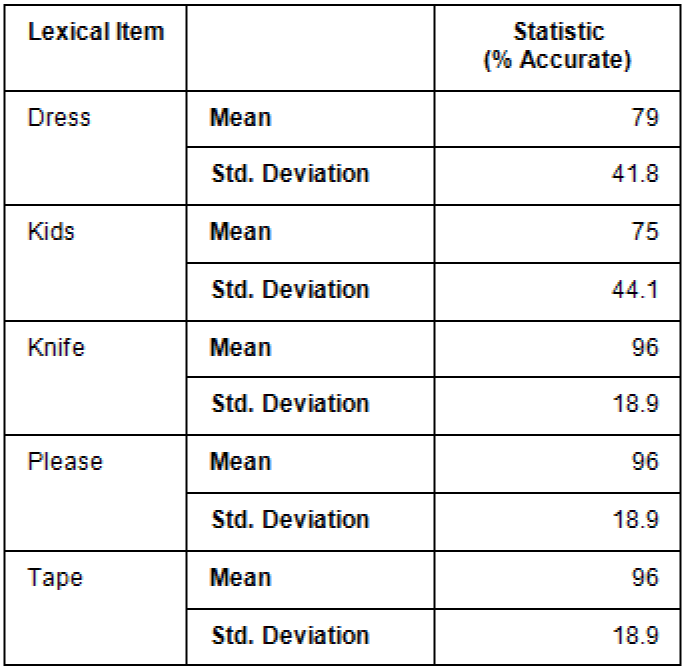
Using additional computer scripts, average percent correct responses by participant were calculated for i) overall accuracy, ii) each of the lexical items, iii) the error category (e.g., consonant errors, vowel errors, combined consonant and vowel errors; Massie, 2014b), and iv) and individual error patterns (e.g., Stopping, Diphthong Reduction; Massie, 2014c).

Lexical Item Analysis

All of the correct forms had a mean accuracy of 100% across participants with a standard deviation of 0.00 except for *dress*, *kids*, *knife*, *please*, and *tape* (see Table 3.1).

Table 3.1

Accuracy and Standard Deviation of the Correct Forms of Lexical Items which were Less than 100% Accurate



Lexical Item		**Statistic (% Accurate)**
Dress	**Mean**	79
	Std. Deviation	41.8
Kids	**Mean**	75
	Std. Deviation	44.1
Knife	**Mean**	96
	Std. Deviation	18.9
Please	**Mean**	96
	Std. Deviation	18.9
Tape	**Mean**	96
	Std. Deviation	18.9

The correct production of *kids* was identified as *kits* 4 times. None of the error production forms were identified as *kits*. The correct production of *dress* was identified as *just* 3 times. Six of the vowel error production form for *dress* ([dɹʌs]) were identified as *just*. As devoicing associated with final obstruents has been observed in some dialects of adult English speech, including Canadian English (Podlubny, 2014; see also Smith, 2012) and Liverpool English (Watson, 2007), the target word *kids* remained in the analysis. Furthermore, perception/production of affrication ([ʤ]) in "dr" clusters is known (Read, 1971); therefore, the target word *dress* also remained for further analysis.

The accuracy of the four forms of each lexical item (e.g., [fæst], [fæs], [fɑst], and [fɑs]) were averaged to determine the overall accuracy of each of the 16 lexical items included in the analysis. There was a considerable range of average accuracy by word (62% to 44%; see Appendix G for details). Given this variability, lexical factors identified during the stimulus creation were verified for appropriate experimental control in the following analysis. That is, the independent variables of: i) lexical frequency, ii) number of syllables, and iii) phonological density as reported by the ClearPOND database were investigated to determine if they were significant factors for predicting average percent accuracy of the lexical items through regression analyses.

The assumption of parametric data such as homogeneity of variance, numerical status, and normal distribution of data were verified (Brace, Kemp & Snelgar, 2013). Lexical frequency was transferred to a log function to meet the assumptions of normality. The word *think* was removed from the first round of analysis, as it was identified as an outlier for the log function of frequency. The regression model consisting of number of syllables, density of phonological neighbourhood, and frequency was found to predict overall accuracy (R^2 change = .475, $F(3, 13) = 3.915$, $p = .034$). However, the number of syllables in a word was the only significant predictor ($\beta = .713$, $p = .007$).

As all but two of the target words were single syllable words, the two syllable words (i.e., *tiny* and *candy*) were removed from final analysis. With syllable length removed as a variable, the remaining regression model considered only density of phonological neighbourhood and lexical frequency. This model did not significantly predict overall accuracy (R^2 change = .148, $F(2, 12) = 1.040$, $p = .383$).

The outlier *think* was reintroduced to the analysis, and log of frequency was still not determined to be a predictor when the regression analysis was run a subsequent time (i.e., R^2

change = .150, *F*(2, 13) = 1.148, *p* = .347). Therefore, all 16 target single syllable words remained for the final analysis.

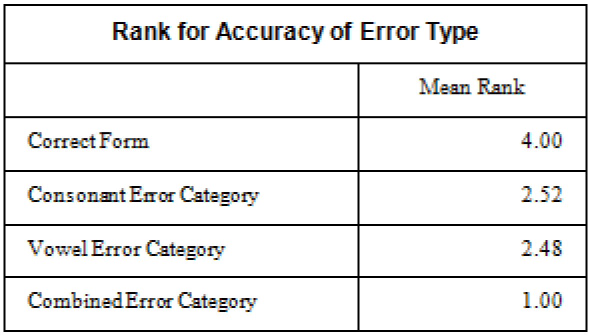
Analysis of Error Categories

The dependent variable "average percent correct for each participant" was investigated for each of the four levels of the independent variable "error category" (i.e., correct, vowel errors, consonant errors, and combined errors). Again, the assumptions of parametric statistics were analyzed. Mean accuracy for the "correct form" was heavily skewed in the positive direction due to the anticipated ceiling effect and therefore, violated the assumptions of parametric data. "Combined errors" violated normal distribution because the mean of all combined error forms was 0.00% with a SD = 0.00 after the removal of 13 identified outliers.

The percent correct for the four error categories were therefore compared using the non-parametric Friedman One-Way Within-Subjects test (Brace, Kemp & Snelgar 2013). This showed significant differences for the four repeated measures: χ^2 (3, *N*=28) = 78.12, *p* <.001. Furthermore, upon inspection of the ranking, "correct form" consistently ranked highest overall and "combined errors" ranked lowest overall. There was minimal difference between the rankings of vowel error category and consonant error category, although consonant average is ranked slightly higher than vowel average (see Table 3.2).

Table 3.2

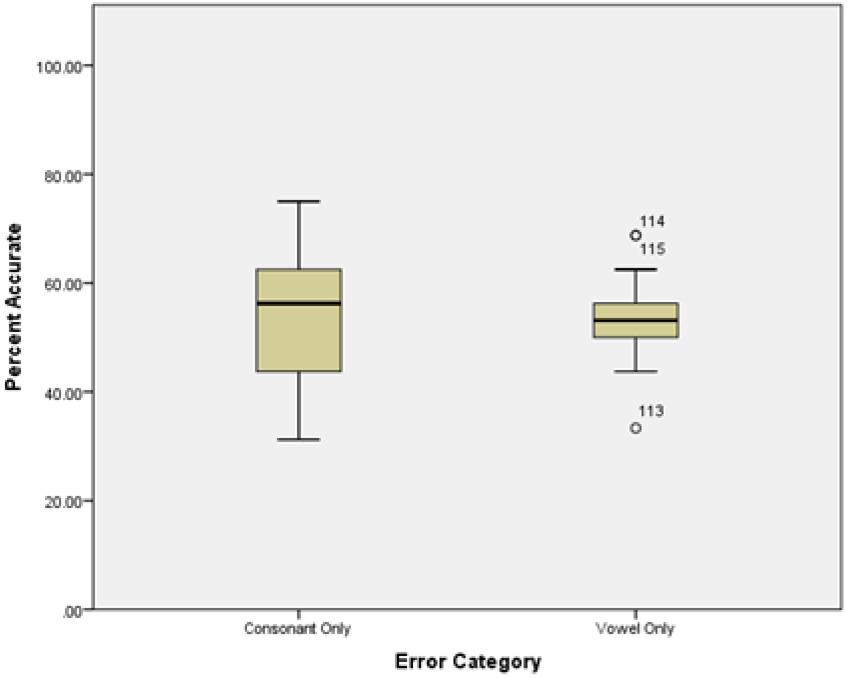
High-to-Low Ranking of Accuracy from the Friedman One-Way Within Subjects Test for the Four Main Error Categories



Rank for Accuracy of Error Type	
	Mean Rank
Correct Form	4.00
Consonant Error Category	2.52
Vowel Error Category	2.48
Combined Error Category	1.00

Visual inspection showed that standard deviation of the consonant error category (SD = 0.10) was larger than the vowel error accuracy distribution (SD = 0.06) despite similar means. As the average percent correct vowel and consonant error categories by participant met the assumptions for parametric data, when three outliers were removed from the vowel error category (see Figure 3.1), a paired t-test was performed to compare the two groups.

Figure 3.1. Boxplot of Mean Percent Accuracy for Vowel and Consonant Error Patterns Including Outliers



Although only one paired t-test was analyzed, there were theoretically 6 comparisons which could have been made and, therefore, significance would only be reached if the *p* value was below .0083 (as per the Bonferroni procedure with the alpha level set at .05; Haynes & Johnson, 2009). No significant differences were seen between vowel error accuracy average and consonant error accuracy average as shown by a post-hoc paired samples t-test (i.e., *t* (24) = 0.57, *p* = .578).

