Title

Phonetic development in an agglutinating language

Author

Cychosz, Margaret E.

Publication Date

2020

Peer reviewed|Thesis/dissertation

Phonetic development in an agglutinating language

by

Margaret E. Cychosz

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Linguistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Keith Johnson, Co-chair
Professor Sharon Inkelas, Co-chair
Associate Professor Mahesh Srinivasan

Summer 2020

Phonetic development in an agglutinating language

Abstract

Phonetic development in an agglutinating language

by

Margaret E. Cychosz

Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor Keith Johnson, Co-chair

Professor Sharon Inkelas, Co-chair

Child speech is highly variable. The speech apparatus—the vocal tract, tongue, teeth, and vocal folds—develop at different rates for different children, which helps explain some of the variability in children's speech. For example, the ratio of the oral to pharyngeal cavities changes as children age, making it difficult to establish reliable articulatory routines (Smith and Goffman 1998; Vorperian et al. 2005). While anatomy does play this undeniable role in child speech development, this dissertation focuses instead on components in the child's *environment* that may explain their speech patterns. To do so, the studies here report on speech development in bilingual children acquiring South Bolivian Quechua (henceforth "Quechua") and Spanish in a mid-size town in Bolivia. Chapters 2 and 3 examine the environmental effect of Quechua's linguistic structure on speech development while chapter 4 examines the role of quantity of language exposure.

Chapter 2 examines how a phonological factor - vowel inventory—interacts with speech development. In Quechua, there are just three phonemic and two allophonic vowels. Chapter 2 asks if vowel inventory size mitigates acoustic variability in children aged four through ten. The study finds that children as young as four approximate adult-like acoustic targets, suggesting that child speech variability is contingent upon the language being learned. Still, children do not necessarily speak like adults. Using these vowel data, chapter 2 additionally finds that the children vary greatly in their ability to articulatorily compensate for their vocal tract morphologies, potentially explaining some of the large amounts of between-speaker variation that characterizes child speech.

Chapter 3 examines how another aspect of Quechua's linguistic structure - its highly agglutinating morphology - may interact with speech development. In Quechua, speakers construct words by supplementing root morphemes with a series of grammatical suffixes. Chapter 3 asks if this word composition could interact with children's coarticulatory patterns. Here coarticulation is quantified using two novel acoustic measures that are less susceptible to the challenges that the child vocal anatomy poses for traditional spectral analysis. In experiment 2 of chapter 3, these measures are validated on a large corpus of four-year-old children acquiring English.

The central results of chapter 3 demonstrate that children and adults distinguish coarticulatorily between word environments: within morpheme (e.g. p**ap**a 'potato') and between morpheme (e.g. pap**a-pi** 'potato-LOC). However, only children compensate for the morphologically complex words'

prosodic structure by shortening word duration. As a result, it remains unclear if the children's spoken language patterns better reflect morphological or prosodic structure.

Finally, chapter 4 asks how children's language exposure and use—in Quechua or Spanish—predicts the speech production outcomes from chapters and 2 and 3. In this study, each child's bilingual language use patterns are computed from daylong audio recordings of the children's language environments. Employing random sampling to annotate the recordings, chapter 4 efficiently estimates the children's bilingual language environments: the annotation method required an average of just 90 minutes of language category annotation from each recording to effectively estimate each child's dual language exposure.

The chapter finds that children's language exposure and use does indeed predict their speech patterns: children with monolingual Quechua mothers have tighter, less variable vowel categories than children with bilingual Quechua-Spanish or Quechua-dominant mothers. Additionally, children who use more Quechua throughout the day tend to distinguish more between the morphological environments tested in chapter 4. This last finding indicates that the more Quechua these children use, the better they are at analyzing and breaking down morphologically complex words.

Overall, the results from this dissertation demonstrate how myriad factors relating to linguistic structure and quantity of language exposure predict child speech variation. In doing so, this work also demonstrates how understudied languages—and novel methodological techniques like child-friendly acoustic measures and daylong audio recordings—can reveal aspects of children's psycholinguistic representations, addressing long-standing questions in the field. Thus, this dissertation concludes that children face anatomical obstacles, such as an unstable oral to pharyngeal cavity ratio, that explain some of their speech variability. However, these anatomical factors co-exist with numerous elements of the children's everyday linguistic environments to predict speech development.

A todas las mamas que hicieron posible este trabajo

To all the mamas who made this work possible,
none more so than my own

Contents

List of Figures

List of Tables

Acknowledgments

There are many people without whom this dissertation, and my development as an independent researcher and fieldworker, would not have been possible.

Thank you to my community in Bolivia. Ana Torres and René Iglesias welcomed me into their home and helped me form a partnership with the primary school. I'm grateful to that school for allowing me to work with and learn from the students and faculty. This dissertation is dedicated to all of the mothers and matriarchs in Bolivia who graciously lent me their time, attention, and expertise to contribute to the daylong recording corpus.

The quantity of data—much of it naturalistic!—analyzed for this project might not have been completed without my stellar group of research assistants. Karina Fong-Hirschfelder, Sierra Adachi, Rachel Elia, Kiran Girish, Hitomi Torng, and Sonnet Phelps all contributed to data pre-processing, coding, sorting, plotting—you name it, in addition to giving me frank and necessary feedback. I loved our monthly team dinners and hope they learned as much from me as I did from them.

I have learned and developed so much from working with the members of the Daylong Audio Recording of Children's Linguistic Environments (DARCLE) community. In particular I thank Adriana Weisleder, Amanda Seidl, and especially Alex Cristia and Middy Casillas. Chapter 4 of this dissertation, and my fieldwork with children in Bolivia, would not have happened without Alex and Middy's guidance and feedback.

To members of my Maryland community: Michele Erskine, Allie Johnson, Tatiana Thonsevanh, and Katie Von Holzen, all of whom helped me get out of my linguistics box. And to my Berkeley community: Sarah Bakst, for always being available to answer any acoustics question, my running inspiration Julia Nee, Myriam LaPierre who always has excellent feedback and helped me tease apart the effects in Chapter 2, Zach O'Hagan for getting me to take much-needed breaks, Ginny Dawson, Raksit Lau, Hannah Sande, Emily Clem, Karee Garvin, Matt Faytak, and Emily Remirez. I had two really low moments in graduate school, and Arik Shams was sitting across the table to talk me through both of them. He has taught me so much about science, communication, and a different world view.

Everything from my fieldwork to teaching at Berkeley was made smoother through the combined efforts of Belén Flores and Paula Floro. Thank you for the supportive words and immediate replies to (my) desperate emails and phone calls. I might still be stuck in that airport in Bolivia without you, Paula!

