The Validity of Diphthong Durations as a Marker for Speech Disorders

Kathleen Russell

University of Amsterdam

Table of Contents

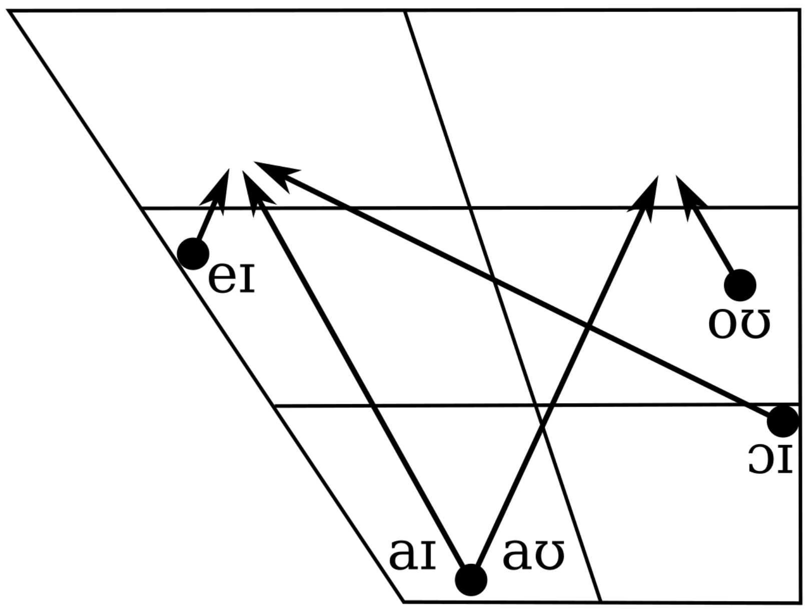
1. Introduction

Speech impediment is a growing problem in children post-pandemic (Khan, Freeman, & Druet, 2023). The present thesis attempts to contribute to research and diagnostics in the speech impediment field by providing measurements and thresholds for standard and disordered diphthong production. I wanted to perform an acoustic segment analysis, and due to the scope of this thesis, I decided to focus on duration measurement. I decided to broaden the field by focusing on diphthongs, which are less frequently studied than consonants or monophthongs.

1.1. Diphthongs

Diphthongs are traditionally defined as being a single phoneme with an onset and offset that match the articulatory properties of two different vowels with no consonant or glottal stop in between. For example, English diphthong /aɪ/ begins with the tongue in the position for /a/ and then gradually moves to the position for /i/. Acoustically, this manifests as formant transitions where the first and second formants will gradually increase or decrease in Hertz. Traditionally, /aɪ/, /aʊ/, /eɪ/, /ɔɪ/, and /oʊ/ are considered the five diphthongs of General American English.

Figure 1.1.1: General American English Diphthongs



https://commons.wikimedia.org/wiki/File:General_American_diphthong_chart.svg

Peterson (2018) outlines various definitions of diphthongs. Some definitions of diphthongs exclude /eɪ/ and /oʊ/ because their endpoints are less contrastive and their durations are shorter. Some argue that in English, /eɪ/ and /oʊ/ are the tense counterparts of /ɛ/ and /ɔ/ respectively, because English does not have the traditional monophthong /e/ and /o/ present in some other languages, and that this means they are not phonologically diphthongs. For the purpose of this thesis however, /eɪ/ and /oʊ/ are both considered diphthongs.

2. Background

I will now present some background information in speech acquisition, speech disorders, and the study of phoneme quality and durations, which will aid in comprehension of my own study on phoneme durations.

2.1. Vowel acquisition

Phonological acquisition begins in infancy as long as the child is exposed to language they can see, hear, or touch. As the focus of this thesis is on oral English diphthongs, this section on speech acquisition will focus on oral speech as acquired by hearing children exposed to aural language from birth.

Hearing babies hear the sounds and speech of the world around them, and their brains start to pattern the sounds into phonemes. Babies can learn to distinguish and understand oral speech earlier than they are able to produce it (Bergelson & Swingley, 2012).

By six months, children start to exhibit language-specific characteristics in their babbled vowels (Lyakso & Silvén, 2002) and vowel perception (Werker & Tees, 1984, Kuhl, 1993). This means that children are more frequently producing vowels that match the articulation qualities of their language's or languages' vowels, as well as producing contrasted vowels that reflect their language's or languages' contrasts. For example, a child acquiring English may produce a vowel like /oʊ/ with much higher occurrence than /o/, because English does not have /o/ as a monophthong, but does have /oʊ/. However, another language, perhaps one that does not have diphthongs but does have monophthongs, would more likely see children at this age producing exclusively monophthongs.

2.1.1 Error patterns

English babblers seem to prefer vowels that are more front and low (MacNeilage & Davis, 1990, Davis & MacNeilage, 1995). Crosslinguisticially, children seem to acquire accurate vowel production before accurate consonant production (Donegan, 2013). Bernhardt & Stemberger (1998) identify closed syllables containing diphthongs, such as "coin" /kɔɪn/, as potentially being more challenging for children to produce. Kehoe & Stoel-Gammon (2001)'s longitudinal subjects successfully produced more consonants after lax vowels, such as /ɛ/ in "bed" /bɛd/", than tense vowels, such /i/ in "bead" /bid/.

Some children's errors involve replacing a phoneme with another phoneme that's different in one or more articulatory properties, such as replacing the lax mid front vowel /ɛ/ with the lower lax front vowel /æ/ Donegan (2013) posits that children prefer to lower vowels in order to increase their sonority. This lowering "applies most strongly to non-palatal non-labial vowels such as /ɨ/ and /ə/, and accounts for their absence in many of the world's languages (Donegan, 2013)." Conversely, raising the vowel decreases sonority (Donegan, 2013). Donegan (2013) notes that cross-linguistically, "only palatal and labial vowels are raised, and tense vowels are favo[u]red for raising".

Even though there are statistical trends in certain patterns of phonological realization, ultimately, each child will have its own pattern. Donegan (2013) notes diphthong monophthongization, lowering, raising, consonant-specific diphthong pre-insertion, tensing, and laxing all being observed in children's error patterns. There is no single way all children pattern (Donegan, 2013).

2.1.1. Development milestones

Kehoe & Stoel-Gammon (2001) looked at the development of phonological distinctions between tense and lax vowels in English. They label lax vowels as "short" and tense as "long". They focused on English-speaking children aged 1;3 to 2;0 from Seattle, USA who have the cot-caught merger—i.e., /ɑ/ and /ɔ/ have merged into /ɑ/. Occasionally, subjects produced /ɔ/, but for the purpose of the study, those productions were coded as the tense-long /ɑ/. Diphthongs "pattern together with tense vowels" as being long (Kehoe & Stoel-Gammon, 2001).

Kehoe & Stoel-Gammon (2000) noted that "no child exhibited equal frequencies of short for long or long for short vowels". This means that every participant, when making substitutions errors, was *either* inclined to sub tense for lax vowels *or* lax for tense vowels.

Eight of Kehoe & Stoel-Gammon (2001)'s ten subjects appeared "to be at the latter stage of rhyme development (Kehoe & Stoel-Gammon, 2001)". The children successfully produced codas "in the majority of target productions (Kehoe & Stoel-Gammon, 2001)" and produced "the correct target vowel length most of the time (Kehoe & Stoel-Gammon, 2001)". The other two children did "not fit the predicted patterns (Kehoe & Stoel-Gammon, 2001)". They produced codas less than 20% of the time but produced vowels of the correct length over 90% of the time (Kehoe & Stoel-Gammon, 2001). Kehoe & Stoel-Gammon (2001) conclude with this evidence that there is "little evidence that coda consonant development precedes acquisition of the vowel length contrast".

They posit, based on the two children who produced few codas but successfully produced most of their vowels, that "children seem to control vowel length earlier than coda position" (Kehoe & Stoel-Gammon, 2000). Salidis & Johnson (1997) came to the same conclusion, in tandem with research by Fikkert (1994) and Femuth & Fee (1995), that English-acquiring children cannot control their vowel output in terms of *length*, i.e. lax-tense distinctions, until they learn to produce coda consonants.

Turunen's 2003 study assessed the vowel realisations of 193 Finnish children aged 2;6. Finnish is a language with diphthongs, short monophthongs, and contrastive long monophthongs (Turunen, 2003). For example, Finnish has a diphthong /yø/ (as in [pyøræ] pyörä 'bike'), where the transition is from a *high* front rounded vowel to a *high-mid* front rounded vowel. Three of the 193 children failed to produce any of the diphthongs correctly, while 60 of the 193 children consistently reached a target-like production of all the diphthongs (Turunen, 2003). Eight to 10% of the children produced a long vowel instead of a diphthong, for example, producing [pyːræ] for pyörä 'bike' instead of the target [pyøræ] (Turunen, 2003).

Kulju & Savinainen-Makkonen (2008) described the middle Finnish vowels /y/ and /ø/, and the front Finnish vowel /æ/ as "the most challenging" for children to learn to reliably produce. They also conclude that "at a certain stage, a child may not be able to produce diphthongs and complex consonant sequences in the same output (Kulju & Savinainen-Makkonen, 2008)."

Kehoe's 1998 study concluded that young children acquiring English are inclined to shift primary stress to syllables with long vowels or diphthongs. Kallay et al.'s 2022 study on 5–9 year olds displays that "speech rhythm continues to develop during the school-age years." Young English-speaking children are more inclined to produce equally-stressed speech rhythms than older English-speakers (Allen & Hawkins, 1980, Grabe et al., 1999). Kallay et al., 2022)'s study suggests that the early school-age years experience substantial "developmental changes in the articulation rate of narrative speech [. . .] with the largest increases occurring between the ages of 7 and 9 years".