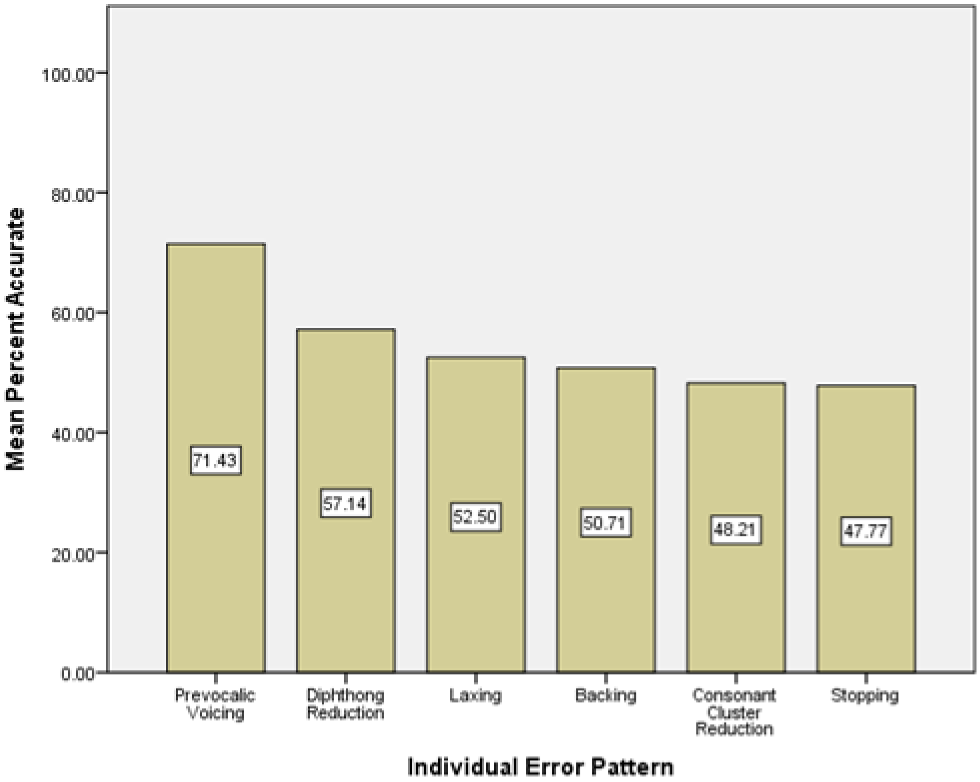
A power analysis between the average percent correct for the consonant error category and the vowel error category revealed the experiment was underpowered (β = .06; Hintze, 2013). However, an investigation into a sufficient sample size stated that a sample size of *N* = 2989 (Hintze, 2013) would be required to detect a potential difference between these two categories. Therefore, the researcher felt confident saying that if a difference between the vowel substitutions and consonant substitutions existed for this group of words, the difference would be small and would likely result in a negligible effect size.

Individual Error Pattern Analysis - Single Substitution Errors

Individual error patterns were investigated by averaging the accuracy by error pattern for each participant. When investigating the dependent variable "average percent correct for each participant" six levels of the independent variable "individual error pattern" were compared. The six levels were checked to see if they met the assumptions of parametric data. It should be noted that the average for Prevocalic Voicing was determined for 4 rather than 6 target words, as *candy* and *tiny* were removed due to the influence of word length. The removal of *candy* and *tiny* also resulted in Backing being reduced to 5 words and Diphthong Reduction being reduced to 5 words.

Prior to removing outliers, Diphthong Reduction and Consonant Cluster Reduction were not visually different from Laxing, Backing, and Stopping (shown in Figure 3.2). With the outliers removed Diphthong Reduction and Consonant Cluster Reduction had a SD = 0.00 thus violating the assumptions of parametric data. Furthermore, Prevocalic Voicing, Stopping, and Laxing were skewed in the positive direction after removal of outliers. Therefore, the outliers were returned to the analysis and a Friedman One-Way Within-Subjects test was performed.

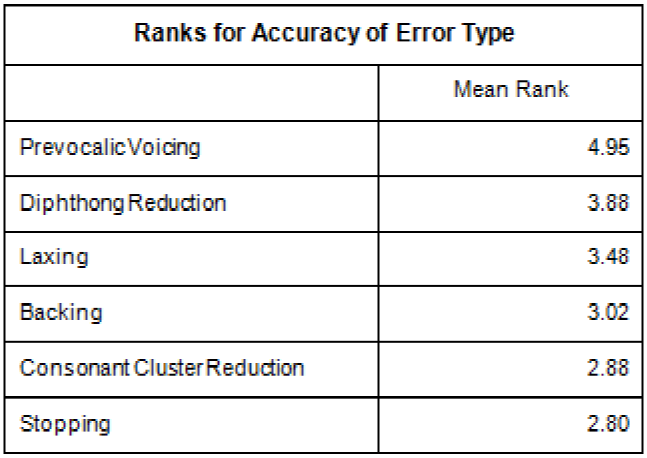
Figure 3.2. Bar Chart Showing Mean Percent Accuracy for 6 Single Substitution Error Patterns



The Friedman One-Way Within-Subjects analysis showed that across the six single substitution error patterns there were significant differences: χ^2 (5, *N*=28) = 28.20, *p* <.001. Five of the six single substitution error patterns ranked fairly closely (see Table 3.3). Upon secondary analysis removing Prevocalic Voicing, there were no significant differences found between these five groups (i.e., χ^2 (4, *N*=28) = 7.56, *p* =.108).

Table 3.3

High-to-Low Ranking of Accuracy from the Friedman One-Way Within Subjects Test for the Single Substitution Error Categories



Ranks for Accuracy of Error Type	
	Mean Rank
Prevocalic Voicing	4.95
Diphthong Reduction	3.88
Laxing	3.48
Backing	3.02
Consonant Cluster Reduction	2.88
Stopping	2.80

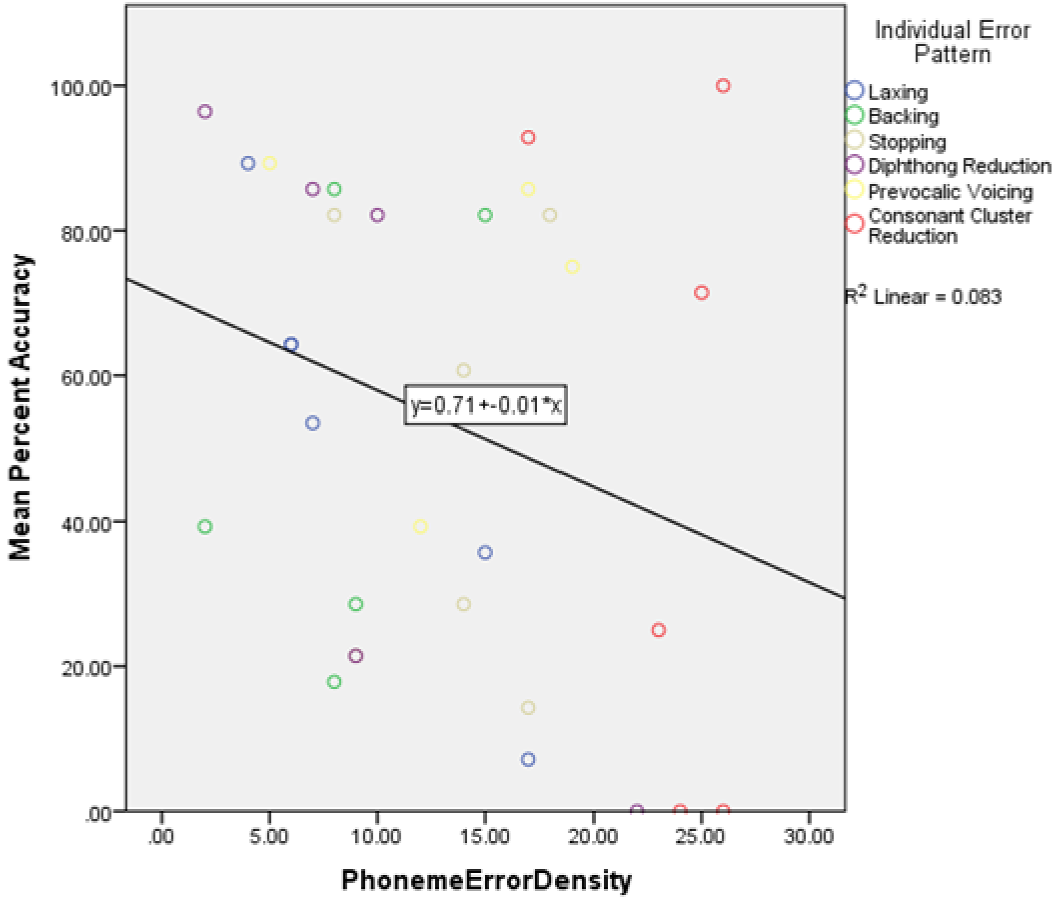
It should also be noted that when the same statistical analysis was run with *candy* and *tiny*, there was an increase in average across Prevocalic Voicing, Laxing, and Backing. This did cause a minor change in the ranking between the single substitution errors. However, Prevocalic Voicing still consistently had the highest average percent accuracy (see Appendices H and I).

Alternative factors for single substitution error patterns. A number of possible factors could account for the difference between these individual error patterns including word position of substitution error, or density of the phonological neighbourhood of the error production.

The phonological density of the error productions resulting from single consonant or vowel error were calculated using the ClearPOND Density Database (Marian, Bartolotti, Chabal & Shook, 2012). A regression analysis was performed to determine if there was a significant relationship between the phonological density of the error productions resulting from single substitution error and the average accuracy of all participants for those same errors (see Figure

3.3). Surprisingly, the relationship between mean accuracy and neighbourhood density of the error productions was not significant (R^2 change = .083, $F(1, 31) = 2.6984$, $p = .111$).

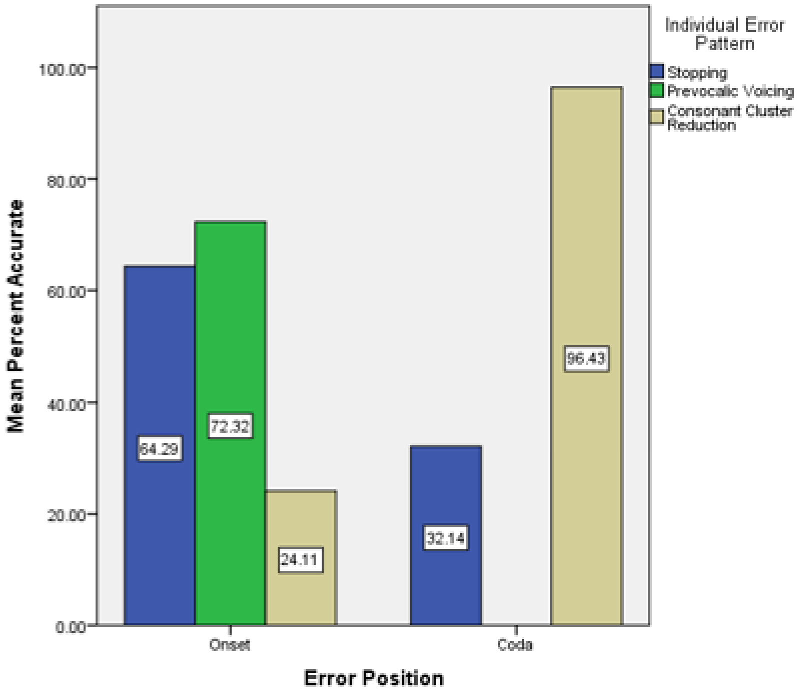
Figure 3.3. Scatterplot of Mean Accuracy by Neighbourhood Density of Error Productions



As mentioned previously, some of the participants reported the perception of [ʤ] for the initial [dɹ] cluster in *dress* through their orthographic word identifications. When the neighbourhood density of the error production was adjusted to reflect possible perception of [ʤ] for the vowel error form of *dress* (i.e., [dɹʌs]), the relationship between the phonological density of the error productions resulting from single consonant or vowel error and the average accuracy of all participants for those same errors was still not significant (R^2 change = .087, $F(1, 31) = 2.865$, $p = .101$).