I've had some incredible teachers and advisors over the past five years. Thank you to Jan Edwards for not just giving me desk space on campus at Maryland, but also involving me as a full lab member. Jan isn't on my committee, but her theoretical ideas are apparent everywhere in this dissertation. At Berkeley, thank you to Lev Michael who supported my idea of doing a project on Quechua speech development early. I have so enjoyed learning from Susan Lin who taught me to use ultrasound and encouraged me to develop my computational chops. My ideas about language were fundamentally shaped by Andrew Garrett's Historical Linguistics class. But more than that, Andrew teaches his students to think outside of themselves—what have linguists done in the past and how do we, as a field, contend with that legacy? It is this message that I will miss. Also his humor. And his apple crisp.

If dissertations, or graduate degrees, could have co-authors, Ronald Sprouse would be mine. Ron spent cumulative *hours* first teaching me computing basics, then getting me to think like an efficient data scientist, and eventually leading me to be an independent developer. He never lost patience with me, even when I wanted to chuck my laptop across the room out of frustration. Our department at Berkeley is lucky to have him.

To my committee: Mahesh Srinivasan has been the perfect outside member, consistently providing input from another field that always makes me stop and reconsider my ideas. More than anything else, Mahesh has taught me how to put myself in a child's shoes and experience language from their perspective.

Of course this whole experience has only been possible because of my co-advisors, Keith Johnson and Sharon Inkelas. It's a testament to them that when I pitched the idea of taking my qualifying exams early, moving to D.C., and working in a new lab for a year, they both agreed that it was what I needed to do for my personal life.

It's been almost five years, but it was Sharon who first encouraged me to explore first language acquisition when she gave me her copy of *Phonological Development: The First Two Years* my first or second week of graduate school. This past semester I taught a (modified version) of the Phonological Development course that Sharon developed. My ability to teach that class, and my still-developing expertise in child phonology, is thanks to many discussions with and reading suggestions from her.

I came to Berkeley because of Keith. He made me feel like I could get a Ph.D. He made me feel like I deserved to work with him. I think my fondest memories of graduate school will be sitting together in his office to hash out predictions about a new vocal tract normalization model or experiment with some new code. Keith never judged me for needing to learn programming. Or stats. Or any of the other hats I've learned to wear in graduate school. He just supported me and trusted that I would develop those skills with time.

To my (non-academic) friends: Anna Strozinsky Pedraza, Halie Vining, Monica Mills, and Bita Eisenhut have been patiently listening to me rant about linguistics and conferences and fieldwork for years. Never has their friendship meant more to me than these last few months, fervently writing under quarantine.

Thank you to my family, especially my sister, who keeps me grounded, and my stepfather for keeping me on my toes. My mother's influence as a lifelong elementary educator is apparent everywhere in this dissertation. Growing up, many a kitchen table discussion centered around children's literacy and the importance of the early language environment. Thank you for supporting me through all my travels (Paris, Berkeley, Bolivia...) and of course for your tireless efforts to sew the more than *eighty* shirts used during fieldwork.

Finally, thank you to Alejandro, who has proofread every single page of this dissertation (except these acknowledgements), multiple times. Finishing the dissertation on the other side of the country from you was never going to be ideal. Doing it under quarantine even less so. But knowing that I got to call you at the end of each writing day was the best thing to look forward to.

Chapter 1

A unified theory of speech development

1.1 Introduction

Phonetic variation in spoken language is widespread. In their iconic visualization of vowel categories, Peterson and Barney (1952) illustrated the inconstancy of speech production, even when reduced to two dimensions in an F2-F1 plane (Figure 1.1). This phenomenon in spoken language is the source of one of the foremost theoretical inquiries in speech science: the search for invariance in the signal (Kleinschmidt 2017; Liberman et al. 1967; Liberman and Mattingly 1985; Lindblom 1990; Perkell and Klatt 2014).

Despite widespread variability, adult speech is also, simultaneously, constrained. Peterson and Barney's formant mapping shows vowel categories that disperse and overlap but, to a listener, remain distinct.[1] It does not matter what acoustic parameter is measured (e.g. duration, kurtosis), or what segment is studied (e.g. glides, fricatives), this conclusion remains. Adult speakers strike a fine balance between sloppy variability and communicative precision, between communicative efficiency and comprehensibility (Lindblom 1990). Their ability to strike this balance is the result of years of trial and error in early language development, combined with the slow mastery of fine motor schemata and accumulation of exemplars that form entrenched phonological categories. The result is that, despite the inevitable articulatory and acoustic variation of speech, adult speakers have relative PHONETIC STABILITY.

Because phonetic stability is a learned characteristic of speech, progressively mastered throughout childhood, then children's spectral and temporal speech patterns should differ from those of adults. Child patterns might be more variable. This is, in fact, one of the most robust findings in developmental phonology and phonetics, replicated innumerable times. Children are more variable than adults. They consistently show more intra- and inter-speaker acoustic and articulatory variation than adults in the production of vowels (Barbier

[1] Results from Hillenbrand et al. (1995) suggest that dynamic information such as fundamental frequency contours, as opposed to only steady-state vowel formants, helps alleviate some of this variability in speech perception tasks.