2.2. Speech disorders

An estimated 5% of American children age 3–17 have a speech disorder (Black et al., 2015). Speech impediment can impair one's ability to communicate and be understood by others, and can furthermore lead to bullying, stigma, and other negative interactions (Gibbon, 2013). According to Gibbon (2013), there is a lack of research and practices developed to target disordered vowels in speech therapy. Gibbon and Beck's 2002 research on speech therapy outlooks concluded that vowel-focused speech therapy can improve vowel production.

This section will overview a variety of sources of speech impairments, including phonological impairments, apraxia of speech, and prosodic disorders.

2.2.1. Phonological impairments

One source of disordered vowel production is phonological disorders. These are disorders in the mental mapping of a child's phonology. For example, if a child fails to map a distinction between two phonemes that are distinct in the language they are acquiring, then that would be a phonological disorder.

Stokes, Lau & Ciocca (2002) assessed English children's production based on tongue movements. The participants were aged 3;0 to 7;3, and 70% of the errors they produced were "simple reductions", which means that the diphthongs were replaced with a monophthong (of the properties of the first half of the diphthong, i.e. /aj/ to [a]). Pollock and Keiser (1990) "discovered that reduction of diphthongs is the most common error pattern, which occurred in all nonrhotic diphthongs".

Torvelainen (2007) found significant correlation in Finnish children between poor phonology at age 3;6 and poor reading skills at age 7. The poor phonological skills were quantified by assessments of "naming, repetition of non-words and phonological processing" (Torvelainen, 2007). The children with the weakest phonological skills "attempted and produced fewer words than the more advanced children" (Torvelainen, 2007). Torevelainen (2007) suggests that there may be a "link between lexicon and the children's phonological development".

Stokes, Lau & Ciocca (2002) analyzed diphthong errors in Cantonese-speaking children with phonological disorders. Cantonese has eleven distinct diphthong phonemes, including the five diphthongs that appear in American English, /aɪ/, /aʊ/, /eɪ/, /ɔɪ/, and /oʊ/ (Zee, 1999). Stokes, Lau & Ciocca (2002)'s "[p]erceptual analysis showed that /ɔɪ/ and /uɪ/ were most frequently in error, whereas /eɪ/, /oʊ/, and /aʊ/ were least frequently in error." This finding could indicate that children may have more trouble producing diphthongs with a movement from front to back (i.e. /ɔ/ to /ɪ/ in /ɔɪ/) than back to front (i.e. /a/ to /ʊ/ in /aʊ/) or along the open-close spectrum (i.e. /eɪ/ and /oʊ/). "The diphthong /eɪ/ was the most accurately produced of all diphthongs, whereas /ɔɪ/ was the least accurately produced (Stokes, Lau & Ciocca, 2002)."

Wellman et al.'s 1931 study found preliminary data on which English vowels are more challenging to acquire than others. Overall, tense monophthongs (along with /ʌ/ seemed to be mastered by age 2, diphthongs and some lax vowels—/ɛ/ and /ɔ/—were mastered by age 3, and the rest of the lax vowels were mastered by age 4 (Wellman et al., 1931). However, Lieberman's 1980 longitudinal study identified English lax /æ/ and /ɛ/ as the first vowels mastered.

Later, Pollock & Keiser (1990) identified English /i/, /u/ and /ɔ/ as having the highest frequency of being produced correctly by their 3–7 year old subjects. /aʊ/, /æ/ and /ʊ/ were identified as being produced correctly the fewest number of times. Additionally, they identified that most of their subjects had difficulties with rhotic vowels and diphthongs. Pollock & Keiser (1990) found that errors often included lowering and backing, but seldom fronting: frequent errors "included [ɛ] for /ɪ/, [ɑ] or [a] for /æ/, [ɑ] for /ʌ/, [ɑ] for /a͡ɪ/ or /a͡ʊ/, and [o͡ʊ] for /ɔ͡ɪ/ (Pollock & Keiser (1990)."

Overall, there is variation in the order of acquisition for vowels. Wellman et al. (1931) and Pollock & Keiser (1990) are consistent in identifying /i/ and /u/ as being two of the first or easiest English vowels mastered and /ʊ/ as being one of the last or most difficult vowels mastered.

2.2.2. Apraxia of speech

Apraxia of speech is a condition where someone has an impairment in speech production due to difficulty with articulative motor accuracies (Guyette & Diedrich, 1981, Allison et al., 2020). This includes childhood or developmental apraxia of speech, which manifests in childhood (and can dissipate in adulthood either with therapeutic intervention or on its own), and acquired apraxia of speech, which manifests in adulthood, often the result of an injury to the head or brain (Allison et al., 2020). Apraxia is a contrastive diagnosis from other neurological-rooted conditions and motor execution impairments such as phonological disorder, aphasia, or dysarthria, as it is specific to motor impairment in linguistic articulators, without affecting mental maps of sound production (Allison et al., 2020).

However, apraxia is often comorbid with aphasia (in adults), specific language impairment (in children), and motor control impairment, so it's difficult to disentangle the two (Allison et al., 2020). Apraxia is a diagnosis specific to *speech articulation*. Aphasia is more strongly linked with word-finding and interpreting difficulties. Phonological impairment contrasts from apraxia by being evident in the mental representations of the phonemes. Motor control impairments are evident in areas of the body other than the linguistic articulators.

Walton & Pollock (1992) studied English speaking children aged 8;2 to 10;9 who have apraxia of speech. Amongst the subjects' productions, /aʊ/ and /ɔɪ/ had a slightly higher proportion of reduction (43% and 42% respectively) than /aɪ/ (32%) (Walton & Pollock, 1992). One of Walton & Pollock (1992)'s five subjects reduced all of their diphthongs across every production.

Kulju & Savinainen-Makkonen (2008) summarized findings in various disordered groups and their first language acquisition of Finnish. Pietarinen (1987) studied 1600 Finnish 5-year-olds, finding that up to 30% may have had articulation disorders, but only 1.4% of those children had problems with vowels. The vowel problems featured substitutions of vowels for other vowels. Mäenpää (1990) found that Finnish children with developmental dyspraxia of speech had the most difficulty with the "highest tongue postures".

Shriberg et al. (1997) identified non-corner English vowels as being more challenging to produce for English speakers with childhood apraxia of speech than in speech delay.

In childhood apraxia of speech, compared to apraxia of speech, "vowel impairment appears to be more central to the diagnosis and indeed may represent a core characteristic for differential diagnosis, although it is not currently included in the ASHA (2007) guidelines (Jacks et al., 2013)". Children with childhood apraxia of speech are often different from each other in the patterning of their vowel errors (Jacks et al., 2013): for example, some children may be inclined to realize monophthongs as diphthongs, while others may be inclined to realize diphthongs as monophthongs.

2.2.3. Prosodic disorders

Rhythm and intonation are important parts of having fluent-sounding speech. Disordered prosody can cause more disordered-sounding speech (Olejarczuk & Redford, 2013; Redford et al., 2018; Shriberg et al., 2001, Kallay, et al., 2022). Improving disordered prosody can improve one's ability to have positive interactions and relationships, as well as improving speech intelligibility (Kallay, et al., 2022, Hawthorne & Fischer, 2020, Kalathottukaren et al., 2015, Peppé, 2009).

Henry (1990) noted that children with speech disorders often also have a deficit in non-linguistic rhythm skills. For example, a disordered child may struggle to produce stress-timed prosody patterns and instead produce equal-timed speech. That same child may also have a hard time producing musical rhythms that involve a variety of note durations. i.e. they may find a rhythm like *fastfastfast slow slow* more challenging than a steady *slow slow slow slow*.

Disordered prosody has been identified in childhood apraxia of speech (Connaghan & Patel, 2012), as well as other conditions outside the scope of this thesis: Down syndrome (Kent & Vorperian, 2013, Stojanovik, 2011, and Wilson et al., 2019), autism (Redford et al, 2018), dysarthria (Patel et al. 2012), and hearing impairment (Chin et al, 2011, and Parkhurst & Levitt, 1978).

2.2.4. Other conditions

Kulju & Savinainen-Makkonen (2008) also summarize findings for disorders outside the scope of this thesis. Vilkman et al,'s 1988 study of Finnish 5 and 8 year olds with fragile X syndrome found that occasionally long vowels were realized as short, but no short vowels were realized as long. Hattunen's 2001 study of 10 Finnish children aged 4–6 with "moderate hearing impairment" concluded that these children had twice as many errors in vowel realization as "normally hearing" children aged 3. These errors included substitution of vowel features, such as substitution front for back. Hattunen's 2000 study of "severely" and "profoundly" hearing-impaired Finnish children found that these children make Finnish vowel harmony violations as well as nasalization and length errors.

Recent studies have also investigated linguistic manifestations of anxiety disorders. Teferra et al. (2022) identified which features of adult speech correlated with anxiety. They assessed voice pitch, total amount of speech/words, articulation rate, and a variety of other acoustic measures. They also assessed the semantic content of participants' speech, discovering a correlation between certain negative and positive keywords and the presence or lack thereof of anxiety. Baird et al. (2020) attempted to predict anxiety by assessing vowel durations.

2.2.5. Error production, identification, and repair

Often when an error is produced in speech, the speaker repeats themself with a correction for the error. For example, one might utter, "She shells sea sells—she *sells* sea *shells*." In this error, the speaker switches around the onsets of the words due to influence from the neighbouring onsets. After speaking, the speaker then corrects their utterance by producing the correct pronunciation of what they were trying to say.