The effect of the position of the error within the word was investigated with only the consonant error forms. All three individual consonant error patterns had lexical items with errors occurring in the onset. The Prevocalic Voicing error pattern occurred in the onset position for all 4 words included in the single syllable (i.e., *tape, toy, kids*, and *treat*). The Stopping error pattern occurred in the onset position for 3 words (i.e., *think, fruit*, and *child*) and the Consonant Cluster Reduction error pattern occurred in the onset position for 4 words (i.e., *spoon, dress, space*, and *brown*). There are 3 words that had Stopping errors in the coda position (i.e., *grass, please*, and *knife*) and 2 words that had Consonant Cluster Reduction in the coda position (i.e., *round* and *fast*).

Figure 3.4. Bar Chart of Average Accuracy for Individual Consonant Error Patterns by Word Position



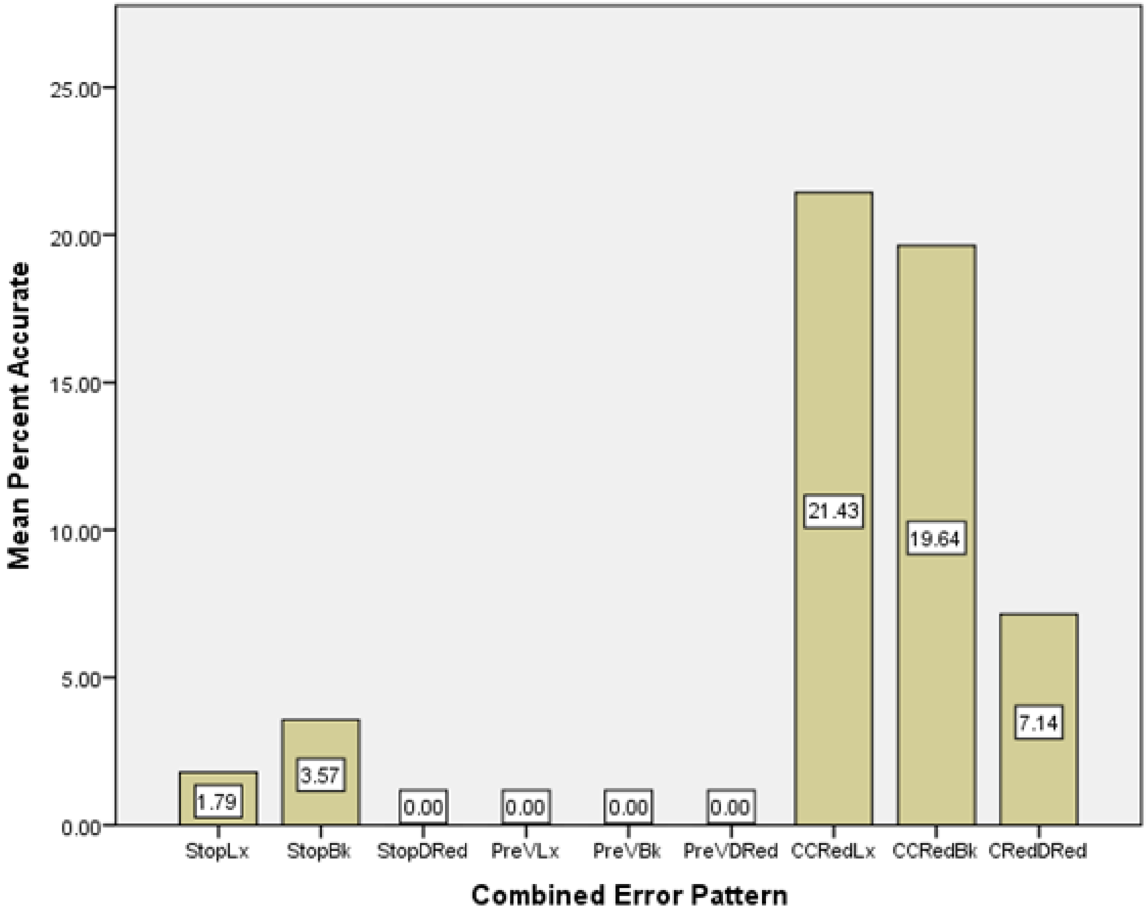
As shown in Figure 3.4, there was a difference between the three levels of the dependent variable average percent accuracy by participant for individual error patterns in the onset position (i.e., χ^2 (2, *N*=28) = 35.80, *p* <.001). The difference between the two levels of the dependent variable "average percent accuracy by participant" for individual error patterns in the coda position was also significant (i.e., χ^2 (1, *N*=28) = 4.48, *p* =.034).

When the same error pattern was compared in separate error positions, different patterns emerge for individual error patterns. That is, for Stopping there was a higher average accuracy for errors in the onset and for Consonant Cluster Reduction there was a higher average for errors in the coda. There was a significant difference between the error positions for Stopping (i.e., χ^2 (1, *N*=28) = 13.76, *p* <.001) and Consonant Cluster Reduction (i.e., χ^2 (1, *N*=28) = 27.00, *p* <.001).

Individual Error Pattern Analysis - Combined Substitution Errors

As children who have vowel error patterns often have consonant error patterns as well, differences between the individual combinations of consonant and vowel errors were also investigated. A separate set of nine levels of the dependent variable "average percent correct for each participant" (e.g., Prevocalic Voicing+Laxing) were analyzed for the assumptions used for parametric statistics. All of the categories violated the assumptions of normality and thus were compared using non-parametric analysis and visual comparison.

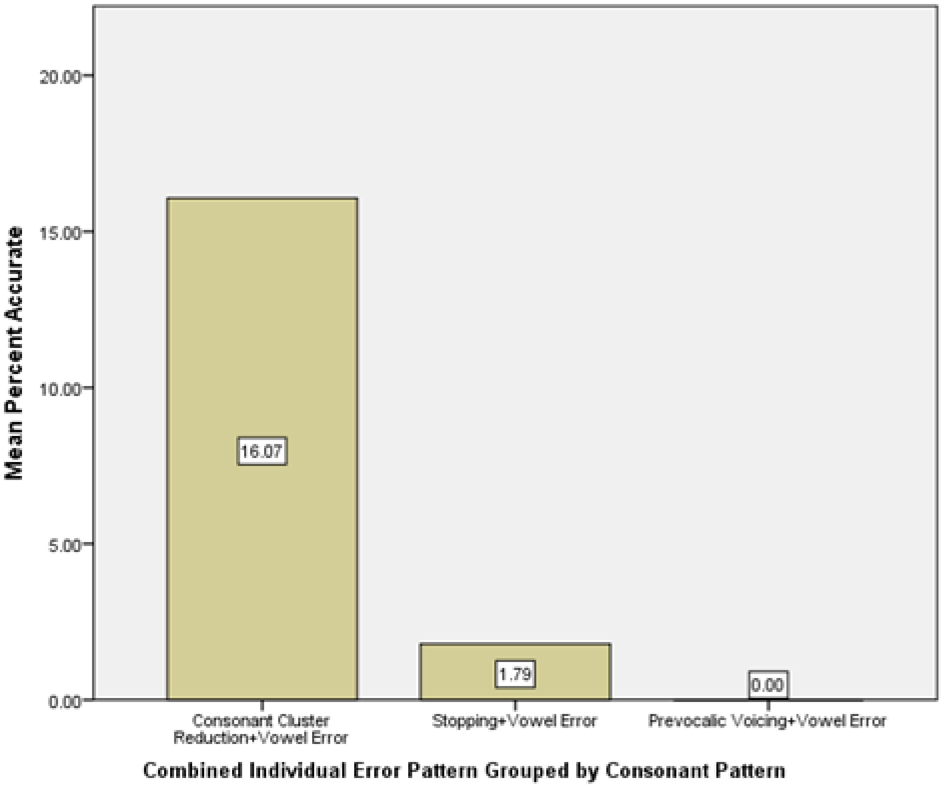
Figure 3.5. Bar Chart of the Mean Accuracy for Combined Vowel and Consonant Error Patterns



Note. StopLx = Stopping + Laxing; StopBK= Stopping + Backing; StopDRed= Stopping + Diphthong Reduction; PreVLx= Prevocalic Voicing+ Laxing; PreVBk= Prevocalic Voicing + Backing; PreVDRed= Prevocalic Voicing + Diphthong Reduction; CCRedLx= Consonant Cluster Reduction +Laxing; CCRedBk = Consonant Cluster Reduction+ Backing; CRedDRed= Consonant Cluster Reduction+ Diphthong Reduction

Again, significant differences were found between the nine combined error patterns (i.e., χ^2 (8, *N*=28) = 61.465, *p* <.001). Consonant Cluster Reduction combined with Laxing, with Backing, and with Diphthong Reduction ranked higher than the other vowel and consonant error groups. Noticing possible patterns in the combination errors, the data was further grouped by consonant pattern and vowel pattern, separately.

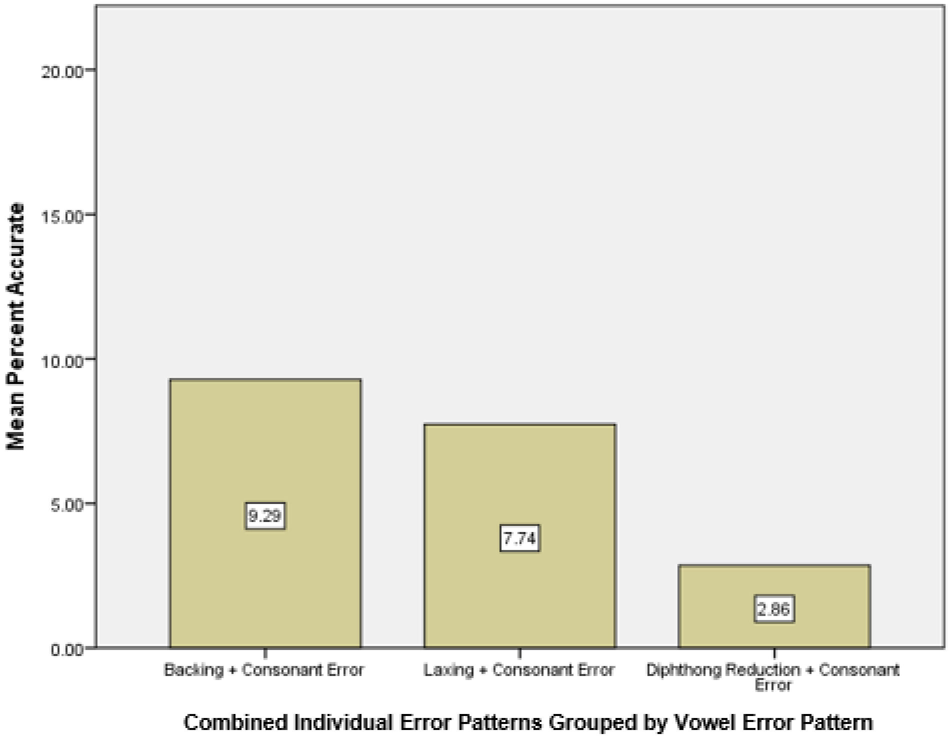
Figure 3.6. Bar Chart of the Mean Average when Combined Error Patterns are Grouped by Consonant Error Pattern



Across the three new categories created by combining consonant patterns with other vowel error patterns, there were significant differences for the dependent variable “average percent correct for each participant”: χ^2 (2, *N*=28) = 36.636, *p* <.001. Consonant Cluster Reduction combined with any of the 3 vowel error patterns had a higher level of accuracy than Prevocalic Voicing or Stopping with the other vowel error patterns (see Figure 3.6).