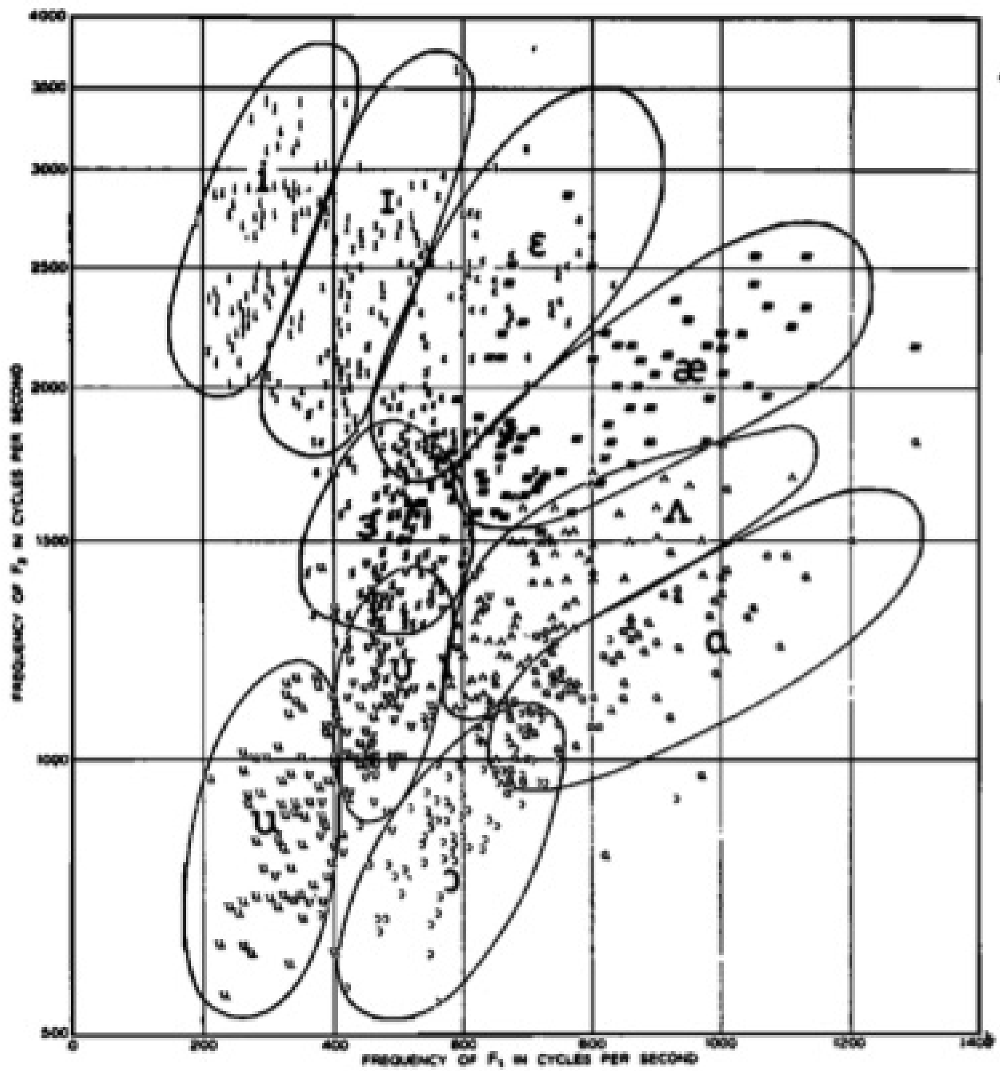


Figure 1.1: Formant frequencies from the N=76 speakers of Peterson & Barney (1952). There is a high amount of inter-speaker variation, but the categories are discernible.

et al. 2015; Eguchi and Hirsh 1969; Hillenbrand et al. 1995; Lee et al. 1999; Peterson and Barney 1952; Pettinato et al. 2016; cf. McGowan et al. 2014; Nittrouer 1993 for F1), fricatives (Nittrouer et al. 1989; Nittrouer et al. 1996; Zharkova et al. 2011), plosives and VOT (Imbrie 2005; Nittrouer 1993), and coarticulation (Gerosa et al. 2006; Goffman et al. 2008; Zharkova et al. 2011). Furthermore, children's acoustic variability generally decreases with age (Lee et al. 1999; Pettinato et al. 2016; Tingley and Allen 1975 [for speaking rate]). Children reach adult-like levels of "normal" speech variability in early adolescence (Hazan and Barrett 2000; McMurray et al. 2018; Vorperian and Kent 2007). I refer to this phenomenon as PHONETIC NARROWING, following the well-known phenomenon of perceptual narrowing in infant speech perception.

Children's speech variability is traditionally attributed to child anatomy (Ménard et al. 2007; Turner et al. 2009), and its changes over development (Vorperian and Kent 2007; Vorperian and Wang 2009), or underdeveloped motor routines and articulator coordination (Barbier et al. 2020; Goffman et al. 2008; Green et al. 2000; Smith and Zelaznik 2004). Children's perceptual faculty can likewise explain some aspects of early speech variation either because children inconsistently incorporate auditory feedback (Caudrelier et al. 2019; Cooper et al. 2018) or because their perceptual attention to acoustic detail changes with development (McMurray et al. 2018).

The developing motor and perceptual systems play undeniable roles in speech development. But many of the findings on child speech variation, and the explanations for it, have been drawn from studies on a handful of closely-related languages such as English and French. This is a significant shortcoming. There are additional factors that could explain the high amount of variability in child speech, but they are dependent upon the child's language learning experience. The goal of this dissertation will be to examine two of these factors: 1) linguistic structure, specifically phonological inventory and word composition, and 2) language exposure. Each explanation constitutes a potential part of phonetic and phonological development beyond those factors that have traditionally been studied. And to study them, we have to examine languages, and cultures, outside of those traditionally studied in child speech development.

1.2 The theory

Speaking like an adult requires mastery of an appropriate ratio of articulatory and acoustic variability to stable, comprehensible acoustic targets. This ratio may promote an efficient speech signal. The crucial point here is that the ratio, while not unique to child speech, is formulated completely differently in children and continually updated throughout development. Why? First, there are obvious, static anatomical differences between adults and children. Second, unlike an adult, a child is undergoing rapid physiological change that interferes with the establishment of articulatory routines and entrenchment of somatosensory phonological representations. These anatomical differences are independent of language and cultural socialization. Finally, the ratio of speech stability to efficiency may differ in children because they have been exposed to and practiced less language than adults. Specifically, children do not possess the same material over which to generalize and formulate linguistic categories. And they are still learning the acoustic effects of their speech output. This is illustrated in Figure 1.2.

Anatomy

The child vocal apparatus is not just a miniature version of the adult apparatus. A two-year-old child's tongue assumes a significantly larger portion of the posterior oral cavity than an adult's tongue (Crelin 1987:95; Fletcher 1973:168-170). Likewise, in a two-year-old, the palate is longer relative to the pharyngeal cavity, (Goldstein 1980:186; see also figures in Crelin 1987:34-39 & 96-99). Both of these anatomical differences limit the fine horizontal movement required to, for example, differentiate /s/ and /ʃ/. They also limit the quick transitions necessary to resist the anticipatory effect of a following vowel on the /ʃ/ in sequences like /ʃu/ or /ʃi/ (Zharkova et al. 2011).

Articulatory routine

Anatomical differences between adults and children resolve as typically- developing children age. However, this also means that a child's anatomy is highly transient with anatomical changes that are both non-linear (Vorperian and Kent 2007) and non-uniform. Different articulators mature at different rates (Nittrouer 1993). These elements of phonological development present a challenge for children who must establish accurate, replicable articulatory-acoustic mappings in the face of rapid, uneven change. This challenge is analogous to shooting an arrow at a bullseye with an ever-changing ratio of arm to bow length.