Error repairs often contain stress, which lengthens durations. In the present study, if a child produced an error and then a repair, neither production was included in the data due to this duration lengthening effect.

The following section outlines research into the patterning of these error productions, identifications and repairs.

Shattuck-Hufnagel (1992) conducted an "error-elicitation experiment" on non-speech-disabled native adult speakers of American English using tongue twisters. She concluded that speech production errors in American English occur most frequently in word onsets. Shattuck-Hufnagel (1992) also posits that assimilative speech errors may be further restricted by position (i.e., the onset of one word influences the onset of another word in the utterance, but an onset cannot influence a medial consonant).

Nooteboom & Quené (2015) assessed the rates of different kinds of substitution errors in Dutch. They concluded with statistical significance that 50% of consonant errors were in the onset, while only 39% of the consonants in the corpus were in onset position. This means that beyond the natural chance based on the distribution of consonants in different positions, errors occur more frequently in the onset position, in Dutch. They additionally concluded that "[a]ll types of segments suffer from substitutions, but generally only consonants, not vowels, suffer from addition and omission." In the entire corpus, there was only one instance of a speech error in an article (Nooteboom & Quené, 2015), even though according to Nooteboom (1973) 14% of words in Dutch speech are articles.

Nooteboom & Quené (2021) investigated the time it takes between a speaker producing an error to producing a repair for the error. When a speaker makes a mistake, sometimes they notice the mistake *before* uttering it (i.e. internally), while other times they notice the mistake after utterance (i.e. external speech) (Levelt et al. 1999, Hartsuiker et al. 2005, Nooteboom & Quené, 2021). The duration between uttering the mistake and uttering the repair has been linked to whether the error was detected internally or externally (Nooteboom & Quené, 2017, 2021).

Guenther (2016, chapter 1) outlines three steps of speech planning: (1) abstract phonemes, (2) auditory perceptual targets, and (3) speech motor commands. (1) *Abstract phonemes* refers to the stage in your mind when your brain puts together the phonemes of what you want to see. (2) *Auditory perceptual targets* then refers to when the brain translates those phonemes into what the speaker wants the listener to hear. This is where the speaker may apply rhythmic intonations for various dramatic effects. (3) *Speech motor commands* is then when the brain sends the instructions to the articulation muscles to move in a specific way to produce the speech.

In terms of errors, if someone noticed a mistake before uttering it, then that mistake is in the abstract phoneme stage or auditory perceptual target stage. i.e., the brain has formed the wrong target and will thus send the wrong instruction to the muscles in phase 3. If the mistake is noticed after production, then that means there was an error either in phase 3, or after phase 3, where the muscles might have received the right instructions but failed to articulate correctly due to some other factor (like perhaps the tongue's starting position wasn't where the speaker thought it was).

Nooteboom & Quené's 2021 study elicited speech errors from participants to analyse the time between uttering the error and uttering the repair. In their first experiment of two, amongst elicited errors that were repaired, there were two peaks in repair time: 146 milliseconds and 398 milliseconds. Their second experiment—with a focus on whether the errors are more phonetic or more phonological—also found two peaks in time-before-repair: 206ms and 646ms.

Nooteboom & Quené (2021) featured two experiments. The first experiment showed that "voicing contrast is relatively weak in Dutch" and that "the contrast between vowels is relatively strong in Dutch" (Nooteboom & Quené, 2021). The second experiment showed that perhaps "speech errors are detected on the basis of more phonetic than phonological contrast" (Nooteboom & Quené, 2021). Experiment 2 allowed Nooteboom & Quené (2021) to conclude that "detection of segmental speech errors involving a weak contrast takes more time than detection of segmental speech errors involving a stronger contrast" (Nooteboom & Quené, 2021).

2.3. Phoneme durations

This thesis focuses on measuring diphthong durations with the goal of distinguishing speech-impaired from non-speech impaired children. This section will provide background into previous research in phoneme durations.

2.3.1. Establishing standards

Nooteboom & Slis's 1972 study on Dutch vowels yielded duration values for each vowel in different positions in a word. The purpose of this study was to assess the rationality behind assigning long and short vowel markers in transcription (Nooteboom & Slis, 1972). The participants of their study—the two authors themselves and one other "phonetically trained member of [their] group"—were required to produce a nonsense utterance of the format /pVpVpVp/, where V is a variable to be filled with each vowel studied in turn (Nooteboom & Slis, 1972). For example, the data for the Dutch diphthong /ɑu/ features the participants saying /pɑupɑupɑup/ (Nooteboom & Slis, 1972). The first syllable was to always be unstressed (Nooteboom & Slis, 1972).

Overall, the acoustic data was consistent with the intended production of an unstressed first syllable (Nooteboom & Slis, 1972). The vowel in the first syllable was always shorter in duration than the vowels in the second and third syllables (Nooteboom & Slis, 1972).

They concluded that Dutch diphthongs and Dutch long monophthongs are similar lengths and should be coded with the same length markers (Nooteboom & Slis, 1972). This finding is consistent with Kehoe & Stoel-Gammon (2001)'s finding that diphthongs and tense vowels pattern similarly in children's error patterns. These findings are a supporting argument for considering diphthongs to be a single segment, rather than two separate vowel segments adjacent to each other (Nooteboom & Slis, 1972).

Gay (1970) assessed phoneme identification as well as comparing whether the more important identification queue is the diphthong's endpoints (i.e. the formant values at the beginning and end of the diphthong) or its transition (i.e. the rate of change). His study focused on /ɔɪ/, /aʊ/, and /aɪ/ because they tend to be distinct in American English. All subjects were second-generation born-raised in New York and spoke in New York dialect (Gay, 1970).

Gay (1970) concluded that English diphthongs are identified by their duration more than the formant values of the end points. This means that the specific formant values at the beginning and end of the diphthong are not as important to perception as the overall duration of the diphthong and its formant trajectory.

For contrasting /ɔ/ and /ɔɪ/ by duration, the shift from /ɔ/ to /ɔɪ/ occurs at 170 milliseconds (Gay, 1970). The shift from /ɛ/ to /ɔɪ/ is earlier, occuring at 130 milliseconds (Gay, 1970). The boundary between /a/ and /aɪ/ is at 180 milliseconds, and the boundary between /a/ and /aʊ/, described by Gay as "rather sharp", is at 150 milliseconds (Gay, 1970).

2.3.2. Phoneme durations in kids

In my previous research, Russell (2023), I attempted to contribute to research on non-speech disabled children's diphthong durations. Russell (2023) referenced Lee et al.'s 2014 research into diphthong lengths, which was a broader study of many age groups, while my research focused on specifically 7;0 to 8;0 year olds. Russell (2023) differed from Lee et al (2014) by having a much smaller sample size (five children, all between ages 7;0 to 8;0), and by annotating vowel boundaries by hand. Annotating vowel boundaries by hand allowed me to avoid machine error by personally listening to every single diphthong and viewing their spectrograms. Additionally, annotating by hand allowed me to more easily identify diphthongs which were untranscribable, usually for the reason of poor audio quality. Automatic transcription systems could possibly attempt to ascribe incorrect values to such diphthongs, influencing the results of the study.

Russell (2023) studied children that had been specifically identified as typically developing (TD). However, the subjects were from a study on narrative production in epilepsy, so the labelling may not have been so strict in terms of phonetic development. Russell (2023) yielded means that were statistically within the close range of Lee et al (2014)'s means for 7-year-olds.

2.3.3. Speech rate and other effects on duration

Standardizing diphthong durations requires consideration of many factors other than speech impediment. Durations can change easily due to infinite extralinguistic features. For example, I can say "cheese" with a very stretched vowel if I want: /tʃiiiiiiiiiiiiiiiiiz/. This doesn't change the meaning of the word. While languages differ in respect to how much room there is for speakers to change durations for various effects, English in particular can have large variation in intonation. English is a stress-timed language, which means the pacing of our speech is based on what words in the utterance take stress, and which words take stress is based on communicative intent rather than phonetic rules (Liu & Takeda, 2021).

This means that one speaker can produce diphthongs with a wide variety of durations depending on communicative circumstance. Additionally, at a thankfully more predictable level, durations of vowels are affected by words' standard stress patterns and by their position in a syllable relative to other segments.

Berns & Jacobs (2012) focused on the impact of surrounding consonants to the realization and perception of vowels—specifically on the Dutch coda /l/ (as well as a bisyllable with open coda in the first syllable and an onset /l/ in the second syllable). They discovered that the coda /l/ does not have an impact on the forments of Dutch tense vowels, but does effect F2 and F1 of diphthongs, so far as even possibly leading the diphthong to be "realized as a monophthong" (Berns & Jacobs, 2012).

Diphthongs' durations increase in the bisyllabic utterances (i.e., [l] comes before the diphthong]). However, this trend could be explained by the stress patterns in the bisyllabic utterances rather than the positioning of the [l] in coda or onset. In the instance of the two participants with very short vowels in the monosyllable, they were inclined to neutralize the diphthongs into monophthongs (/ɑu/ to [ɑ] and /ɛɪ/ to [ɛ]).

In addition to word and syllable structures, there are non-linguistic factors into a segment's duration. Eaton & Ratner (2013) assessed children's reactiveness to being told to "speak clearly". Their study featured an invigilator modelling speech styles for the participant, and then assessed how closely the participants adopted that speech (Eaton & Ratner, 2013). The subjects were asked to help a puppet named Penelope "speak better" by modelling speech for Penelope. They discovered "a significant effect of hearing fast versus slow speech", and they noticed that "3-year-olds produced far fewer phonological reductions" than 4-year-olds did (Eaton & Ratner, 2013). They ultimately concluded that "children as young as 3 years of age do imitate the articulatory characteristics of a model without cueing" (Eaton & Ratner, 2013).