When “average percent correct for each participant” for the combined error patterns was grouped by vowel error patterns, the three levels were statistically different (i.e., χ^2 (2, *N*=28) = 5.851, *p* =.05). That is, Diphthong Reduction combined with one of the three consonant error patterns had a consistently lower level of accuracy than Laxing or Backing (see Figure 3.7).

Figure 3.7. Bar Chart of the Mean Accuracy when Combined Error Patterns are Grouped by Vowel Error Pattern

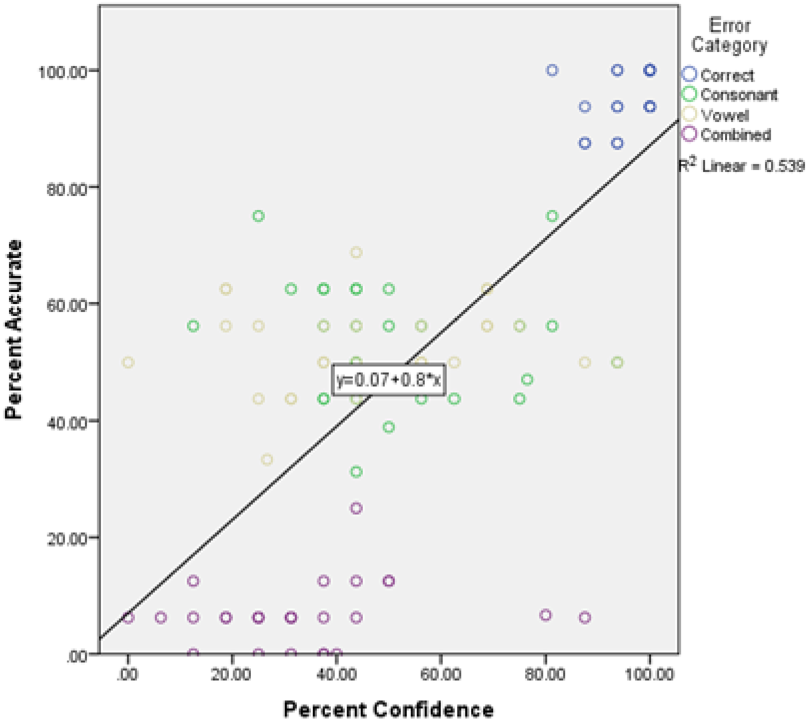


Analysis of Confidence

The high or low confidence ratings were investigated to determine if any of the error types (i.e., vowel, consonant, and combined) were associated with higher (or lower) levels of uncertainty. The participants indicated if they had "high" or "low" confidence in their response by pressing either "1" or "0". Confidence was investigated as a second dependent variable. It was compared in three ways: i) error pattern category averages by participant (4 levels), ii) single substitution error pattern averages by participant (6 levels), and iii) combined error pattern averages by participant (9 levels; Massie, 2014d). Each of the levels for error pattern category fit the assumptions for parametric data.

A regression analysis was performed to determine the relationship between "average confidence by participant" and "average percent accuracy by participant." As shown in Figure 3.8, higher levels of confidence were associated with error patterns that had higher overall accuracy (R^2 change = .492, $F(1, 133) = 129.051$, $p < .001$).

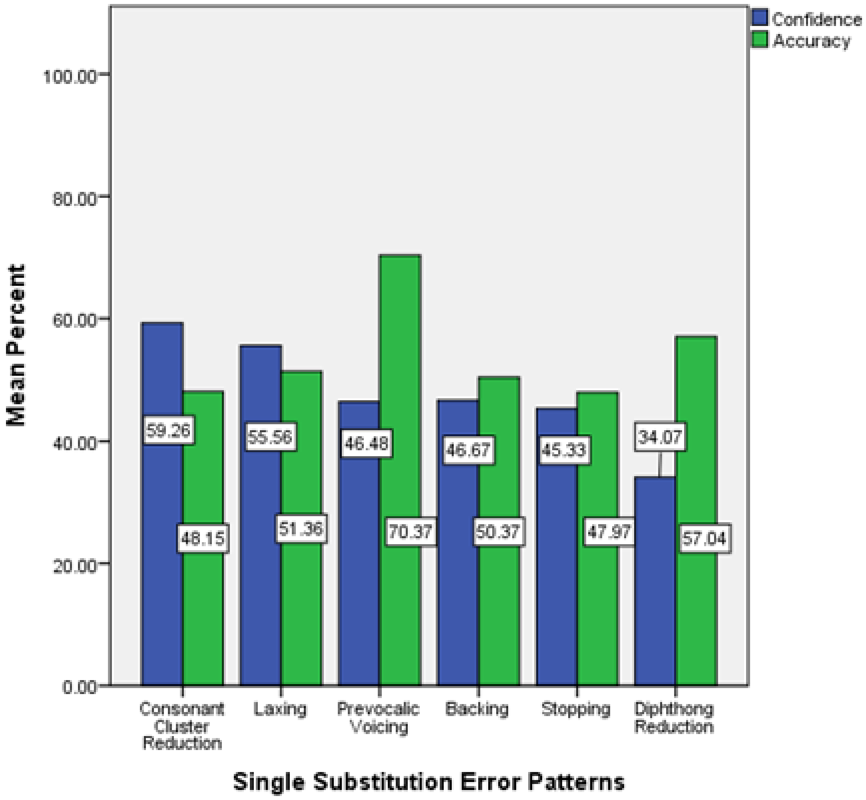
Figure 3.8. Regression Line Showing Relationship Between Confidence and Accuracy with Error Categories Identified



Next, the averages by individual error pattern were investigated to see if there was a unique pattern for any of the individual error patterns. Again, there were violations for parametric data across many groups, so nonparametric analysis was used. See Appendix J for SPSS output on Friedman One-Way Within-Subjects test.

Each of the average confidence individual error pattern by participant was compared to average accuracy of the same using Friedman One-Way Within-Subjects test. Prevocalic Voicing and Diphthong Reduction had low confidence despite higher levels of accuracy (see Figure 3.9). These differences were significant (χ^2 (1, N=27) = 5.762, p =.016, and χ^2 (1, N=27) = 11.636, p =.001, respectively). The difference between confidence and accuracy for Consonant Cluster Reduction was also significantly different although did not appear to be so on initial visual inspection (χ^2 (1, N=27) = 5.762, p =.016). Other differences between single substitution errors' accuracy and confidence were not significant. In general, the pattern for higher accuracy associated with higher confidence held with much lower confidence in the combined errors section (see Figure 3.9).

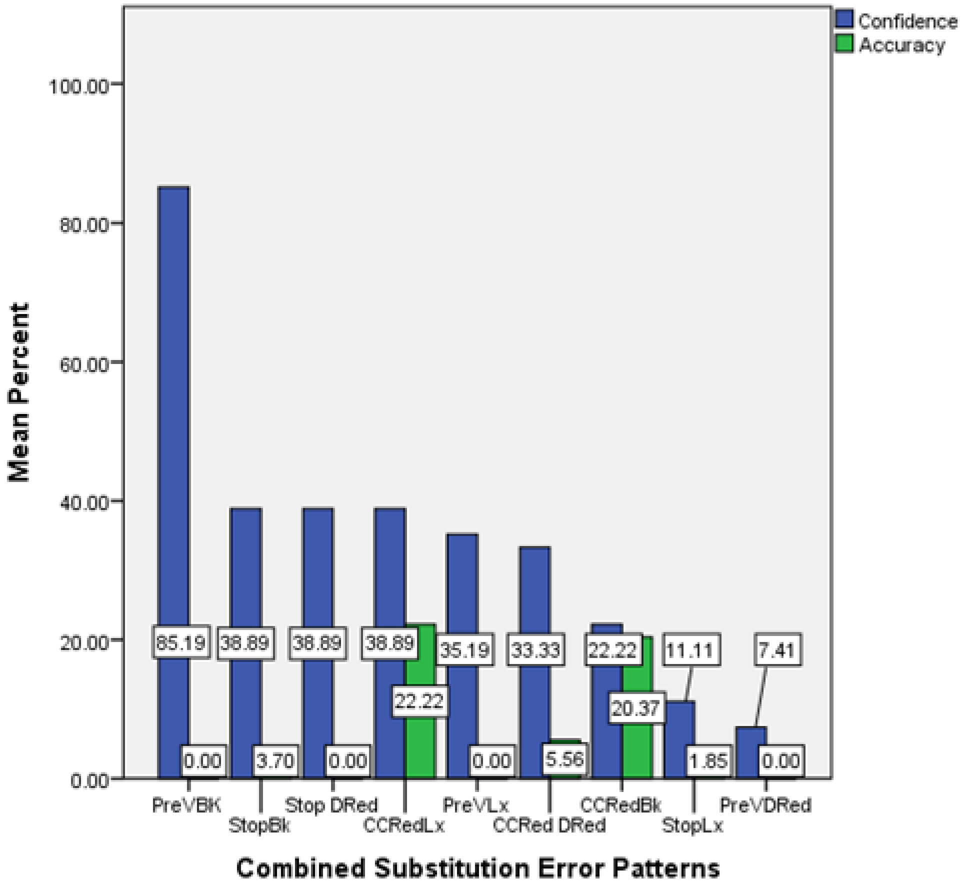
Figure 3.9. The Average Percentage Accuracy and Average Percent Confidence for Single Substitution Errors



Note. The single substitution error patterns are arranged by confidence from high-to-low. One participant was removed from this analysis due to invalid responses for confidence therefore N = 27

The same procedure as described above was performed for the nine levels of "average confidence by participant" for the combined substitution errors. Prevocalic Voicing+Backing had a very high confidence despite having an overall average accuracy of 0.00% (see Figure 3.10). All of the combined error substitutions have significant differences between the confidence and accuracy except for Consonant Cluster Reduction+Backing (χ^2 (1, *N*=27) = 077, *p* =.782).