Studies on orofacial articulator development typically conclude that children's articulatory gestures are less stable than adults' gestures. Children vary more from one gesture to the next (Goffman et al. 2008; Green et al. 2000; Smith and Goffman 1998; Smith and Zelaznik 2004). Furthermore, like acoustical narrowing, this articulatory variability decreases as children age (Green et al. 2000; Grigos 2009).[2] For example, using "minimal sentence pairs," Goffman et al. (2008) demonstrated that the intra-subject timing and magnitude of upper lip movements were more variable over repeated productions for children than adults. Children may also employ different production strategies for equivalent sounds (Smith and Goffman 1998). The articulatory force behind a gesture that distinguished /s/ and /ʃ/ at 36 months may no longer serve to adequately contrast these sounds at 48 months.

The effect of articulatory routine on representation

The lack of articulatory routine in child speech has repercussions for children's phonological representations. For this connection between children's phonological representation

[2] Grigos (2009) only found significant differences by age for jaw movement variability, not lower or upper lip (longitudinal measurements for 12-21 weeks beginning at 1;7).

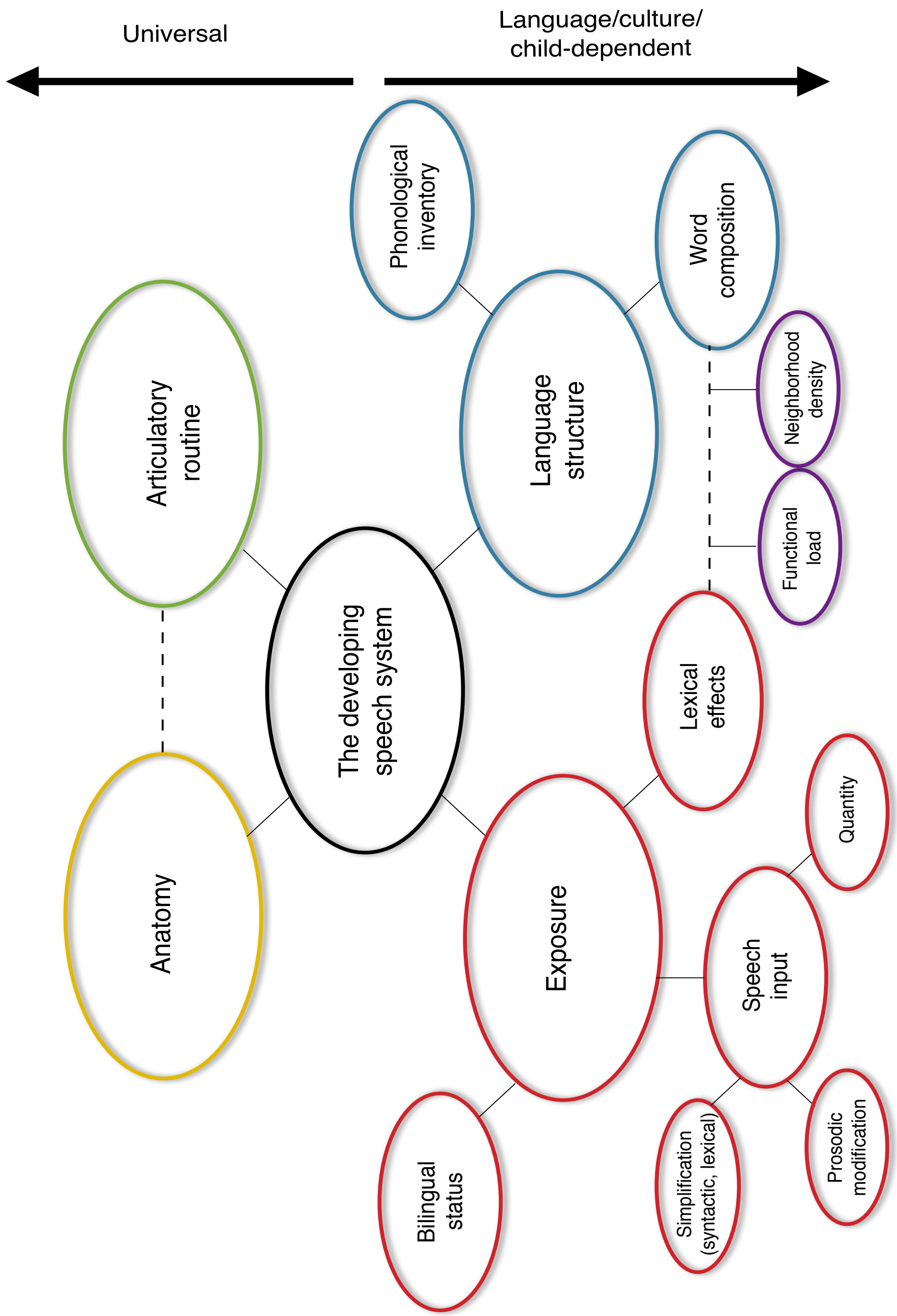


Figure 1.2: A unified model of speech development

and articulatory habit, I assume that speech is an emergent, dynamic system (Thelen and Smith 1996; Smith and Thelen 2003). The speech system is dynamic because it changes, both ontogenically and phylogenically. It is emergent, at least in part, because it is the result of self-organization, a phenomenon "generate[d]...through [its] own activity" (Smith and Thelen 2003:343). For phonological acquisition, Menn et al.'s Linked-Attractor Model (Menn et al. 2013) illustrates a dynamic systems approach well.[3] Throughout this chapter, I incorporate and acknowledge ideas from several theories of cognitive and phonological development, including Menn's Linked-Attractor Model, Redford's Core Model (Davis and Redford 2019), McAllister-Byun et al.'s A(rticulatory)-Map model (McAllister Byun et al. 2016b) and Thelen and Smith.

Infants and young children are exposed to an endless stream of speech that they must parse into linguistic constituents like words, and then eventually articulate. For this formidable task, infants are provided a multi-dimensional acoustic landscape, where they can carve out perceptual representations, and a motor landscape where articulatory routine can be established and represented. In landscapes such as these, everything from phonological categories to language-specific word ordering can be established. (One could optionally choose to invoke a limited hypothesis space of Universal Grammar into this representation - it would certainly make the task of language acquisition easier for the child).[4] With time, exposure, and practice, "craters" are carved into these landscapes (Menn et al. 2013:475) and these constitute phonological categories.

Perception. Early perceptual exposure from the ambient language establishes scratches upon the metaphorical acoustic landscape. Perceptual scratches commence as the infant employs statistical learning techniques to parse words from the speech stream, such as noting phone and syllable co-occurrence patterns in the ambient language (Saffran 2003). For example, in English, some syllables such as /li/ are much more likely to occur word-finally (e.g. [ˈlʌvli] 'lovely') than word-initially (e.g. [ˈlimɚ] 'lemur') or word-medially (e.g. [bəˈliv] 'believe'). On the basis of this information, a young learner could suppose that the syllable or phone that followed [li] in the speech stream marked the beginning of a word. If that word was a member of a minimal pair, the child could begin to delimit relevant phonological contrasts. This explains the phenomenon of perceptual narrowing throughout the first year of life. Neonates have near universal perceptual contrast abilities that gradually taper down into native language phonemes in the first 6-12 months (Kuhl 1991; Maye et al. 2002; Werker and Tees 1984).[5]

Once words have been segmented from the speech stream, phonological categories also emerge as the infant or child organizes episodic traces of each word on the basis of overlapping

[3]Note that Menn et al. (2013) do not reference the theories outlined in Thelen and Smith (1996); I have drawn those connections independently (in part because I saw so many parallels).