Mefferd (2017) assessed the displacement of the tongue and jaw in four different speech styles: typical, slow, loud, and clear. They were interested in the production of English diphthong /aɪ/ in kite. They do not discuss if there was any presence of Canadian Raising, the phenomenon in North American English where /aɪ/ and /aʊ/ are realized as [ʌɪ] and [ʌʊ] before voiceless consonants. Canadian Raising is typical in Canada, but affects speakers in the United States more variably.

Mefferd (2017) concludes that all participants "increased their diphthong durations from typical to slow speech". "Slow speech" had the longest durations by around double the mean durations of "clear speech". "Loud speech" had slightly shorter durations than "clear speech". "Typical speech" had the shortest durations.

Mefferd (2017) also assessed formant transitions and compared their results to previous results yielded by Tasko & Greilick (2010). Tasko & Greilick's 2010 study yielded a formant transition duration increase of 50ms during clear speech. Mefferd (2017) concluded however that formant transition duration only increased by an average of 26ms.

Mefferd (2017) reports that "acoustic vowel contrast tended to be smaller during clear speech than during slow speech, whereas tongue composite movements were rather similar for the two speech conditions".

Gay (1978) assessed the acoustic properties of [i] under different stress conditions. His results showed that the vowel's durations and loudness increase when produced with stress. F0 also seemed to increase, but F1 and F2 both variably experienced increases *and* decreases from unstressed to stressed.

Nooteboom concluded in his 1997 paper on prosody that "the more syllables preced[ing] the stressed syllable, the shorter the stressed syllable becomes".

Lee et al. (2019) attempted to quantify how acoustics change in voices (both within and between speakers), and how they stay the same. Their study measured formants and harmonics. The participants were adults aged 18–29 from the University of California. They found that "variability [. . .] in source spectral shape and spectral noise [accounts] for 18% and 20% of variance across females and males, respectively" (Lee et al. 2019). Second to spectral shape and noise, formant frequency accounted for 11% of variance in female participants while spectral slope accounted for 10% of variance in male participants. Third, in female participants, spectral shape accounted for 10%, and in male participants, formant frequencies accounted for 9%. F0 only accounted for 6% and 7% of variance in female and male participants respectively.

Lee et al. (2019) summarized trends in findings in other papers to conclude that individuals with conditions such as ataxic dysarthria (Kent et al., 1979), apraxia of speech (Collins et al, 1983), aphasia (Baum et al., 1990) or apraxia (Nijland et al., 2003) have an inclination to have longer vowel durations than typically developing controls.

According to Peter & Stoel-Gammon (2005), children with childhood apraxia of speech have less distinction in the durations of their tense and lax vowels than do typically developing children. The two participants with childhood apraxia of speech produced longer vowels on average than the typically developing controls.

2.4. Technical points in duration analysis

Analyzing segment durations requires accurate-to-the-millisecond timestamp labelling as well as bulk processing of statistics.

Adi et al. (2016) sought to find an objective, fast/automated way to measure vowels to counteract the variability and time involved in subjective manual annotation. They've created a software that they claim is the most accurate automated vowel duration measurer to-date, which I considered using, but chose not to because I wanted to make sure I could avoid machine error. No matter how accurate they claim to be, in order for me to assess their accuracy, I would have to manually go through the data anyway.

Howard and Heselwood (2013) describe a standardized method of measuring vowel durations in which the duration is measured from the onset to the offset of the second formant. They also indicate that transcriptions of disordered speech should always use phonetic notations, rather than phonological. In order to maintain this consistency, in the present thesis, target productions are transcribed phonologically, and subject's realizations are transcribed phonetically but rather broadly.

3. The Present Study

This thesis focuses on the analysis of diphthong segment durations in speech produced by children identified as having a speech sound disorder (and comparing it to typically developing counterparts).

3.1. Data

The data used in this thesis is sourced from existing audio samples collected by PERCEPT-R (Benway et al, 2022) and PERCEPT-GFTA (Benway et al, 2023). Both corpora collected audio samples from smaller studies on speech production with subjects who natively speak American English from New York or New Jersey (Benway et al., 2022, Benway et al., 2023). Most of the studies had subjects with an age range inside of 6–17, but some featured adults (Benway et al., 2022, Benway et al, 2023). PERCEPT-GFTA was divided into subcorpora based on the subject's identified condition: a history of SSD affecting rhotic production in prekindergarten, suspected childhood apraxia of speech, suspected SSD in rhotic production, and typical developing (Benway et al., 2023). The typically developing subjects from these corpora were not used in the present study due to study scope and my preexisting analysis of typically developing data from another source. PERCEPT-R only included subjects with pre-kindergarten or suspected SSD (speech sound disorder) affecting rhotic production (Benway et al., 2022).

There were a total of 42 children found in PERCEPT-R and PERCEPT-GFTA who were between the ages of 7;0 and 8;0, who were suspected to have either SSD or CAS (childhood apraxia of speech), or had been identified as having SSD in pre-kindergarten. Due to the scope of this study, I selected a subsection of these children. I sorted each eligible child in reverse alphabetical order by file directory, and selected every third child to be included in my data set. I then selected the first audio file for each subject, which was always labelled as "pre-treatment". This resulted in a total of 14 audio files across 14 subjects.

One subject, PrestonHullEdwards2013_286, was removed due to a problem in the processing or labelling of their audiofile—the audio contained a voice with a pitch far too deep to reasonably be a 7-year-old. This left 13 subjects. Then, two of the subjects, one from PERCEPT-R and one from PERCEPT-GFTA but both labelled PerceptionRCT_88, were discovered to actually be the same child from the same study. However, as it was two unique audio files, I decided to keep both in my data set, and simply apply statistics by considering the data from both files to be linked to the same participant predictor. So in the end, I had 12 subjects and 13 audio files.

The audio files for each of these speech-disordered children consisted of the child reading out words from a list. Because the children came from different studies, some of them read the same lists as each other and some of them read different lists. Overall, each word list featured multiple instances of each diphthong /aɪ/ (or /ʌɪ/), /aʊ/ (or /ʌʊ/), /eɪ/, and /oʊ/. Notably, /ɔɪ/ was largely absent and ended up only being produced a single time by a single subject.

Due to the scope of this thesis, the data for the typically developing children was sourced from my previous paper on diphthong duration (Russell, 2023), which analyzed data from POLER. Plasticity of Language in Epilepsy Research (POLER) studied the effect of various kinds of epilepsy on narrative production in children aged 7 to 11 (Berl et al, 2005, Gaillard et al, 2007, Mbwana et al., 2009, Strekas et al., 2013, Steinberg et al., 2013).

While the subjects of PERCEPT-R and PERCEPT-GFTA produced words from a word list, the subjects of POLER produced a spontaneous narrative based on images (Berl et al, 2005, Gaillard et al, 2007, Mbwana et al., 2009). As a result, the words used by each POLER child had less consistency between subjects than the word list data for the speech-impaired children. Additionally, this narrative production data featured 5/6 subjects producing /ɔɪ/ with the highest frequency of any diphthong due to using "the boy" as the main character of the story. The sixth child told the story from first person and did not produce /ɔɪ/ a single time.

3.2. Experiment design

I opened each audio file in Praat and used a TextGrid to mark the beginning and end of each diphthong the child produces (excluding any phonologically erroneous diphthongs or target-diphthongs that were produced off-target). I also marked the phonological environment of each diphthong (i.e. /bihaɪv/ as the environment of one specific production of /aɪ/).

The following North American English diphthongs were counted: /aɪ/, /aʊ/, /eɪ/, /ɔɪ/, and /oʊ/. When /aɪ/ and /aʊ/ were realized as the Canadian Raised forms /ʌɪ/ and /ʌʊ/ they were labelled with the Canadian Raised forms. This labelling allowed for the option to assess if Canadian raising had a substantial effect on diphthong durations (and should thus not be included in the means of the unraised forms). As the population for this study was entirely from the United States, Canadian Raising was rare, but still present. /ʌʊ/ was only produced a single time, so this study has been unable to yield any valid result for that sound. However, /ʌɪ/ was yielded a sufficient number of times to produce valid statistics.

Additionally, of the disordered children, only one child produced /ɔɪ/, and only a single time. So despite /ɔɪ/'s abundance in the control data, there are also no valid results for the impact of disorder on this vowel.

After annotating the diphthongs in textGrids, I wrote a Praat script that automatically outputted the durations in a way that could be imported to a spreadsheet. I included in the spreadsheet the age, gender, and condition of each subject, as well as the specific diphthong that goes with each duration.

Next, I calculated means and standard deviations for five groups: all disordered, suspected SSD, pre-kindergarten SSD, suspected CAS, and typically developing. Then I ran a 1m model on the data to determine the effects and their significance, if any.