Figure 3.10. The Average Percentage Accuracy and Average Percent Confidence for Combined Substitution Errors



Note . The single substitution error patterns are arranged by confidence from high-to-low. One participant was removed from this analysis due to invalid responses for confidence therefore *N* = 27

Note. StopLx = Stopping + Laxing; StopBK = Stopping + Backing; StopDRed = Stopping + Diphthong Reduction; PreVLx = Prevocalic Voicing+ Laxing; PreVBk = Prevocalic Voicing + Backing; PreVDRed = Prevocalic Voicing + Diphthong Reduction; CCRedLx = Consonant Cluster Reduction+Laxing; CCRedBk = Consonant Cluster Reduction+ Backing; CRedDRed = Consonant Cluster Reduction+ Diphthong Reduction

Chapter 4: Discussion

This study was a partial replication of a previous study examining the relative effects of consonant and vowel errors on intelligibility. Single word targets produced by a child with and without errors were recorded and played to a group of naive adult listeners. Errors represented common consonant and vowel error patterns, alone or in combination, and stimulus words were controlled for phonological density and lexical frequency. Accuracy of listener identification of the target words was analyzed across the four conditions (correct, vowel error only, consonant error only, consonant and vowel error combined), six single substitution error patterns (Prevocalic Voicing, Stopping, Cluster Reduction, Laxing, Backing, and Diphthong Reduction), and nine combinations of error patterns.

Question 1: Is There A Difference In The Impact Of Vowel Error Patterns Compared To Consonant Error Patterns On Intelligibility?

When single vowel error and single consonant errors were compared, no differences were found in the ability for participants to identify the target word. Words containing both consonant and vowel errors had substantially lower levels of accuracy. These findings are in agreement with the original findings from the Vaughn and Pollock (1997) study which also failed to find significant differences between vowel and consonant error categories but found a lower level of accuracy for combination errors.

These data suggest that vowel errors have the same potential to affect intelligibility as consonant errors. Because clear data supporting why vowels could be selected as targets has been lacking in literature, this is an important finding. These data provide the evidence-based support needed to encourage clinicians to investigate and target vowel errors more closely. If a clinician's goal is to impact intelligibility as much as possible, examination of vowel errors for a client is important. Furthermore, if vowel errors are identified, they should be considered as targets for remediation along with consonant errors.

Vowel and consonant errors were represented by 3 different error patterns each. These were only a small subset of common error patterns used by children. Inclusion of additional error patterns could show different outcomes and may result in a different relationship between these two larger categories of sounds. It is possible that, had additional patterns been included as error types, differences between vowel errors and consonant errors might have been detected. However, the error patterns selected represented a range of categories including subjectively smaller changes (e.g., tensing), and subjectively larger changes (e.g., diphthong reduction). Therefore it seems likely that additional consonant and vowel patterns would likely have resulted in equal impact between vowel and consonant errors.

These findings suggest that a distinction between the vowels and consonants when considering intelligibility may be arbitrary and fails to capture stronger predictors. Yet in order to ensure an equal level of empirical support for the sound classes, adequate inclusion of all sound classes in assessment such as standardized tests, remediation, and research should be encouraged. Otherwise the continued building of empirical support for one sound class will result in the continued neglect of the other.

Question 2: Is There A Difference In The Impact Of The Individual Error Patterns On Intelligibility, Regardless Of Consonant Or Vowel Status?

Analysis of the specific individual consonant-only and vowel-only substitution patterns found no significant difference between the individual patterns except for Prevocalic Voicing. As mentioned previously, with or without the inclusion of *tiny* and *candy*, the Prevocalic Voicing error pattern had a significantly higher accuracy than the other five single substitution errors. This is a very different result from Vaughn and Pollock (1997) who found greater differences between the individual error patterns. These data show that individual error patterns alone may not be strong predictors of intelligibility although there are some significant effects. The individual patterns may represent a single distinctive feature change (e.g., voicing or manner) or a change to syllable shape (e.g., CCVC to CVC). Smaller changes such as voicing may be easier to interpret and this may explain why Prevocalic Voicing was the pattern with the highest level of accuracy. Yet, why other relatively small single feature vowel changes (e.g., Laxing) correlated with lower levels of intelligibility is not understood.

Average accuracy of all combined consonant single substitution errors in initial position was higher. Ultimately this reinforced that for these words Prevocalic Voicing was, the most accurate error pattern of the six single substitution error patterns explored. When separated by position of the error, Stopping and Consonant Cluster Reduction had opposite impacts on intelligibility. Given other linguistic theories, such as uniqueness point of Cohort Model of speech processing (Taft & Hambly, 1986), the impact on intelligibility would have been predicted as greater for the beginning of a word regardless of error type. If listeners are processing phonetic information sequentially, then a phoneme misidentification at the earlier stages of processing should result in incorrect recognition and lower accuracy overall. Yet this evidence shows greater intelligibility for Stopping errors at the beginning of the word. As this was not incorporated into the original design of the experiment, there are a relatively small number of lexical items (between 2 to 4 words) per error pattern in a given word position. A greater number of lexical items would be required to explore this relationship more clearly.

Neighbourhood density of the target word was controlled for and, as hoped for, did not play a role in overall accuracy. Although there was considerable variability in the neighbourhood density of the error productions, it did not predict accuracy. As a predictive factor, phonological density did not have a significant main effect for predicting intelligibility of single substitution error patterns. In word initial position the theoretical ranking of the error patterns' intelligibility, as predicted by the phonological density of the error productions, did not align with the current findings. That is, greatest impact on intelligibility came from Consonant Cluster Reduction in initial position (cf. Leinonen-Davies, 1988). These data show different patterns for Stopping and Consonant Cluster Reduction in unique word positions suggesting that it may not be the error pattern or the phonological density alone that impacts intelligibility. Therefore, further investigation into the interaction effect between error patterns and phonological density may provide greater amounts of information.

For the single substitution error patterns, a distinctive ranking of the patterns was not identified, as had been expected. This evidence cannot be generalized beyond these six individual error patterns but it does suggest that if an SLP intends to impact intelligibility maximally during therapy, they should target all of these six patterns equally, except perhaps for Prevocalic Voicing which could be targeted last. Although not conclusive, investigation into factors of word position and phonological density that interact with the individual error patterns may be warranted for future investigation. It is recommended that additional lexical items be included to develop a fuller picture of the interaction of these factors in future investigations.

Like the ad-hoc investigation into word position, the combination errors were also composed of very few lexical items (between 1-2 words per individual category). However, the empirical evidence collected from the combination errors is representative of what would be the most likely productions for a child with a phonological disorder characterized by vowel errors. That is, vowel error patterns are most often produced by children who also have consonant error patterns but the opposite is not necessarily true. Something which may be worth investigating in the future is the impact of multiple error patterns from the same category on intelligibility. That is, does the impact of vowel error patterns have a more cumulative on intelligibility than the cumulative impact of multiple consonant patterns? This could further aid our understanding of the relative effect of vowel and consonant errors on intelligibility by comparing those values to the values of the combined error patterns with both vowel and consonant errors.

Participants' ranking of the their confidence did not hold any unanticipated answers except for Prevocalic Voicing+Backing having a high average confidence despite having an overall average accuracy of 0.00%. For that pattern only one error production was included in the final analysis. Despite efforts to create only nonword error productions, the only error form for this combination pattern was produced as /gʊdz/ or *goods*; a low-frequency, real word unlikely to be spoken by a child. Hence it is possible that people felt more confident about the answer despite the instruction that the words were common words spoken by children.

When the combined errors were pooled into groups of a single substitution error with an error from the opposite sound class (e.g., all combined errors involving Backing were averaged in a single value) there were some analogous and some unexpected patterns are formed with the single substitution error data. For example, with a co-occurring vowel error Prevocalic Voicing has the lowest level of intelligibility. This is the opposite ranking of the single substitution error pattern of Prevocalic Voicing. With a co-occurring vowel error, Consonant Cluster Reduction had the highest level of intelligibility, yet it has equal ranking with all other single substitution error patterns except Prevocalic Voicing. With a co-occurring consonant error the three vowel error patterns have closer level of intelligibility, and they have equal ranking amongst the other 5 single substitution errors. This shows us that combination error patterns have to be considered separately from single substitution error patterns. We can even rank the combined error patterns included in this investigation in a way that might assist SLPs in selecting targets that would maximize their therapeutic impact on intelligibility. See Table 4.1.

Although the combined patterns are ranked differently when vowel and consonant errors are alone, there is one parallel. There is increased variability for the consonant error category, and a wider range of intelligibility values for the combined errors when they were grouped by consonant error pattern (see Figure 3.6 and 3.7). This suggests that consonant errors, when present in speech, may need to be investigated more closely to determine how much they are impacting a child’s speech. From a theoretical perspective it suggests that further, and likely more comprehensive investigations across more error patterns may need to be performed for a ranking such as this to be a useful tool for future SLPs.

Table 4.1

Suggested Ranking of Priority of Combined Error Patterns for Therapeutic Remediation

Suggested Priority Ranking	**Combined Error Pattern**	**Percent Intelligible**	**Standard Deviation for Percent Intelligible**
First	Prevocalic Voicing + Vowel Error	0.00	N/A
Second	Stopping + Vowel Error	1.79	13.28
Third	Diphthong Reduction + Consonant Error	2.38	16.72
Fourth	Laxing + Consonant Error	7.74	26.80
Fifth	Backing + Consonant Error	9.29	25.02
Sixth	Consonant Cluster Reduction + Vowel Error	16.07	36.83

Limitations and Future Directions of the Present Study

As alluded to previously, the inclusion of /dɹ/ in *dress* as a target for Cluster Reduction introduced confusion for listeners as the child produced this cluster with affrication (i.e., [d͡ʒɹɛs] or [d͡ʒɛs]), a commonly accepted variant in English (e.g., Read, 1971). Thus it is not surprising that listeners often identified these productions as a word beginning with the letter "j," (e.g., "jest" or "just"). In future experiments, the production and perception of consonant clusters such as /dɹ/ and /tɹ/ as affricates should be considered and avoided.

The overall accuracy for single error substitution averaged around 60% showing that listeners had difficulty in correctly interpreting the target words. The addition of filler or distractor stimuli in the experimental condition may have provided more difficulty for connecting the different forms of the experimental stimuli. For example, if the participants heard the correct form "candy" first, they may have been able to more easily determine what the error form was supposed to be. However, randomization of the experimental stimuli for each participant should have minimized any learning effects in the study.