[4]For example, Werker and Curtin (2005) outline some innate components in their PRIMR model.

[5]The landscape-crater metaphor closely resembles Kuhl (1991)'s Perceptual Magnet Effect, a fact that Menn et al. (2013) note as well.

acoustics and semantic referent (Figure 1.4).[6] Where the traces overlap a great deal, a lexical, syllabic, or phone target emerges (Pierrehumbert 2003).

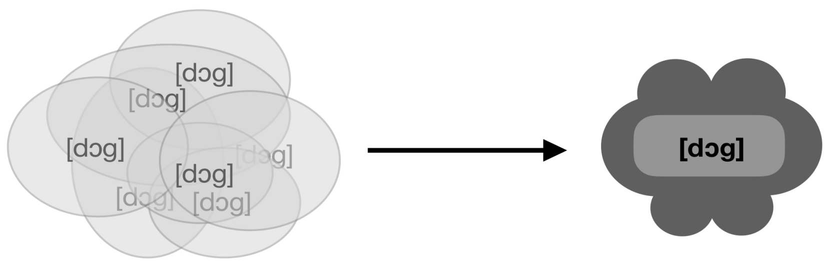


Figure 1.3: An emergent perceptual target: each round object represents an episodic trace of the word ‘dog’ [dɔg] that the child has parsed from the speech stream.

The first target to emerge from overlapping traces is likely to be a word or word-like unit (e.g. a high-frequency collocation like ‘on top’), not a phone or syllable. But with time, as the child’s vocabulary grows, phonological neighborhoods that revolve around words and syllables will become more dense (Storkel 2004). Dense neighborhoods mean more competition between perceptually-confusable words (Charles-Luce and Luce 1990; Storkel 2002), putting pressure on the representations to become more segmental and abstract. Thus, the more frequently that a sound or syllable occurs across distinct lexical items, the more likely that a child will abstract the sound or syllable away from the original lexical context (the episodic traces). Note that just because syllables and segments have been abstracted away from the original episodic traces does not mean that the episodic traces have dissipated, are no longer accessible, or are no longer relevant. Laboratory studies on sociophonetic perception demonstrate that this cannot be the case (Drager 2011; Johnson et al. 1999). Instead, children and adults develop *redundant*, and at times conflicting, representations, spread across multiple levels of the grammar (phone, syllable, word, multiword unit, etc.).

Articulation. Articulation is also critical for establishing native phonological representations. Consider that infants’ speech perception of non-native contrasts can be compromised when articulation is inhibited and the infants can’t employ an internal feedback model. A particularly striking example of this is the inability of infants aged 0;6 to distinguish between a retroflex-dental contrast when they have a teething toy in their mouth that inhibits lingual movement Bruderer et al. (2015). Acknowledging this, Menn et al. (2013) argue

[6]Semantic referent included because remember that the model has to account for why ’thyme’ and ’time’ are not real homophones (Gahl 2008).

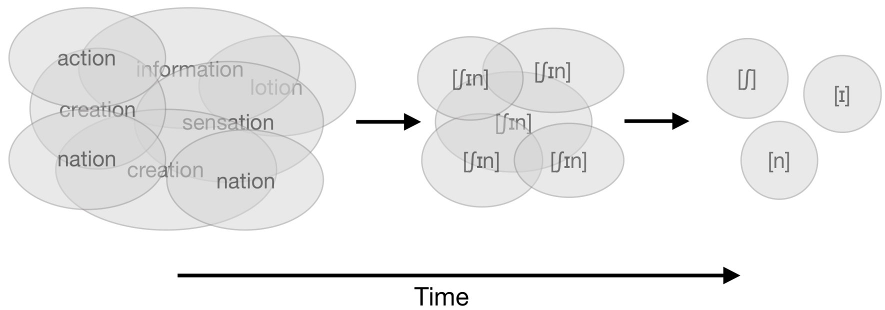


Figure 1.4: Perceptual representations emerge from generalizations made over acoustic traces of lexical and sub-lexical chunks.

for "perceptual/input templates," such as the perceptual craters described above, as well as "production/output templates" where articulatory habit can be encoded (2013:474). For articulatory representations, scratches upon the motor landscape commence with newborn's vegetative sounds and crying (for the first 1-2 months) but continue to evolve through vocal exploration and the marginal babbling (around 4-5 months) and canonical babbling stages (7 months and beyond) (Davis and Redford 2019; Warlaumont 2015). For every vocalization that an infant makes - crying, raspberries, babble - an *articulatory* trace of the vocal tract configuration employed to make the sound, here simply referred to as a schema, is reflected in the motor space. And an *acoustic* scratch in the form of the infant's vocalization is reflected on the perceptual landscape. Thus, articulatory schemata form a critical component of phonological representations.

In adults, routine may entrench phonological categories further. In children, it might instead resemble scratches around a single point in the n-dimensional landscape (McAllister Byun et al. 2016b). Figure 1.5 visualizes how this may play out. The first assumption is that the emerging perceptual category is the child's acoustic goal (here I assume this category is a word). When the child first attempts to utter the word, they will default to articulatory routines established in early babbling (Vihman 2017). The acoustic output of the child's production, inevitably deviating from the adult model, provides acoustic-auditory feedback that the infant can incorporate into their next attempt. Thus, with time, the child's productions slowly begin to approximate the emerging perceptual category, which is still developing as more episodic traces are accumulated. However, the key here is that the initial production attempts will mimic established articulatory routines. In addition, while the approximation of the perceptual category is inevitable in typically-developing children, the development of acoustic-articulatory mapping is not unidirectional (i.e. children's approximation of the perceptual target will not improve linearly over time). Children may prefer to rely upon established articulatory routines (i.e. Vihman's vocal motor schemes [McCune and Vihman

2001; Vihman 2017] or highly rank the PRECISE constraint of the A-Map model [McAllister Byun et al. 2016b]) at the expense of accurate acoustic output. Alternatively, children may approximate the perceptual target, but do so at the expense of articulatory ease and habit, likely resulting in high variability in their attempt (i.e. highly ranking the ACCURATE constraint in the A-Map model [McAllister Byun et al. 2016b]).