3.3. Results

3.3.1. Means

Mean (Standard Deviation) (number of subjects who produced the diphthong at least once/the total number of subjects in the disorder category)

Figure 3.3.1.: Means

Diphthong	All Disordered	Suspected CAS	PreKHistorySSD	SuspectedSSD	TD

aɪ	259.9557ms (116.2612ms) (11/12 subjects)	206.8546ms (89.62216ms) (1/1 subjects)	225.9537ms (146.4426ms) (4/4 subjects)	286.9712ms (109.3875ms) (7/8 subjects)	194.5979ms (54.5369ms) (6/6 subjects)
aʊ	253.1273ms (126.4025ms) (11/12 subjects)	199.9115ms (46.03449ms) (1/1 subjects)	207.2209ms (31.44299ms) (3/4 subjects)	280.4037ms (154.616ms) (7/8 subjects)	227.8803ms (28.66231ms) (6/6 subjects)
eɪ	194.1409ms (65.91575ms) (10/12 subjects)	160.4701ms (18.44531ms) (1/1 subjects)	190.0597ms (16.92526ms) (3/4 subjects)	201.7932ms (86.01861ms) (6/8 subjects)	158.3447ms (28.69303ms) (6/6 subjects)
ɔɪ	308.4106ms (N/A) (1/12 subjects)	308.4106ms (N/A) (1/1 subjects)	N/A (N/A) (0/4 subjects)	N/A (N/A) (0/8 subjects)	287.1324ms (287.1324ms) (5/6 subjects)
oʊ	235.8094ms (93.14203ms) (10/12 subjects)	173.082ms (21.79454ms) (1/1 subjects)	181.2245ms (56.08768ms) (3/4 subjects)	273.5564ms (100.3644ms) (6/8 subjects)	177.5576ms (40.58823ms) (6/6 subjects)
ʌj	209.1379ms (82.69033ms) (7/12 subjects)	253.1551ms (N/A) (1/1 subjects)	195.5062ms (9.8486 37ms) (2/4 subjects)	204.9495ms (113.3576ms) (4/8 subjects)	134.236ms (18.92837ms) (2/6 subjects)
ʌʊ	121.467ms (N/A) (1/12 subjects)	N/A (N/A) (0/1 subjects)	N/A (N/A) (0/4 subjects)	121.467ms (N/A) (1/8 subjects)	N/A (N/A) (0/6 subjects)

3.3.2. Effects

I ran a `lm` model on the data, comparing non-speech-disordered durations with speech disordered durations (as a group, rather than the individual disorders).

Considering all diphthongs aside from /ɔɪ/ and /ʌʊ/ (due to too few productions), disorder increases diphthong length by an estimated 36.982ms. The p-value is 0.11065, which is > 0.05, which means that this difference is not significant. This study cannot reject or accept the null hypothesis that speech disordered and non-speech disordered children produce diphthongs of the exact same length. Additionally, no significant effect of diphthong property (frontness and height of the initial vowel, /j/ versus /w/ end-vowel, and Canadian Raised versus unraised) was found. There was also no significant effect of gender or age.

I have included the estimated effects of different factors in the following chart, as well as the p-values. However, as the p-values are consistently over 0.05, there is no evidence to show if the estimated effects are significant or not. In order to better determine significance or lack thereof, more data and studies would need to be conducted.

Figure 3.3.2.: Effects

Characteristic	Estimated Effect	p-value
End target /j/ or /w/	/w/ yielded longer duration by 7.610ms	0.8250

Canadian Raising	Canadian Raising yielded shorter duration by 5.947ms	0.9114
Initial target front or back	Back initial target yield longer duration by 26.839ms	0.5842
Initial target height mid or low	Low initial target yielded longer duration by 55.030ms	0.2083
Condition of speech disabled or non-speech disabled	Speech disability yielded longer duration by 36.982ms	0.1107
Age (range 7;0 to 8;0)	Higher age yielded longer duration by 3.202ms	0.3533
Gender male or female	Male gender yielded longer duration by 28.832ms	0.1742

3.4. Discussion

This study found a statistically insignificant correlation between speech disorder and diphthong duration. indicator for speech impairment. However, further study would need to be conducted to validate this conclusion. It's possible that the selection of subjects in this thesis was too small, or contained many outliers. Future study should focus on comparing children who've performed the same task (i.e., all the subjects performed word list). Future study should also utilize typically developing controls that have been explicitly identified as *linguistically* typically developing. Comparing diphthong durations with the children's articulation rate should also be considered.

4. Conclusion & Summary

Overall, the present thesis has established that diphthong duration may not be a reliable indicator of speech disorder, and that future study is required to validate it.

I've analyzed past research in the field of speech acquisition, disorders, and analysis. English diphthongs are among the earlier vowels acquired by English-speaking children, and their development and disorders often pattern with tense monophthongs. Diphthong errors often involve diphthongs being monophthongized, i.e. /aj/ realized as [aː]. The durations of English diphthongs tend to be more similar to the durations of tense monophthongs than lax monophthongs.

The present thesis showed, with statistical *insignificance*, that children with speech disorders on average produce diphthongs 36.982 milliseconds longer than those without speech disorders. Further study needs to be performed to determine if the difference is truly insignificant, or if this study's data is an outlier in the grand scheme of diphthong durations. Larger data sets from a larger number of children will improve the accuracy of the means. Additionally, care should be taken to compare disordered subjects to control subjects who speak the same dialect.

APPENDIX A: Subjects

Subjects

- CAS_511 through PERCEPT-GFTA
- CPF_32 through PERCEPT-R
- EPG_50 through PERCEPT-R
- PerceptionRCT_88 through PERCEPT-GFTA and PERCEPT-R
- PerceptionRCT_103 through PERCEPT-GFTA
- PerceptionRCT_109 through PERCEPT-R
- PerceptionRCT_520 through PERCEPT-GFTA
- POLER/Match/146
- POLER/Match/162
- POLER/Match/163
- POLER/Match/164
- POLER/Match/173
- POLER/Match/174
- PrestonHullEdwards2013_281 through PERCEPT-GFTA
- PrestonHullEdwards2013_286 through PERCEPT-R—removed due to audio or labelling problem
- PrestonHullEdwards2013_289 through PERCEPT-GFTA
- PrestonHullEdwards2013_297 through PERCEPT-R
- PrestonHullEdwards2013_298 through PERCEPT-GFTA
- US2014_360 through PERCEPT-R

APPENDIX B: Praat Script

The following is the script used to extract the durations from the annotated textGrids.

```
clearinfo

ejCounter = 0
ajCounter = 0
ojCounter = 0
awCounter = 0
owCounter = 0
vwCounter = 0
vjCounter = 0


preKHistorySSD$ = "phon - Clinical - PERCEPT-GFTA - PreKHistorySSD/"
suspectedCAS$ = "phon - Clinical - PERCEPT-GFTA - SuspectedCAS/"
suspectedSSD$ = "phon - Clinical - PERCEPT-GFTA - SuspectedSSD/"
pERCEPTR$ = "phon - Clinical - PERCEPT-R/"

for i from 1 to 13
    if i = 1
       child$ = preKHistorySSD$ + "PrestonHullEdwards2013_281"
    elsif i = 2
       child$ = preKHistorySSD$ + "PrestonHullEdwards2013_289"
    elsif i = 3
       child$ = preKHistorySSD$ + "PrestonHullEdwards2013_298"
    elsif i = 4
       child$ = suspectedCAS$ + "CAS_511"
    elsif i = 5
       child$ = suspectedSSD$ + "PerceptionRCT_88"
    elsif i = 6
       child$ = suspectedSSD$ + "PerceptionRCT_103"
    elsif i = 7
       child$ = suspectedSSD$ + "PerceptionRCT_520"
    elsif i = 8
       child$ = pERCEPTR$ + "CPF_32_-_PRE-1"
    elsif i = 9
       child$ = pERCEPTR$ + "EPG_50_-_PRE-1"
    elsif i = 10
       child$ = pERCEPTR$ + "PerceptionRCT_88_-_PRE-1"
    elsif i = 11
       child$ = pERCEPTR$ + "PerceptionRCT_109_-_PRE-1"
    elsif i = 12
       child$ = pERCEPTR$ + "PrestonHullEdwards2013_297_-_PRE-1"
    elsif i = 13
       child$ = pERCEPTR$ + "US2014_360_-_PRE-1"
    endif


    textGrid = Read from file: child$ + ".TextGrid"
    numberOfIntervals = Get number of intervals: 2

    for interval to numberOfIntervals
       diphthong$ = Get label of interval: 2, interval
       if diphthong$ <> ""

             ; getting length of the diphthong
             start = Get start time of interval: 2, interval
             end = Get end time of interval: 2, interval
```

```
            length = 1000 * (end - start)    ;in ms

            if diphthong$ = "eɪ"
                  ejCounter = ejCounter + 1
                  ejLengths[ejCounter] = length
                  ejLabels$[ejCounter] = child$
            elsif diphthong$ = "aɪ"
                  ajCounter = ajCounter + 1
                  ajLengths[ajCounter] = length
                  ajLabels$[ajCounter] = child$
            elsif diphthong$ = "ɔɪ"
                  ojCounter = ojCounter + 1
                  ojLengths[ojCounter] = length
                  ojLabels$[ojCounter] = child$
            elsif diphthong$ = "aʊ"
                  awCounter = awCounter + 1
                  awLengths[awCounter] = length
                  awLabels$[awCounter] = child$
            elsif diphthong$ = "oʊ"
                  owCounter = owCounter + 1
                  owLengths[owCounter] = length
                  owLabels$[owCounter] = child$
            elsif diphthong$ = "ʌʊ"
                  vwCounter = vwCounter + 1
                  vwLengths[vwCounter] = length
                  vwLabels$[vwCounter] = child$
            elsif diphthong$ = "ʌɪ"
                  vjCounter = vjCounter + 1
                  vjLengths[vjCounter] = length
                  vjLabels$[vjCounter] = child$
            endif

        endif
    endfor

    ; cleaning up
    removeObject: textGrid

endfor

appendInfoLine: "/eɪ/ lengths"
for i to ejCounter
    appendInfoLine: ejLengths[i]
endfor
appendInfoLine: "/eɪ/ labels"
for i to ejCounter
    appendInfoLine: ejLabels$[i]
endfor

appendInfoLine: "/aɪ/ lengths"
for i to ajCounter
    appendInfoLine: ajLengths[i]
endfor
appendInfoLine: "/aɪ/ labels"
for i to ajCounter
    appendInfoLine: ajLabels$[i]
endfor

appendInfoLine: "/ɔɪ/ lengths"
for i to ojCounter
    appendInfoLine: ojLengths[i]
```

```
endfor
appendInfoLine: "/ɔɪ/ labels"
for i to ojCounter
    appendInfoLine: ojLabels$[i]
endfor

appendInfoLine: "/aʊ/ lengths"
for i to awCounter
    appendInfoLine: awLengths[i]
endfor
appendInfoLine: "/aʊ/ labels"
for i to awCounter
    appendInfoLine: awLabels$[i]
endfor

appendInfoLine: "/oʊ/ lengths"
for i to owCounter
    appendInfoLine: owLengths[i]
endfor
appendInfoLine: "/oʊ/ labels"
for i to owCounter
    appendInfoLine: owLabels$[i]
endfor

appendInfoLine: "/ʌʊ/ lengths"
for i to vwCounter
    appendInfoLine: vwLengths[i]
endfor
appendInfoLine: "/ʌʊ/ labels"
for i to vwCounter
    appendInfoLine: vwLabels$[i]
endfor

appendInfoLine: "/ʌɪ/ lengths"
for i to vjCounter
    appendInfoLine: vjLengths[i]
endfor
appendInfoLine: "/ʌɪ/ labels"
for i to vjCounter
    appendInfoLine: vjLabels$[i]
endfor
```