All of the factors outlined in the intelligibility section could be investigated in tandem with individual error patterns. Of considerable interest to the present study is the relationship between word position and individual error pattern as this could confirm the limited findings of the present study.

The present study used adult participants as listeners. Consideration of intelligibility as judged by peers has been investigated for children with SSDs (Speake, Stackhouse & Pascoe 2012). This is an important question from a functional perspective, as children who have speech difficulties require peer interaction for social development. It would be worthwhile in the future to determine if intelligibility loss as a result of specific error patterns was different for children of the same age group compared to the adult listeners.

The errors produced by the child speaker in this study were identified by the transcribers and the researcher as full phonemic substitutions. That is, while the “incorrect” phonemes were produced in “error” from the original target word they may not be representative of more subtle sub-phonemic errors that might be produced by children who have SSDs.

Furthermore, the full phonemic substitutions means that different words could be created by the substitutions. Many speech sound errors for children are distortions (e.g., lateral lisps), rather than complete phonemic substitutions. Distortions do not necessarily generate the perception of new words; rather they tend to affect speech in more subtle ways. Investigating phonetic distortions would answer questions regarding intelligibility for different populations such as children with articulation difficulties or motor speech disorders, rather than children with phonological disorders.

Conclusion

The findings of the present study offer additional information for clinicians regarding the effect of vowel and consonant errors on intelligibility, which may be useful when selecting potential remediation targets. For example, results suggest giving a lower priority to remediating Prevocalic Voicing over other single substitution errors, as this error pattern had significantly less impact on intelligibility. The lack of significant differences between the other single substitution error patterns suggests that relative impact on intelligibility may not need to be considered when choosing among these error patterns. In the case of vowel errors and error patterns, this study showed that their importance with regard to intelligibility has been undervalued in the past and that when vowel errors occur, they should be considered as potential targets for remediation. Regardless, continued exploration of the factors influencing intelligibility can further deepen our understanding of how to best plan intervention for children with SSDs.

References

Alberta College of Speech-Language Pathologists and Audiologists (2008). *Hearing Screening Guidelines*. Edmonton, A.B.: Author. Retrieved from: http://acslpa.ab.ca/download/college/Hearing%20Screening%20Guideline.pdf

Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International 5*(9-10), 341-345.

Bowen, C. (1998). *Developmental phonological disorders. A practical guide for families and teachers*. Melbourne: ACER Press.

Bowen, C. (Ed.) (2009). *Children's speech sound disorders.* Oxford, England: Wiley-Blackwell, 2009.

Brace, N., Kemp, R., & Snelgar, R. (2013). *SPSS for psychologists* (Fifth ed.). New York, NY: Routledge.

Brumbaugh, K. M., & Smit, A. B. (2013). Treating children ages 3-6 who have speech sound disorder: A survey. *Language, Speech & Hearing Services in Schools, 44*(3), 306-319. doi:10.1044/0161-1461(2013/12-0029)

Connolly, J. H. (1986). Intelligibility: A linguistic view. *British Journal of Disorders of Communication, 21*, 371-376.

Dagenais, P. A., Adlington, L. M., & Evans, K. J. (2011). Intelligibility, comprehensibility, and acceptability of dysarthric speech by older and younger listeners. *Journal of Medical Speech - Language Pathology, 19*(4), 37-49.

Davis, B. L., Jacks, A., & Marquardt, T. P. (2005). Vowel patterns in developmental apraxia of speech: Three longitudinal case studies. *Clinical Linguistics and Phonetics, 19*(4), 249-274. doi:10.1080/02699200410001695367

Davis, B. L., & MacNeilage, P. F. (1990). Acquisition of correct vowel production: A quantitative case study. *Journal of Speech and Hearing Research, 33*(1), 16-27.

Dawson, J. & Tattersall, P. (2001). *Structured Photographic Articulation Test II Featuring Dudsberry: Second Edition*. DeKalb, IL: Janelle Publications, Inc.

Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A. (2009). Diagnostic Evaluation of Articulation and Phonology — U.S. edition (DEAP). San Antonio, TX: Pearson.

Donegan, P. (2013). Normal vowel development. In M. J. Ball, & J. S. Damico (Eds.), *Handbook of vowels and vowel disorders* (pp. 24-60). New York: Taylor & Francis, LLC.

Eisenberg, S. L., & Hitchcock, E. R. (2010). Using standardized tests to inventory consonant and vowel production: A comparison of 11 tests of articulation and phonology. *Language, Speech, & Hearing Services in Schools, 41*(4), 488-503. doi:10.1044/0161-1461(2009/08-0125)

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual - Second Edition*. Baltimore: Brookes Publishing.

Fey, M. E. (1986). *Language intervention with young children*. San Diego, California: College-Hill Press, Inc.

Fisher, H. B., & Logemann, J. A. (1971). *The Fisher-Logemann Test of Articulation Competence*. Boston: Houghton Mifflin.

Garcia, J. M., & Cannito, M. P. (1996). Influence of verbal and nonverbal contexts on the sentence intelligibility of a speaker with dysarthria. *Journal of Speech & Hearing Research, 39*(4), 750.

Gibbon, F. E. (2009). Vowel therapy. In C. Bowen (Ed.), *Children's speech sound disorders* (pp. 147-152). Oxford, England: Wiley-Blackwell.

Gierut, J. A. (2005). Phonological intervention: The how or the what? In A. G. Kamhi, & K. E. Pollock (Eds.), *Phonological disorders in children: Clinical decision making in assessment and intervention* (2nd ed., pp. 201-210). Baltimore: Paul H. Brookes.

Gierut, J. A., & Hulse, L. E. (2010). Evidence-based practice: A matrix for predicting phonological generalization. *Clinical Linguistics and Phonetics, Vol. 24*(4-5), 323-334. doi:10.3109/02699200903532490

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation* (Second ed.) Pearson.

Haynes, W. O., & Johnson, C. E. (2009). In Dragin S. D. (Ed.), *Understanding research and evidence-based practice in communication disorders.* USA: Pearson Education.

Hintze, J. (2013). PASS 12. NCSS, LLC. Kaysville, Utah, USA. www.ncss.com.

Hodge, M. M., & Gotzke, C. L. (2011). Minimal pair distinctions and intelligibility in preschool children with and without speech sound disorders. *Clinical Linguistics and Phonetics, 25*(10), 853-863. doi:10.3109/02699206.2011.578783

Howe, T. J. (2008). The ICF contextual factors related to speech-language pathology. *International Journal of Speech Language Pathology, 101-2*, 27-37. doi:10.1080/14417040701774824

Hustad, K. C. (2008). The relationship between listener comprehension and intelligibility scores for speakers with dysarthria. *Journal of Speech, Language, and Hearing Research, 51*(3), 562-573. doi:10.1044/1092-4388(2008/040)

Irwin, J., & Wong, S. (Eds.). (1983). *Phonological development in children 18 to 72 months*. Carbondale: Southern Illinois University Press.

Kamhi, A. G. (2006). Treatment decisions for children with speech-sound disorders. *Language, Speech & Hearing Services In Schools*, *37*(4), 271-279.

Kent, R. D., Miolo, G., & Bloedel, S. (1994). The intelligibility of Children's speech: A review of evaluation procedures. *American Journal of Speech-Language Pathology, 3*(2), 81-95. doi:10.1044/1058-0360.0302.81

Kent, R. D., Weismer, G., Kent, J., & Rosenbek, J. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders, 54*, 482-499.

Kim, H., Hasegawa-Johnson, M., & Perlman, A. (2011). Vowel contrast and speech intelligibility in dysarthria. *Folia Phoniatrica et Logopaedica, 63*(4), 187-194. doi:10.1159/000318881

Landa, S., Pennington, L., Miller, N., Robson, S., Thompson, V., & Steen, N. (2014). Association between objective measurement of the speech intelligibility of young people with dysarthria and listener ratings of ease of understanding. *International Journal of Speech Language Pathology, 16*(4), 408-416. doi:10.3109/17549507.2014.927922

Leinonen-Davies, E. (1988). Assessing the functional adequacy of children's phonological systems. *Clinical Linguistics and Phonetics, 2*(4), 257-270. doi:10.3109/02699208808985259

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2007). Speaker-independent factors affecting the perception of foreign accent in a second language. *Journal of the Acoustical Society of America, 121*(4), 2327-2338. doi:10.1121/1.2537345

Marian, V., Bartolotti, J., Chabal, S., Shook, A. (2012). ClearPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE, 7(8): e43230.* doi:10.1371/journal.pone.0043230

Massie, C. (2014a). *extract_april3.py.* [source code] Available at https://github.com/massiecb/Vowel-Errors

Massie, C. (2014b). *averages_group_acc11.* [source code] Available at https://github.com/massiecb/Vowel-Errors

Massie, C. (2014c). *averages_acc.* [source code] Available at https://github.com/massiecb/Vowel-Errors

Massie, C. (2014d). *averages_con.py.* [source code] Available at https://github.com/massiecb/Vowel-Errors

McLeod, S., & Baker, E. (2014). Speech-language pathologists' practices regarding assessment, analysis, target selection, intervention, and service delivery for children with speech sound disorders. *Clinical Linguistics & Phonetics, 28*(7-8), 508-531. doi:10.3109/02699206.2014.926994

Miller, N. (2013). Measuring up to speech intelligibility. *International Journal of Language & Communication Disorders, 48*(6), 601-612. doi:10.1111/1460-6984.12061

Monsen, R. B. (1976). Normal and reduced-phonological space: The production of English vowels by Deaf adolescents. *Journal of Phonetics, 4*, 189-198.

Neel, A. (2008). Vowel space characteristics and vowel identification accuracy. *Journal of Speech, Language, and Hearing Research, 51*(3), 574-585.

Otomo, K., & Stoel-Gammon, C. (1992). The acquisition of unrounded vowels in English. *Journal of Speech and Hearing Research,* (3), 604.

Paschall, L. (1983). Development at 18 months. In J. Irwin, & S. Wong (Ed.), *Phonological development in children* (pp. 27-54). USA: Southern Illinois University.

Podlubny, R. (2014). *Effortful devoicing: Word-final obstruents in Western Canadian English.* Unpublished manuscript, Department of Linguistics, University of Alberta, Edmonton, Canada.

Pollock, K. E. (1991). The identification of vowel errors using traditional articulation or phonological process test stimuli. *Language, Speech & Hearing Services in Schools, 22*, 39-50.

Pollock, K. E. (2002). Identification of vowel errors: Methodological issues and preliminary data from the Memphis Vowel Project. In M. J. Ball, & F. E. Gibbon (Eds.), *Vowel disorders* (pp. 83-113) Butterworth-Heinemann.