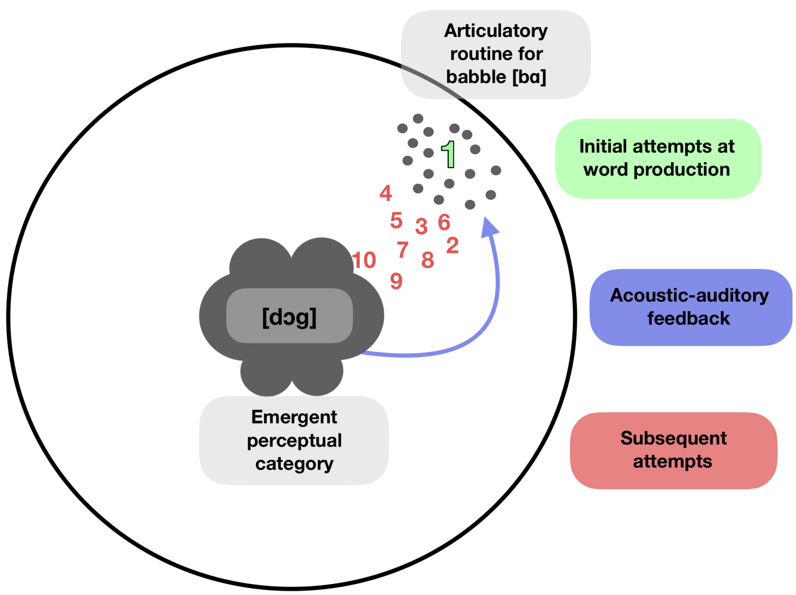


Figure 1.5: Initial word attempts are based on previous motor schemata, practiced during babbling. Acoustic-auditory feedback from these productions causes the child to update the articulatory representation of the word and closer approximate the perceptual target.

In early development, articulatory feedback is limited to sensory consequences of actual gestures that the child produces (Tilsen 2016) or EXTERNAL FEEDBACK (Grossberg 1978). But, eventually, the child can merely "compar[e] the predicted consequences of motor commands to sensory targets," without ever producing a gesture, in INTERNAL FEEDBACK (Grossberg 1978; Tilsen 2016:59). This has the benefit of speeding up phonological predictability.[7] The child does not have to actually produce the segment to comprehend its sensory implications; instead, the implications are hard-coded within each segment's representation. But such a rich, stable phonological representation takes time and practice.

[7] Tilsen (2016) would probably limit this to gestural predictability instead of expanding to acoustic-articulatory representations. I think the consequence of the feedback is the same, regardless of how the category is represented.

Perception-articulation link. The untouched acoustic and articulatory landscapes are universally shared and accessible from birth. In these spaces, acoustic and articulatory categories emerge with experience. However, the entire acoustic space is available to the child from birth - there are some changes in auditory sensitivity but the human auditory range does not change significantly over the first few years of life (Schneider et al. 1985). Unlike the acoustic space, the articulatory landscape *available* to the infant or child learner is dynamic. The developing anatomy effectively prohibits children from accessing certain parts of the motor landscape. Some clear examples of this are lingual contact with the pharyngeal wall or hard palate, neither of which is accessible to an infant whose tongue subsumes such a large portion of the oral cavity (Crelin 1987; Fletcher 1973). Motor planning advances greatly over the first few years of life as well, as demonstrated on a range of speech (Green et al. 2000; Goffman et al. 2008) and non-speech tasks (Chen et al. 2010). Consequently, the inaccessibility of the motor landscape, and the infant's physiological inability to articulate speech, is one reason that perception precedes production in early language development.

Early perception and articulation are intrinsically linked (Bruderer et al. 2015); note the important finding that the early babble routines of children mimic their ambient language (de Boysson-Bardies et al. 1984; de Boysson-Bardies et al. 1991). This connection may even strengthen with age.[8] With experience, it becomes progressively more difficult to emerge from entrenched perceptuo-articulator routines.

Each triangle in Figure 1.6 represents production of a single language chunk, be it a segment, syllable, or word. The slope of the crater represents the articulatory effort behind the production. Initially the child is merely constructing the category; the articulatory plan is not entrenched very deeply so the child produces the chunk variably from one token to the next. Yet, with exposure and practice, it becomes nearly impossible to propel production out of the deepened crater. Each production is consistent. Of course this explains well-known findings from second language phonology (Best and Tyler 2007). But it also addresses the tendency for children and adults who stutter to do so in more infrequent words (Anderson 2007; Ronson 1976). Smooth speech articulation is reserved for well-practiced, high-frequency items.

Naturally, as Menn et al. (2013) point out, our representational craters are not uniform. At the phonological level, some representations are more entrenched than others because we have different experiences with different phones, syllables, and words. Some phones are infrequent and more prone to merge during sound change with a neighbor that is more entrenched (Hay et al. 2015). Some phones contrast a lot of words and are reticent to variation, and thus change (Martinet 1952; Wedel et al. 2013).

[8] It is obvious that children have distinct perceptual and articulatory representations. Phonological production cannot entirely reflect a child's representation because, for example, a child who habitually stops their fricatives can nevertheless distinguish [t] from [s]. But fricative stopping and other child-specific phonological patterns are also not merely performance errors (cf. Hale and Reiss 1998). They are subject to grammatical constraints (McAllister Byun et al. 2016b; Rose and Inkelas 2011). Furthermore, an entirely unified perceptuo-articulator representation would also have to explain u-shaped phonological development and phonological idioms.

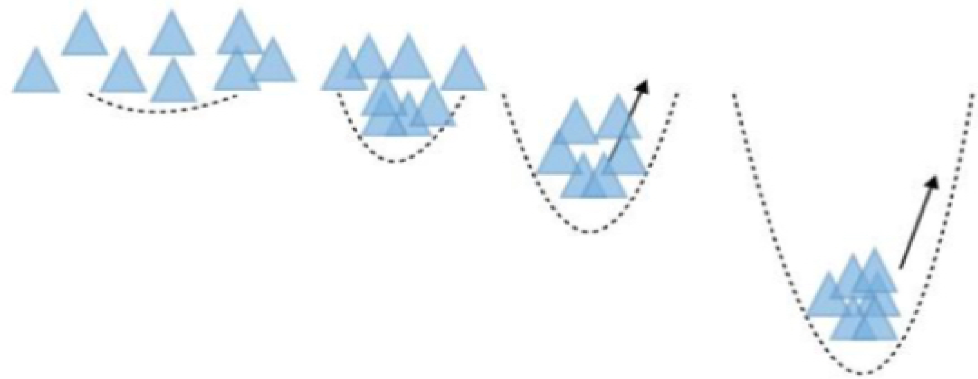


Figure 1.6: Greater perceptuo-articulator entrenchment with age. Childhood (L) to adulthood (R). *Concept adapted from Thelen and Smith (1996)*