References

Colour coding for editing purposes: part of a book, whole book, article in journal, thesis/dissertation, other

Adi, Y., Keshet, J., Cibelli, E., Gustafson, E., Clopper, C., & Goldrick, M. (2016). Automatic measurement of vowel duration via structured prediction. *J. Acoust. Soc. Am.*, *140*(6), 4517–4527. https://doi.org/10.1121/1.4972527

Allen, G. D., & Hawkins, S. (1980). Phonological rhythm: Definition and development. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology: Volume 1: Production* (pp. 227–256). Academic Press. https://doi.org/10.1016/B978-0-12-770601-6.50017-6

Allison, K. M., Cordella, C., Iuzzini-Seigel, J., & Green, J.R. (2020). Differential Diagnosis of Apraxia of Speech in Children and Adults: A Scoping Review. *Journal of Speech, Language, and Hearing research*, *63*(9), 2952–2994. https://doi.org/10.1044/2020_JSLHR-20-00061

Baird, A., Cummins, N., Schnieder, S., Krajewski, J., & Schuller, B. W. (2020). An evaluation of the effect of anxiety on speech — computational prediction of anxiety from sustained vowels. *Proc. Interspeech 2020*, 4951–4955. https://doi.org/10.21437/interspeech.2020-1801

Baum, S.R., Blumstein, S.E., Naeser, M.A., & Palumbo, C.L. (1990). Temporal dimensions of consonant and vowel production: An acoustic and CT scan analysis of aphasic speech. *Brain and Language*, *39*, 33–56. https://doi.org/10.1016/0093-934X(90)90003-Y

Benway, N.R., Preston, J.L., Hitchcock, E., Rose, Y., Salekin, A., Liang, W., & McAllister, T. (2023). Reproducible Speech Research With the Artificial Intelligence–Ready PERCEPT Corpora. *Journal of Speech, Language, and Hearing Research*, *66*, 1986–2009. https://doi.org/10.1044/2023_JSLHR-22-00343

Benway, N.R., Preston, J.L., Hitchcock, E., Salekin, A., Sharma, H., & McAllister, T. (2022). PERCEPT-R: An Open-Access American English Child/Clinical Speech Corpus Specialized for the Audio Classification of /ɹ/. *Proc. Interspeech 2022*, 3648–3652. https://doi.org/10.21437/Interspeech.2022-10785

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America, 109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109

Berl, M. M., Balsamo, L. M., Xu, B., Moore, E. N., Weinstein, S. L., Conry, J. A., ... & Ritter, F. J. (2005). Seizure focus affects regional language networks assessed by fMRI. *Neurology*, *65*(10), 1604–1611. https://doi.org/10.1212/01.wnl.0000184502.06647.28

Bernhardt, B. H., & Stemberger, J. P. (1998). *Handbook of phonological development from the perspective of constraint-based nonlinear phonology.* Academic Press. http://doi.org/10.1017/S0305000900214633

Berns, J., & Jacobs, H. (2012). A first glance at the role of length in production and perception of diphthongs before Dutch coda *l*. *Linguistics in the Netherlands*, *29*(1). https://doi.org/10.1075/avt.29.02ber

Black, L. I., Vahratian, A., & Hoffman, H. J. (2015). Communication Disorders and Use of Intervention Services Among Children Aged 3-17 Years: United States, 2012. *NCHS data brief*, (205), 1–8. https://pubmed.ncbi.nlm.nih.gov/26079397/

Boersma, P., & Weenink, D. (2023). Praat: doing phonetics by computer [Computer program]. Version 6.2.19, retrieved 12 September 2022 from http://www.praat.org

Boë, L. J., Heim, J. L., Honda, K., & Maeda, S. (2007). The vocal tract of humans and neanderthals: Acoustic capabilities and consequences for the debate on the origin of language. A reply to Lieberman (2007a). *Journal of Phonetics*, *35*(4), 564–581. https://doi.org/10.1016/j.wocn.2007.06.006

Chevrot, J. P, Nardy, A., & Barbu, S. (2011). Developmental dynamics of SES-related differences in children's production of obligatory and variable phonological alternations. *Language Sciences*, *33*, 180–191. https://doi.org/10.1016/j.langsci.2010.08.007

Chin, S. B., Bergeson, T., & Phan, J. (2012). Speech intelligibility and prosody production in children with cochlear implants. *Journal of Communication Disorders*, *45*(5), 355–366. https://doi.org/10.1016/j.jcomdis.2012.05.003

Ciocca, V., & Whitehill, T. L. (2013). Chapter 4: The acoustic measurement of vowels. In M. J. Ball, & F. E. Gibbon (Eds.), *Handbook of vowels and vowel disorders* (pp. 113–137). Psychology Press. https://doi.org/10.4324/9780203103890

Collins, M., Rosenbek, J. C., & Wertz, R. T. (1983). Spectrographic analysis of vowel and word duration in apraxia of speech. *Journal of speech and hearing research*, *26*(2), 224–230. https://doi.org/10.1044/jshr.2602.224

Connaghan, K. P., & Patel, R. (2012). Impact of prosodic strategies on vowel intelligibility in childhood motor speech impairment. *Journal of Medical Speech-Language Pathology*, *20*(4), 133–139. https://go.gale.com/ps/i.do?id=GALE%7CA328852894&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=10651438&p=HRCA&sw=w&userGroupName=anon%7E824ba2e8

Davis, B., & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech and Hearing research*, *38*(6), 1199–1211. https://doi.org/10.1044/jshr.3806.1199

Demuth, K. & E. J. Fee. (1995). Minimal words in early phonological development. *Brown University & Dalhousie University*. https://scholar.google.nl/scholar?hl=nl&as_sdt=0%2C5&as_vis=1&q=Minimal+prosodic+words+in+early+phonological+development&btnG=

Donegan, P. (2013). Chapter 2: Normal vowel development. In M. J. Ball, & F. E. Gibbon (Eds.), *Handbook of vowels and vowel disorders* (pp. 24–60). Psychology Press. https://doi.org/10.4324/9780203103890

Eaton, C.T., & Ratner, N.B. (2013). Rate and phonological variation in preschool children: effects of modeling and directed influence. *Journal of Speech, Language, and Hearing Research*, *56*(6), 1751–1763. https://doi.org/10.1044/1092-4388(2013/12-0171)

Emerich, G. H. (2012). *The Vietnamese Vowel System*. [Doctoral dissertation, University of Pennsylvania]. Publicly Accessible Penn. Dissertations. https://repository.upenn.edu/edissertations/632

Fikkert, P. (1994). On the acquisition of prosodic structure. *Glot International*, *2*(4), 9–10. https://hdl.handle.net/2066/17308

Flemming, E. S. (2004). Contrast and perceptual distinctiveness. In E. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically-based phonology* (pp. 232–276). Cambridge, U.K.: Cambridge University Press. https://doi.org/10.1017/CBO9780511486401.008

Fromkin, V. (1970). Tips of the slung, or to err is to human. *Univ. of California at Los Angeles: Working Papers in Phonetics*, *13*, 65. https://scholar.google.com/scholar_lookup?title=Tips+of+the+slung%2C+or+to+err+is+human&publication_year=1970&journal=UCLA+Working+Papers+in+Phonetics&pages=65

Gaillard, W. D., Berl, M., Moore, E., Ritzl, E., Rosenberger, R., Weinstein, S., Conry, J., Pearl, P., Ritter, F., Sato, S., Vezina, L., Vaidya, C., Wiggs, E., Fratalli, C., Risse, G., Ratner, N. B., Gioia, G., & Theodore, W. (2007). Atypical language in lesional and non-lesional complex partial epilepsy. *Neurology*, *69*(18), pp. 1761–71. https://doi.org/10.1212/01.wnl.0000289650.48830.1a

Gay, T. (1968). Effect of speaking rate on diphthong formant movements. *J. Acoust. Soc. Am.*, *44*(6), 1570–1573. https://doi.org/10.1121/1.1911298

Gay, T. (1970). A Perceptual Study of American English Diphthongs. *Language and Speech*, *3*(2), 65–88. https://doi.org/10.1177/002383097001300201

Gay, T. (1978). Physiological and acoustic correlates of perceived stress. *Language and Speech*, *21*(4), 347–353. https://doi.org/10.1177/002383097802100409

Gibbon, F. E. (2013). Chapter 17: Therapy for abnormal vowels in children with speech disorders. In M. J. Ball, & F. E. Gibbon (Eds.), *Handbook of vowels and vowel disorders* (pp. 429–446). Psychology Press. https://doi.org/10.4324/9780203103890

Gibbon, F. E., & Beck, J. M., (2002). Therapy for abnormal vowels in children with phonological impairment. In M. J. Ball & F. E. Gibbon (Eds.), *Vowel disorders* (pp. 217–248). Butterworth Heinemann. https://eresearch.qmu.ac.uk/handle/20.500.12289/2264

Grabe, E., Post, B., & Watson, I. (1999). The acquisition of rhythmic patterns in English and French. In *Proceedings of the 14th international congress of phonetic sciences* (pp. 1201–1204). University of California. https://scholar.google.com/scholar?cluster=5672175603430787771&hl=nl&as_sdt=0,5

&scilib=1&scioq=Grabe,+E.,+Post,+B.,+%26+Watson,+I.+(1999).+The+acquisition+of+rhythmic+patterns+in+English+and+French.