Pollock, K. E. (2013). The Memphis Vowel Project: Vowel errors in children with and without phonological disorders. In M. J. Ball, & F. E. Gibbon (Eds.), *Handbook of vowels and vowel disorders* (1st ed., pp. 260-287). New York, NY: Psychology Press.

Pollock, K. E., & Berni, M. C. (2001). Transcription of vowels. *Topics in Language Disorders, 21*(4). pp. 22-40.

Pollock, K. E., & Berni, M. C. (2003). Incidence of non-rhotic vowel errors in children: Data from the Memphis Vowel Project. *Clinical Linguistics and Phonetics, 17*(4-5), 393-401. doi:10.1080/0269920031000079949

Pollock, K. E., & Hall, P. K. (1991). An analysis of the vowel misarticulations of five children with developmental apraxia of speech. *Clinical Linguistics & Phonetics, 5*(3), 207-224. doi:10.3109/02699209108986112

Pollock, K. E., & Keiser, N. J. (1990). An examination of vowel errors in phonologically disordered children. *Clinical Linguistics and Phonetics, 4*(2), 161-178. doi:10.3109/02699209008985479

Powell, T. W. (1991). Planning for phonological generalization: An approach to treatment target selection. *American Journal of Speech-Language Pathology, 1*, 21-27. doi:10.1044/1058-0360.0101.21

Read, C. (1971). Pre-school children's knowledge of English phonology. *Harvard Educational Review, 41*(1), 1-34.

Reynolds, J. (1990). Abnormal vowel patterns in phonological disorder: Some data and a hypothesis. *The British Journal of Disorders of Communication, 25*(2), 115-148.

Sander, E. K. (1972). When are speech sounds learned? *Journal of Speech and Hearing Disorders, 37*(1), 55-63.

Shriberg, L. D. (1997). Developmental phonological disorders: One or many. In B. W. Hodson, & M. L. Edwards (Eds.), *Perspectives in applied phonology* (pp. 105-127). Gaithersburg, MD: Aspen.

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997a). The percentage of consonants correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research, 40*(4), 708-722.

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997b). The speech disorders classification system (SDCS): Extensions and lifespan reference data. *Journal of Speech, Language, and Hearing Research, 40*(4), 723-740.

Shriberg, L. D., & Kwiatkowski, J. (1982). Phonological disorders II: A conceptual framework for management. *Journal of Speech and Hearing Disorders, 47*, 242-256. doi:10.1044/jshd.4703.242

Sims, M. (2010). *Scale the intensity of all .wav files in subdirectories (normalize intensity).* [source code] Available at http://ualberta.ca.ca/~mnsims

Skahan, S. M., Watson, M., & Lof, G. L. (2007). Speech-language pathologists' assessment practices for children with suspected speech sound disorders: Results of a national survey. *American Journal of Speech-Language Pathology, 16*(3), 246-259. doi:10.1044/1058-0360(2007/029)

Smith, J. (2012). *Phonetic and perceptual variation of word-final stop voicing in Toronto English* (Doctoral Dissertation). Retrieved from Academia.edu.

Smith, N. (1973). *The acquisition of phonology: A case study*. Cambridge: Cambridge University Press.

Speake, J., Stackhouse, J., & Pascoe, M. (2012). Vowel targeted intervention for children with persisting speech difficulties: Impact on intelligibility. *Child Language Teaching & Therapy*, *28*(3), 277-295.

Stoel-Gammon, C., & Dunn, C. (1985). *Normal and disordered phonology in children*. Baltimore: University Park Press.

Stoel-Gammon, C., & Herrington, P. B. (1990). Vowel systems of normally developing and phonologically disordered children. *Clinical Linguistics and Phonetics, 4*(2), 145-160. doi:10.3109/02699209008985478

Taft, M. & Hambly, G. (1986). Exploring the cohort model of spoken word recognition (English). *Cognition*, *22*(3), 259-282.

Templin, M. C. (1957). *Certain language skills in children*. Minneapolis: University of Minnesota Press.

Templin, M. C., & Darley, F. (1969). *The Templin-Darley Tests of Articulation*. Iowa City, IA: Bureau of Educational Research and Service, University of Iowa.

Tjaden, K., & Liss, J. M. (1995). The influence of familiarity on judgments of treated speech. *American Journal of Speech-Language Pathology, 4*, 39-47.

Vaughn, A., & Pollock, K. E. (1997). *The relative contribution of vowel and consonant errors to intelligibility.* Unpublished manuscript.

Vigouroux, J., & Miller, N. (2007). Intelligibility testing: Issues in closed versus open format scoring. *Newcastle and Durham Working Papers in Linguistics, 12*, 83-95.

Walton, J. H., & Pollock, K. E. (1993). Acoustic validation of vowel error patterns in developmental apraxia of speech. *Clinical Linguistics and Phonetics, 7*(2), 95-111. doi:10.3109/02699209308985547

Watson, K. (2007). Liverpool English. *Journal of the International Phonetic Association, 37*(3), 351-360. doi:10.1017/S0025100307003180

Watts, N. (2004). Assessment of vowels summary. *Acquiring Knowledge in Speech, Language and Hearing, 6*(1), 22-25.

Weismer, G., & Martin, R. E. (1992). Acoustic and perceptual approaches to the study of intelligibility. In R. D. Kent (Ed.), *Intelligibility in speech disorders: Theory, measurement, and management* (pp. 67-118). Amsterdam: John Benjamins.

World Health Organisation, (2001). *International classification of functioning, disability and health (ICF).* Geneva: Author.

Wilcox, K. & Morris, S. (1999). *Children's speech intelligibility measure.* USA: The Psychological Corporation.

Williams, A. L. (2005). From developmental norms to distance metrics: Past, present and future directions for target selection practices. In A. G. Kamhi, & K. E. Pollock (Eds.), *Phonological disorders in children: Clinical decision making in assessment and intervention* (pp. 101-108). Baltimore: Paul H. Brookes Publishing Co., Inc.

Yavas, M., & Lamprecht, R. (1988). Processes and intelligibility in disordered phonology. *Clinical Linguistics & Phonetics, 2*(4), 329-345. doi:10.3109/02699208808985264

Yorkston, K. M., & Beukelman, D. R. (1980). Clinician-judged technique for quantifying dysarthric speech based on single word intelligibility. *Journal of Communication Disorders, 13,* 15–31. doi:10.1016/0021-9924(80)90018-0

Yorkston, K., & Beukelman, D. R. (1981). Ataxic dysarthria: Treatment sequences based on intelligibility and prosodic considerations. *Journal of Speech and Hearing Disorders, 46,* 398-404.

Yorkston, K. M., Strand, E. A. & Kennedy, M. R. T. (1996). Comprehensibility of dysarthric speech: Implications for assessment and treatment planning. *American Journal of Speech–Language Patholog*y, *5*, 55–66. doi:10.1044/1058-0360.0501.55

Zarifian, T., Tehrani, L. G., Salavati, M., Modaresi, Y., & Kazemi, M. D. (2014). The percentage of vowel correct scale in Persian speaking children. *Iranian Rehabilitation Journal, 12*(19), 5-8.

Appendix A

Descriptive Statistics for the Between-Subjects Factors of Participants

Between-Subjects Factors		
		N
SLP Services	Services Reported	3
	No Services Reported	25
Gender	Female	23
	Male	5
Experience Working with Children (Years)	None Reported	14
	1.50	2
	2.00	1
	3.00	1
	6.00	4
	7.00	3
	8.00	1
	10.00	1
	11.00	1
Age (Years)	18	7
	19	5
	20	8
	21	1
	22	2
	23	2
	26	1
	27	2

Appendix B

Multiple Regression Analysis of Between-Subjects Factors for Participants

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.205[a]	.042	-.035	.04395

a. Predictors: (Constant), Years of Experience Working with Children, Age

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.002	2	.001	.549	.584[b]
	Residual	.048	25	.002		
	Total	.050	27			

a. Dependent Variable: Overall Participant Accuracy

b. Predictors: (Constant), Years of Experience Working with Children, Age

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.615	.066		9.261	.000
	Age	-.002	.003	-.147	-.751	.460
	Years of Experience working with Children	.002	.002	.140	.713	.482

a. Dependent Variable: Overall Participant Accuracy

Appendix C

Reported Means and Independent Samples Test for Difference between Participants with or Without Reported SLP Services

Group Statistics

	SLP Services Reported	N	Mean	Std. Deviation	Std. Error Mean
Overall Accuracy	None Reported	25	.5694	.04409	.00882
	Services Reported	3	.5812	.04121	.02379

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Overall Accuracy	Equal variances assumed	.039	.846	-.438	26	.665	-.01175	.02681	-.06686	.04335
	Equal variances not assumed			-.463	2.583	.679	-.01175	.02538	-.10041	.07690

Appendix D

Reported Means and Independent Samples Test for Gender Differences

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Overall Accuracy	Female	23	.5753	.04041	.00843
	Male	5	.5498	.05437	.02431

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Overall Accuracy	Equal variances assumed	.340	.565	1.204	26	.239	.02546	.02115	-.01800	.06893
	Equal variances not assumed			.990	5.005	.368	.02546	.02573	-.04066	.09159