All of these facts are exacerbated in children because their changing anatomy does not allow them to establish consistent articulatory routines. Indeed, Smith and Goffman (1998) point out that the child's anatomy is so mercurial that it is likely not even advantageous for children to settle into an articulatory routine, even if they could. Consequently, the neural pathways leading to stable articulatory routines are less ingrained (Smith and Goffman 1998). The less practiced, and thus ingrained, a movement trajectory, the more variable we anticipate it to be from one production to the next. And this process is cyclical. A child who fronts their velars mostly for articulatory reasons (Inkelas and Rose 2007; McAllister Byun 2012) consistently updates their phonological representations with traces of [tæt] and [dɪv] for 'cat' and 'give,' respectively (McAllister Byun et al. 2016b; McAllister Byun and Tessier 2016). This may even worsen with age as children master internal articulatory feedback! This explanation also alleviates the conundrum of phonological idioms (Ferguson and Farwell 1975): regressive phonological idioms may persist in child speech if they have a high output frequency in the child's speech relative to other output lexical items or adult input.

Exposure

Exposure also contributes to child speech patterns. The relationship is intuitive: the more a child is exposed to a sound, syntactic relation, or semantic domain, the faster they master the category. In speech development, category "mastery" means achieving acoustic stability or:

- stable, adult-like phonetic categories in perceptuo-motor space
- a stable, adult-like ratio of efficiency to comprehensibility in coarticulation patterns

In short, mastery means talking like an adult. So while the debate concerning the construction and organization of children's phonological representations may be unresolved, most contemporary models, abstractionist or emergentist, readily acknowledge the role that input and statistical inference play in phonological development (Demuth 2006; Edwards and Beckman 2008; Fikkert and Levelt 2008; McAllister Byun et al. 2016b; McAllister Byun and Tessier 2016; Seidl et al. 2014; see also Lidz and Gagliardi (2015); Meylan et al. (2017); Yang (2004); 2017)) for relevant arguments pertaining to morphosyntax). Really, our theoretical interest for phonetic and phonological development lies in understanding the relative contribution of input-related measures (e.g. raw frequency, biphone frequency, statistical inference) on the one hand and predispositions (e.g. learning biases, markedness hierarchies, language acquisition device) on the other.

The lexicon

Exposure to language plays a critical role in speech development, and thus acoustic stability. In particular, it affects which consonants first emerge in a child's speech, and when they do so. (de Boysson-Bardies et al. 1991; Stokes and Surendran 2005; Zamuner et al. 2005). It predicts children's consonant mastery (Edwards and Beckman 2008; Edwards et al. 2015; Jarosz et al. 2017), with cascading ramifications for word learning (Storkel 2004) and vocabulary development (Edwards et al. 2004).

But ascribing speech development to a purely bottom-up interpretation is disingenuous. For one thing, raw frequency often fails to explain developmental phenomena. Input frequency alone cannot explain the order of infants' perceptual attunement to vowel categories in Dutch (Tsuji et al. 2017) and even multiple, combined frequency-based metrics (segment token/type) cannot predict children's production accuracy in Polish (Jarosz et al. 2017). These studies demonstrate that, to improve models of phonological development, we should refine our definition of frequency.

To that end, a new frequency-based measurement has emerged in infant phonology: lexical interaction. The PROTO-LEXICON HYPOTHESIS, or the idea that infants employ known words to bootstrap into phonological categories, is compelling. In an interactive model of phonological development, both the unparsed speech stream (raw frequency, biphone co-occurrence) and the lexicon (prosodic word boundaries, semantic mappings) may interact to strengthen the boundaries of an infant's phonological categories (Feldman et al. 2013; Martin et al. 2013; Ngon et al. 2013; Swingley 2009; cf. Bergmann et al. 2017). Edwards et al. (2004)'s Lexical Scaffolding Hypothesis makes similar predictions for speech production and mastery in toddlers: the more words a child knows, the more accurate their segmental productions will be (see also Edwards et al. (2015)). This is because children abstract segments further and further from word types and sublexical chunks such as syllables (Cychosz et al. 2020a)

For phoneticians, the idea of lexical feedback in phonological development should be highly attractive. Recall the famous Peterson and Barney (1952) visual. In a major development for speech science, they found that categories are not static points in 2D acoustical space. But they also uncovered the true extent of category overlap. From the perspective of a child, this greatly complicates the phonological learning task: How to acquire a phonological system when the categories supplied in the input are not distinct?

Of course different parameters could be substituted for the first two formants, which would perhaps elucidate some differences between vowel categories.[9] For example, while F1 and F2 overlap greatly, perhaps a young child could rely more on durational cues to distinguish tense and lax vowels. For infants, however, Swingley (2009) proposed that an accessible lexicon might simplify the daunting phonological learning task as "words, which are identifiable by infants, might serve as rough indicators of where vowel category boundaries lie" (3624). I extend this from words to chunks of speech, both larger (e.g. 'What's 'at?') and smaller (e.g. single syllable) than the typical English word.

But why do I mention lexical effects in infant phonology when discussing input and environmental effects on speech development? And are these results from infants applicable to young children? To the first point, the role of the lexicon in phonological development indirectly contributes to a prominent debate in language development: the role of child-directed speech, which I address below. To the second, it is impossible to discuss the effects of exposure on acoustic stability without acknowledging the foundation upon which children have constructed their phonology in infancy. Furthermore, Edwards et al. (2004; 2015) have extended lexical effects into models of phonological development in older children. Finally, a comprehensive theory of developmental phonology should incorporate findings from the infant phonology literature.

Multilingualism

Children acquiring two languages offer a unique opportunity to evaluate the role of exposure upon phonetic and phonological development. Bilingual children are rarely exposed to both of their languages equally. This fact about bilingual development provides a method to manipulate exposure frequencies within the same child. Moreover, the roles of the *type* of language that children are exposed to - receptive versus expressive - can also be evaluated in these environments. For example, in communities undergoing language shift, children may frequently receive input in the parents' (minority) language - which the children do learn to speak - but the children express themselves at school or with peers in the majority language. As a result, the roles of expressive and receptive language can be evaluated within individual children.

Child-directed speech

We know that children form phonological representations despite fuzzy, overlapping acoustic categories provided in the input. Vowels in particular have been used to illustrate that while deriving phonological categories from messy input is not an impossible task

[9] For example, Adriaans and Swingley (2012) found that the prosodic dynamicity of CDS aided vowel category learning.