Guenther, Frank H. (2016). *Neural control of speech.* The MIT Press. https://doi.org/10.7551/mitpress/10471.001.0001

Guyette, T. W., & Diedrich, W. M. (1981). A Critical Review of Developmental Apraxia of Speech. *Speech and Language*, *5*, 1–49. https://doi.org/10.1016/B978-0-12-608605-8.50007-6

Hartsuiker, R. J., Kolk, H. H. J., & Heike Martensen. (2005). Division of labor between internal and external speech monitoring. In R. J. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech* (pp. 187–205). Psychology Press. https://doi.org/10.4324/9780203506196

Hattunen, K. H. (2000). *Early childhood hearing impairment: Speech intelligibility and late outcome* [Doctoral dissertation, University of Oulu]. Acta Universittatis Oulensis, B 35. https://books.google.nl/books/about/Early_Childhood_Hearing_Impairment.html?id=05HjNQAACAAJ&redir_esc=y

Hattunen, K. H. (2001) Phonological development in 4–6-year-old moderately hearing impaired children. *Scandinavian Audiology* 30 (53), 79–82. https://doi.org/10.1080/010503901750166745

Hawthorne, K., & Fischer, S. (2020). Speech-language pathologists and prosody: Clinical practices and barriers. *Journal of Communication Disorders*, *87*(2). https://doi.org/10.1016/j.jcomdis.2020.106024

Henry, C. E. (1990). The development of oral diadochinesia and non-linguistic rhythmic skills in normal and speech disordered young children. *Clin. Linguist Phon.*, *4*(2), 121–137. https://doi.org/10.3109/02699209008985476

Hodge, M. M. (2013). Chapter 1: Development of the vowel Space. In M. J. Ball, & F. E. Gibbon (Eds.), *Handbook of vowels and vowel disorders* (pp. 1–23). Psychology Press. https://doi.org/10.4324/9780203103890

Howard, S., & Heselwood, B. (2013). Chapter 3: The contribution of phonetics to the study of vowel development and disorders. In M. J. Ball, & F. E. Gibbon (Eds.), *Handbook of vowels and vowel disorders* (pp. 61–112). Psychology Press. https://doi.org/10.4324/9780203103890

Jacks, A., Thomas, P. M., & Davis, B. L. (2013). Chapter 12: Vowels in apraxia of speech. In M. J. Ball, & F. E. Gibbon (Eds.), *Handbook of vowels and vowel disorders* (pp. 326–346). Psychology Press. https://doi.org/10.4324/9780203103890

Kalathottukaren, R. T., Purdy, R., McCormick, S. C., & Ballard, E. (2015). Behavioural measures to evaluate prosodic skills: A review of assessment tools for children and adults. *Contemporary Issues in Communication Science and Disorders*, *42*, 138–154. https://doi.org/10.1044/cicsd_42_S_138

Kallay, J. E., Dilley, L., & Redford, M. A. (2022). Prosodic Development During the Early School-Age Years. *Journal of Speech, Language, and Hearing Research*, *65*, 4025–4046. https://doi.org/10.1044/2022_JSLHR-22-00124

Kehoe, M, M. (1998). Support for metrical stress theory in stress acquisition. *Clinical Linguistics & Phonetics*, *12*, 1–23. http://dx.doi.org/10.3109/02699209808985210

Kehoe, M, M. & Stoel-Gammon, C. (2001). Development of syllable structure in English-speaking children with particular reference to rhymes. *Journal of Child Language*, *28*, 393–432. https://doi.org/10.1017/S030500090100469X

Kent, R. D., Netsell, R., & Abbs, J. H. (1979). Acoustic characteristics of dysarthria associated with cerebellar disease. *Journal of speech and hearing research*, *22*(3), 627–648. https://doi.org/10.1044/jshr.2203.627

Kent, R. D. & Vorperian, H. K. (2013). Speech impairment in Down syndrome: A review. *Journal of Speech, Language, and Hearing Research*, *56*(1), 178–210. https://doi.org/10.1044/1092-4388(2012/12-0148)

Khan, T., Freeman, R., & Druet, A. (2023). Louder Than Words: Pediatric Speech Disorders Skyrocket Throughout Pandemic. *Komodo Health*. https://www.komodohealth.com/hubfs/2023/Speech_Pathology_Research_Brief.pdf?hsCtaTracking=e40f8354-b93c-4306-9160-2bcce2f146de%7Cf835bb14-ede9-434f-a4c0-c55564520191

Kuhl, P. K. (1993). Early linguistic experience and phonetic perception: implications for theories of developmental speech perception. *Journal of Phonetics*, *21*, 125–139. https://doi.org/10.1016/S0095-4470(19)31326-9

Kulju, P. (2003). *Production of word structures: A constraint-based study of 2;6 year old Finnish children at-risk for dyslexia and their controls*. [Doctoral thesis, University of Jyväskylä]. Jyväskylä Studies in Languages, 52. http://urn.fi/URN:ISBN:951-39-1419-4

Kulju, P., & Savinainen-Makkonen, T. (2008). Normal and disordered phonological acquisition in finnish. In A. Klippi & K. Launonen (Eds.) *Research in logopedics: Speech and language therapy in Finland* (pp. 35–62). Multilingual Matters. https://doi.org/10.21832/9781847690593-005

Ladefoged, P. (2006). *A course in phonetics*. Thomson Wadsworth. https://books.google.nl/books/about/A_Course_in_Phonetics.html?id=jRmoQgAACAAJ&redir_esc=y

Lee, S., Potamianos, A., & Narayanan, S. (2014). Developmental acoustic study of American English diphthongs. *J. Acoust. Soc. Am.*, *136*(4), 1880–1884. https://doi.org/10.1121/1.4894799

Lee, Y., Keating, P., & Kreiman, J. (2019). Acoustic voice variation within and between speakers. *J. Acoust. Soc. Am.*, *146*, 1568–1579. https://doi.org/10.1121/1.5125134

Lehiste, I., & Peterson, G. E. (1961). Transitions, glides, and diphthongs. *J. Acoust. Soc. Am.*, *33*(3), 268–277. https://doi.org/10.1121/1.1908638

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *The Behavioral and brain sciences*, *22*(1), 1–75. https://doi.org/10.1017/s0140525x99001776

Lieberman, P. (1980). On the development of vowel production in young children. In G. H. Yeni-Komshian, J. S. Kavanagh, & C. A. Ferguson (Eds), *Child phonology, Volume 1: Production*. Academic Press. https://doi.org/10.1016/B978-0-12-770601-6.50012-7

Liu, S., & Takeda, K. (2021). Mora-timed, stress-timed, and syllable-timed rhythm classes: Clues in English speech production by bilingual speakers. *Acta Linguistica Academica*, *68*(3), 350–369. https://doi.org/10.1556/2062.2021.00469

Lyakso, E. E., & Silvén, M. (2002). Comparative characteristics of early vocalizations in Finnish and Russian infants. *Journal of Sensory System*, *16*, 65–74. https://researchportal.tuni.fi/en/publications/comparative-characteristics-of-early-vocalizations-in-finnish-and

Mäenpää, M. (1990). *Kehityksellinen verbaali dyspraksia kuudella esikouluikäisellä lapsella [Developmental verbal dyspraxia: A multi case study of six pre-school children]*. [Master's thesis, University of Helsinki]. Unpublished, cited through Kulju & Savinainen-Makkonen (2008).