Appendix E

Lexical Details for the Correct Form of Target Stimuli

Target Word	**# of Syllables**	**Source[a]**	**Clearpond (Frequency per million)**	**V #**	**Vowel Neighbours**	**C #**	**Consonant Neighbours**	**Removed Items[b]**	**Total # Phono Neighbours**	**V Pattern[c]**	**C Pattern[d]**
Brown	1	M.B.	60.1176	4	brain; bran; brawn; Brownie;	5	crown; drown; frown; browse; brow;	browns	9	DR	CC
Candy	2	M.B.	35.7843	3	candor; canned; candle;	5	handy; dandy; caddy; bandy; randy	candies Sandy; Caddie/caddy	8	BK	PV
Child	1	CP	157.6471	1	chilled	5	wild; filed; mild; piled; riled;		6	DR	ST
Dress	1	M.B.	87.1961	3	address; dresser; dressy	4	press; dread; dredge; heiress;	dressed Des;	7	BK	CC
Fast	1	M.B.	137.451	7	first; forced; fist; feast; faced; fest; faster;	9	last; past; fact; vast; mast; caste; gassed; fat; faxed	past/passed; Faust; fairest; cast/caste	16	BK	CC
Fruit	1	CP	21.7255	6	freight; frat; fright; fret; fraught; fruity	3	brute; flute; root;	fruits; Groot	9	LX	ST
Grass	1	CP	16.7843	5	grace; gross; grease; grouse; grassy	10	grab; glass; brass; grad; gram; Gran; graph; gas; grasp; crass;	Greece;	15	BK	ST
Kids	1	CP	301.098	6	cards; cords/chords; kiddie/kiddy; kidder; orchids; kiddies;	6	bids; lids; skids; kills; codes; kings;	kid; kiddo; cords/chords; kiddie/kiddy;	12	BK	PV
Knife	1	M.B.	46.8039	0		9	night; life; nice; wife; nine; Nile; Fife; rife; nigh;	knifed Knight; Nike;	9	DR	ST
Please	1	CP	1100.9608	2	plays; plows;	9	plead; fleas; sleaze; plebe/plebes; pees/peas; pleased; plea;	plebe/plebes flees/fleas; pees/peas; plead/pleads;	11	LX	ST
Round	1	CP	66.5294	3	rained; Rand; around;	11	found; sound; bound; pound; hound; mound; downed; ground; drowned; crowned; frowned	rounds;	14	DR	CC
Space	1	CP	66.0588	4	spice; spouse; airspace; Spacey	3	Spain; Spade; pace;	spaced;	7	LX	CC
Spoon	1	M.B.	7.6078	6	spin; Spain; spine; span; spawn; spun;	4	spook; swoon; spool; soon;	spoons; Poon; Spooner;	10	LX	CC

								sewn;			
Tape	1	CP	68.8431	6	top; type; tip; tap; tarp; taupe;	7	shape; rape; Cape; tame; ape; tale/tail; take;	tapes; taped; Tate; tail/tale	13	LX	PV
Think	1	CP	2691.3922	2	Thank; thinker	12	things; pink; sink/sync; link; mink; tink; wink; Chink; zinc; kink; thick; rink; thing;	thinks; Dink; Fink; sink/sync; thingy	15	BK	ST
Tiny	2	M.B.	32.2157	3	teeny; Tawny; Tunney	3	shiny; tidy; whiny;	Tony;	6	DR	PV
Toy	1	M.B.	16.8431	8	to; tea; tie; tour; toe; tore; tar; Tao;	7	boy; joy; soy; coy; Choy; hoy; toil;	toys; toyed to/too/two; tie/Thai/Tai; Hoi; boy/buoy; Tae; joy/Joie; Loy; Tau; Goy; Troy; coy/koi; toe/tow; tore/tor; tee/tea;	15	DR	PV
Treat	1	CP	51.8824	5	trout; trot; trait; trite; treaty;	4	tree; greet; tweet; street;	tree/trees; treats; Crete;	9	LX	PV

[a] M.B. = MacArthur Bates Communicative Development Inventories; CP = Clinician Selected and verified frequency on Clearpond Database

[b] Removed Items included Inflectional Morphemes, Homonyms, Proper names, and Words which were more than one phoneme off

[c] CC = Consonant Cluster Reduction; PV = Prevocalic Voicing; ST= Stopping;

[d] DR = Diphthong Reduction; BK = Backing; LX = Laxing;

Appendix F

IPA Error Forms for All Target Stimuli

Target Item	Consonant Error Type [a]	IPA Form	Vowel Error Type [b]	IPA Form	Combined Error Type [a,b]	IPA Form
brown	CC	/ba͡ʊn/	DR	/bɹɑn/	CC + DR	/bɑn/
candy	PV	/gændi/	BK	/kɑndi/	PV + BK	/gɑndi/
child	ST	/ta͡ɪld/	DR	/tʃɑld/	ST + DR	/tɑld/
dress	CC	/dɛs/	BK	/dɹʌs/	CC+BK	/dʌs/
fast	CC	/fæs/	BK	/fɑst/	CC+BK	/fɑs/
fruit	ST	/pɹut/	LX	/fɹʊt/	ST + LX	/pɹʊt/
grass	ST	/gɹæt/	BK	/gɹɑs/	ST +BK	/gɹɑt/
kids	PV	/gɪdz/	BK	/kʊdz/	PV + BK	/gʊdz/
knife	ST	/na͡ɪp/	DR	/nɑf/	ST + DR	/nɑp/
please	ST	/plid/	LX	/plɪz/	ST + LX	/plɪd/
round	CC	/ɹa͡ʊn/	DR	/ɹɑnd/	CC + DR	/ɹɑn/
space	CC	/se͡ɪs/	LX	/spɛs/	CC + LX	/sɛs/
spoon	CC	/pun/	LX	/spʊn/	CC + LX	/pʊn/
tape	PV	/de͡ɪp/	LX	/tɛp/	PV + LX	/dɛp/
think	ST	/tɪŋk/	BK	/θʊŋk/	ST +BK	/tʊŋk/
tiny	PV	/da͡ɪni/	DR	/tɑni/	PV + DR	/dɑni/
toy	PV	/dɔ͡ɪ/	DR	/tɑ/	PV + DR	/dɑ/
treat	PV	/dɹit/	LX	/tɹɪt/	PV + LX	/dɹɪt/

[a] CC = Consonant Cluster Reduction; PV = Prevocalic Voicing; ST= Stopping;
[b] DR = Diphthong Reduction; BK = Backing; LX = Laxing;

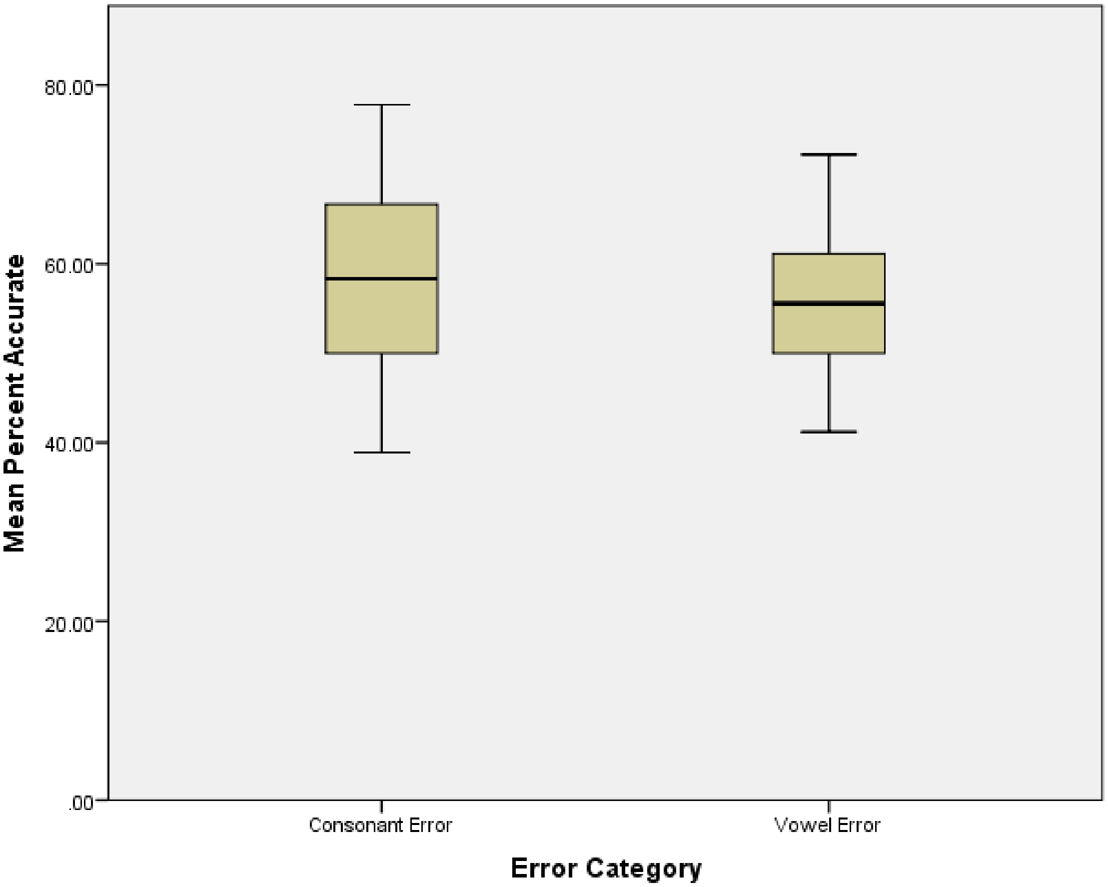
Appendix G

Accuracy of Lexical Target with All Productions Combined

Report			
Accuracy Of Word (All Productions Combined)			
Answer	Mean	N	Std. Deviation
Candy	61.607	28	24.9835
Fruit	60.714	28	23.9874
Grass	58.929	28	22.7855
Toy	58.929	28	25.6529
Knife	58.036	28	23.6228
Brown	58.036	28	23.6228
Please	58.036	28	25.5074
Spoon	57.143	28	25.3285
Tape	57.143	28	23.4295
Fast	56.25	28	23.199
Round	56.25	28	21.1093
Tiny	56.25	28	25.1155
Treat	56.25	28	23.199
Think	55.357	28	23.9184
Child	54.464	28	21.5741
Dress	53.571	28	23.2879
Kids	52.679	28	22.9122
Space	43.75	28	19.9826
Total	56.300	504	23.4616

Appendix H

Boxplot of Mean Percent Accuracy for Vowel and Consonant Only Error Patterns with the Inclusion of Tiny *and* Candy



Appendix I

High-to-Low Ranking of Accuracy from the Friedman One-Way Within Subjects Test for the Single Substitution Error Patterns with the Inclusion of Tiny *and* Candy

Ranks

	Mean Rank
Prevocalic Voicing	5.45
Diphthong Reduction	3.70
Backing	3.54
Laxing	3.16
Stopping	2.59
Consonant Cluster Reduction	2.57

Test Statistics[a]

N	28
Chi-Square	52.721
df	5
Asymp. Sig.	.000

a. Friedman Test

Figure F. Prevocalic Voicing had the highest ranking accuracy with the inclusion of *tiny* and *candy*. Backing and Laxing switched relative ranking but continue to be ranked equally with respect to the other single substitution error patterns.

Appendix J

High-to-Low Ranking of Confidence from the Friedman One-Way Within Subjects Test for All Individual Error Patterns

Ranks For Confidence

	Mean Rank
Prevocalic Voicing + Backing	12.78
Consonant Cluster Reduction	11.04
Laxing	10.48
Prevocalic Voicing	9.20
Backing	9.11
Stopping	9.00
Stopping + Backing	8.24
Stopping + Diphthong Reduction	7.65
Consonant Cluster Reduction + Laxing	7.65
Prevocalic Voicing + Laxing	7.24
Diphthong Reduction	6.98
Consonant Cluster Reduction + Diphthong Reduction	6.98
Consonant Cluster Reduction + Backing	5.69
Stopping + Laxing	4.19
Prevocalic Voicing + Diphthong Reduction	3.78

Test Statistics[a]

N	27
Chi-Square	123.981
df	14
Asymp. Sig.	.000

a. Friedman Test