MacNeilage, P. F., & Davis, B. (1990). Acquisition of speech production: Frames, then content. In M. Jeannerod (Ed.), *Attention and performance 13: Motor representation and control* (pp. 453–476). Psychology Press. https://doi.org/10.4324/9780203772010

Mbwana, J., Berl, M. M., Ritzl, E. K., Rosenberger, L., Mayo, J., Weinstein, S., ... & Sato, S. (2009). Limitations to plasticity of language network reorganization in localization related epilepsy. *Brain*, *132*(2), 347–356. https://doi.org/10.1093/brain/awn329

Menyuk, P. (1980). The role of context in misarticulations. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology* (pp. 211–226). Academic Press. https://doi.org/10.1016/B978-0-12-770601-6.50016-4

Nijland, L., Maassen, B., Van Der Meulen, S., Gabreëls, F., Kraaimaat, F. W., & Schreuder, R. (2003). Planning of syllables in children with developmental apraxia of speech. *Clinical linguistics & phonetics*, *17*(1), 1–24. https://doi.org/10.1080/0269920021000050662

Nooteboom, S. G. (1973). The tongue slips into patterns. Victoria A. Fromkin (Ed.), *Speech errors as linguistic evidence* (pp. 144–156). De Gruyter Mouton. https://doi.org/10.1515/9783110888423

Nooteboom, S. G. (1997). The prosody of speech: Melody and rhythm. *Utrecht University*. https://www.researchgate.net/publication/46675980_The_prosody_of_speech_Melody_and_rhythm

Nooteboom, S. G. & Quené, H. (2017). Self-monitoring for speech errors: Two-stage detection and repair with and without auditory feedback. *Journal of Memory and Language*, *95*, 19–35. https://doi.org/10.1016/j.jml.2017.01.007

Nooteboom, Sieb G. & Hugo Quené. (2021). Why are some speech errors detected by self-monitoring "early" and others "late"? *Proceedings of DiSS*. https://www.researchgate.net/publication/354330721_Why_are_some_speech_errors_detected_by_self-monitoring_early_and_others_late

Nooteboom, S. G. & Slis, I. H. (1972). The phonetic feature of vowel length in Dutch. *Language and Speech*, *15*(4), 301–316. https://doi.org/10.1177/002383097201500401

Olejarczuk, P. & Redford, M. A. (2013). The relative contribution of rhythm, intonation and lexical information to the perception of prosodic disorder. *Proceedings of Meetings on Acoustics*, *19*. https://doi.org/10.1121/1.4800625

Parkhurst, B. G. & Levitt, H. (1978). The effect of selected prosodic errors on the intelligibility of deaf speech. *Journal of Communication Disorders*, *11*(2–3), 249–256. https://doi.org/10.1016/0021-9924(78)90017-5

Patel, R., Hustad, K., Connaghan, K. P., & Furr, W. (2012). Relationship between prosody and intelligibility in children with dysarthria. *Journal of Medical Speech-Language Pathology*, *20*, 1–5. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4299669/

Peppé, S. J. E. (2009). Why is prosody in speech-language pathology so difficult? *International Journal of Speech-Language Pathology*, *11*(4), 248–271. https://doi.org/10.1080/17549500902906339

Peppé, S. J. E. (2018). Prosodic development in atypical populations. In P. Prieto, & N. Esteve-Gilbert (Eds.), *The development of prosody in first language acquisition* (pp. 343-362). John Benjamins. https://doi.org/10.1075/tilar.23.17pep

Peter, B., & Stoel-Gammon, C. (2005). Timing errors in two children with suspected childhood apraxia of speech (sCAS) during speech and music-related tasks. *Clinical linguistics & phonetics*, *19*(2), 67–87. https://doi.org/10.1080/02699200410001669843

Peterson, S. J. (2018). *Accounting for diphthongs: Duration as contrast in vowel dispersion theory*. [Doctoral dissertation, Georgetown University]. Georgetown University Graduate Theses and Dissertations - Linguistics. https://repository.library.georgetown.edu/handle/10822/1053075

Pietarinen, A. (1987). *Vantaalaisten v. 1980 syntyneiden lasten viisivuotisseulalla mitatut kielelliset häiriöt ja niiden yhteydet kehityksen muihin osatekijöihin [Speech and language disorders and their connection to other developmental features – Screening test for five-year-old children born in 1980 in Vantaa city]*. [Master's thesis, University of Helsinki]. Unpublished, cited through Kulju & Savinainen-Makkonen (2008).

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

Polyanskaya, L., & Ordin, M. (2015). Acquisition of speech rhythm in first language. *The Journal of the Acoustical Society of America*, *138*(3), EL199–EL204. https://doi.org/10.1121/1.4929616

Reford, M. A., Kapatsinski, V., & Cornell-Fabiano, J. (2018). Lay listener classification and evaluation of typical and atypical children's speech. *Language and Speech*, *61*(2), 277–302. https://doi.org/10.1177/0023830917717758

Reynolds, J. (2013). Chapter 9: Recurring patterns and idiosyncratic systems. In M. J. Ball, & F. E. Gibbon (Eds.), *Handbook of vowels and vowel disorders* (pp. 229–259). Psychology Press. https://doi.org/10.4324/9780203103890

RStudio Team (2020). RStudio: Integrated Development for R [Computer program]. Version 4.1.1, retrieved 10 August 2021 from http://www.rstudio.com

Russell, K. A. (2023). Standard lengths of American 7-year-old's diphthongs. *University of Amsterdam*. Unpublished.

Salidis, J., & Johnson, J. S. (1997). The production of minimal words: a longitudinal case study of phonological development. *Language Acquisition*, 6, 1–36. https://www.jstor.org/stable/20011452

Savinainen-Makkonen, T. (1998). Ensisanojen kauden fonologiaa: Tapaustutkimus [Phonology of the first words: A case study]. In M. Karjalainen (Ed.), *Kielen ituja: Ajankohtaista lapsenkielen tutkimuksesta* (pp. 44–83). Suomen ja saamen kielen ja logopedian laitoksen julkaisuja 10. Oulun yliopistopaino. https://scholar.google.com/scholar_lookup?title=Ensisanojen+kauden+fonologiaa%3A+tapaustutkimus+%5BPhonology+at+the+first-words+stage%3A+a+case+study%5D&author=T.+Savinainen-Makkonen&publication_year=1998&pages=44-83

Shattuck-Hufnagel, Stefanie. (1992). The role of word structure in segmental serial ordering, *Cognition*, *42*(1–3), 213–259. https://doi.org/10.1016/0010-0277(92)90044-I

Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. (1997). Developmental apraxia of speech: II. Toward a diagnostic marker. *Journal of speech, language, and hearing research*, *40*(2), 286–312. https://doi.org/10.1044/jslhr.4002.286

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research*, *44*(5), 1097–1115. https://doi.org/10.1044/1092-4388(2001/087)

Steinberg, M., Bernstein Ratner, N., Berl, M. & Gaillard, W. (2013). Fluency patterns in narratives from children with localization-related epilepsy. *Journal of Fluency Disorders*, *38*(2) 193–205. http://dx.doi.org/10.1016/j.jfludis.2013.01.003

Stojanovik, Vesna. (2011). Prosodic deficits in children with Down syndrome. *Journal of Neurolinguistics*, *24*(2), 145–155. https://doi.org/10.1016/j.jneuroling.2010.01.004

Stokes, S. F., Lau, J. T., & Ciocca, V. (2002). The interaction of ambient frequency and feature complexity in the diphthong errors of children with phonological disorders. *Journal of Speech, Language, and Hearing Research*, *45*(6), 1188–1201.

https://link.gale.com/apps/doc/A96267572/AONE?u=amst&sid=bookmark-AONE&xid=13cdf191

Strekas, A., Bernstein Ratner, N., Berl, M. & Gaillard, W. D. (2013). Narrative abilities of children with epilepsy. *International Journal of Language and Communication Disorders*, *48*(2), 207–219. https://doi.org/10.1111/j.1460-6984.2012.00203.x

Tasko, S. M., & Greilick, K. (2010). Acoustic and articulatory features of diphthong productions: A speech clarity study. *Journal of Speech, Language, and Hearing Research*, *53*, 84–99. https://doi.org/10.1044/1092-4388(2009/08-0124)

Teferra, B. G., Borwein, S., DeSouza, D. D., Simpson, W., Rheault, L., & Rose, J.. (2022). Acoustic and linguistic features of impromptu speech and their association with anxiety: Validation study. *JMIR Mental Health*, *9*(7) https://doi.org/10.2196/36828

Thomas, E. R. (2011). *Sociophonetics: An introduction*. Palgrave Macmillan. https://books.google.nl/books?id=GyNHEAAAQBAJ&printsec=frontcover&source=gbs_ViewAPI&redir_esc=y#v=onepage&q&f=false

Torvelainen, P. (2007). *Kaksivuotiaiden lasten fonologisen kehityksen variaatio : puheen ymmärrettävyyden sekä sananmuotojen tavoittelun ja tuottamisen tarkastelu [Variation in Phonological Development of Two-Year-Old Finnish Children: A study of speech intelligibility and attempting and production of words]*. [Doctoral dissertation, University of Jyväskylä]. Jyväskylä studies in humanities. https://jyx.jyu.fi/handle/123456789/13420

Vorperian, H. K., & Kent, R. D. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1510–1545. https://doi.org/10.1044/1092-4388(2007/104)

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

Wellman, B. E., Case, I. M., Mengert, I. G., & Bradbury, D. E. (1931). Speech sounds of young children. *University of Iowa Studies in Child Welfare*, *5*(2), 7–80. https://www.mpi.nl/publications/item_2406326

Werker, J. F. & Tees, R. C. (1984). Cross-Language Speech Perception: Evidence for Perceptual Reorganization During the First Year of Life. *Infant Behavior and Development*, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Vilkman, E., Niemi, J., & Ikonen, U. (1988). Fragile X speech phonology in Finnish. *Brain and Language*, *34*, 203–221. https://doi.org/10.1016/0093-934X(88)90133-2

Wilson, E. M., Abbeduto, L., Camarata, S., & Shribberg, L. D. (2019). Speech and motor disorders and intelligibility in adolescents with Down syndrome. *Clinical Linguistics and Phonetics*, *33*(8), 790–814. https://doi.org/10.1080/02699206.2019.1595736

Zee, E. (1999). Chinese (Hong Kong Cantonese). In *Handbook of the IPA* (pp. 58-60). Cambridge University Press. https://www.internationalphoneticassociation.org/content/handbook-ipa

Þráinsson, H. (2004). *Faroese: An overview and reference grammar*. Føroya Fróðskaparfelag. https://books.google.nl/books/about/Faroese.html?id=VrKPQAAACAAJ&redir_esc=y