(Vallabha et al. 2007), there are ways to reduce the degrees of learning freedom for the child. An interactive model, where children receive feedback from the lexicon as their phonology develops, is one way to delimit the learning space. Another proposed method is child-directed speech (CDS).

Acoustically, CDS is typically characterized by hyper-articulated, temporally longer phones (Kuhl et al. 1997; Ratner 1984; cf. Cristia and Seidl 2014; Martin et al. 2015)[10] and more dynamic pitch contours (Fernald and Simon 1984; Liu et al. 2007), as well as shorter utterances containing fewer constituents (Fernald et al. 1989). These characteristics seem to bolster vowel category learning because longer, hyperarticulated tokens are further apart and less prone to overlap, providing more unique exemplars for the child to learn from. CDS appears to be similar, though not identical, cross-linguistically, at least in the languages studied (see Fernald et al. (1989) for a standardized comparison of French, German, Italian, Japanese, and British and American English and Kuhl et al. (1997) for Russian, Swedish, and American English). Models of infant learners find the more hyperarticulated tokens of CDS easier to disambiguate (Adriaans and Swingley 2012) and real infants favor CDS in the head-turn preference procedure (Fernald 1985; The Many Babies Consortium 2020). It would seem that, if we entertain both the possibility of bootstrapped information from the lexicon, as well as CDS, that we may identify a few of the critical tools that infants and children employ to construct phonological categories.

But the outstanding problem with this line of research should now be apparent: CDS is not universal. Some reports do not find reliable differences between CDS and adult-directed speech. A study on Norwegian CDS found that caregivers actually *underspecified* the vowel space compared to adult-directed speech (Englund and Behne 2006). Perhaps most critically, we know adults in many cultures do not speak directly to children either with the frequency that English-, Japanese-, or French-speaking caregivers do, or at all (Lieven et al. 1997). Speaking with children could be taboo within the cultural context or parents may view the children as invalid conversational partners. Elinor Ochs documented this element of language socialization in a Samoan village (Ochs 1988). And while some contemporary research presumes an almost universality of CDS, it was demonstrated years ago that Mayan mothers do not raise their pitch in speech directed to their children compared to adult-directed speech (Ratner 1984). Recently, large-scale quantitative analyses have begun to document the diversity of children's language exposure, though analysis can be limited to factors such as the number of words in the input and overheard versus child-directed speech (Casillas et al. 2019; Cristia et al. 2017; Mastin and Vogt 2016; Shneidman and Goldin-Meadow 2012; Shneidman et al. 2013; Vogt et al. 2015).

Shneidman and Goldin-Meadow (2012) and Shneidman et al. (2013) studied children's language environments in a Yucatec Mayan village and found that children are exposed primarily to overheard speech, in place of directed speech or a CDS register. Yet speech spoken directly to children remained the best predictor of word learning (see Shneidman and

[10] As Cristia and Seidl (2014) point out, the vowel space may be hyperarticulated because caregivers speak slower and use longer duration segments.

Woodward 2016 for alternative arguments). Vogt et al. (2015) found differences in the type of intentions in the CDS of Dutch versus Mozambique caregivers. In the same Mozambique community, Mastin and Vogt (2016) found that joint attention, or when two actors are focusing upon a mutual object or experience, is not required for vocabulary development. However, it was for Dutch-learning children in the Netherlands. More recently, Cristia et al. (2017) documented that children in Tsimane-speaking communities in Bolivia are exposed to fewer directed words (though not observed) per day than peer children in other pre-industrial societies. The authors compared their results to those from smaller-scale investigations in a Guatemalan village (Klein et al. 1977) and an !Kung tribe (Konner 1977). Casillas et al. (2019) likewise found that children aged 0;2-3;0 in a Tseltal Mayan community were exposed to very little directed speech, at least in comparison to North American samples (Bergelson et al. 2019b). Nevertheless, the authors conclude that the children acquiring Tseltal Mayan reached key linguistic milestones, such as early word combinations.

The overarching theme of this inquiry is that if caregivers direct less speech to their children, or none at all, then those children clearly receive proportionately less CDS. But to say that CDS universally helps establish phonological categories is to say that children who are exposed to the sing-songy components typical of CDS learn their phonological categories better and faster. Yet Tsimane- and Tseltal Mayan-speaking children all eventually attain adult-like phonological categories. Recent work also suggests that they hit key milestones in early phonological development: Cychosz et al. (under review) compared the babbling development of children from many of the aforementioned speech communities - Tsimane, Tseltal Mayan - with children learning English and Spanish in the United States. All of the children reached a key .15 ratio of canonical to non-canonical babbles in their speech by approximately 0;10, regardless of the language learning context.

So what is the role of CDS, and thus culture, in phonological development? This is an open line of inquiry (Cristia 2020). But one thing remains certain: the contribution of CDS, at least with its current definition, is not universal. Two options remain: either CDS is a helpful bootstrapping mechanism in phonological development in only some sociolinguistic settings, or we are mistaken that children employ it in phonological development at all.

Though Shneidman and colleagues (Shneidman and Goldin-Meadow 2012; Shneidman et al. 2013) primarily studied word learning as the outcome variable, their conclusions resonate for speech development as well. CDS may indeed aid phonological learning, just as Shneidman and Woodward (2016) agree that joint attention fosters lexical growth for children learning English in the United States. What Shneidman concludes, however, is that infants and children must learn to *use* CDS and joint attention to acquire language. The same could apply for speech development. CDS can provide clear, dispersed phonological category exemplars (cf. Cristia and Seidl 2014; Martin et al. 2015). Infants can use these exemplars to bootstrap into categories. This acquisition, in turn, helps infants parse those categories from the speech stream, such that the process becomes cyclical. But this does not resolve what environmental tools Tsimane- or Mayan-learning children, who do not receive as much CDS, use to disambiguate overlapping acoustic categories from their input.

Language Structure

Anatomy, articulatory routine, and exposure form a nearly-complete model of speech development. Yet this model misses a critical component of phonological development: language structure. Though language structure does not directly explain child speech phenomena, after all children and their adult interlocutors speak the same language, language structure interacts heavily with other factors in the model of speech development outlined here. For example, as previously outlined, children's inability to establish reliable articulatory routines means that they have more dispersed perceptuo-motor targets than adults. This does not have to vary by the language that the child is learning, but it can (Edwards and Beckman 2008). Will children have more stable vowel categories if vowels mark semantic distinctions, as in Hebrew or Arabic? CV syllables emerge before CVC in child speech - could this be mitigated if, as in Japanese, nasals are the only segment that occur in coda position? Thus, while language structure alone does not explain differences in adult and child speech, it interacts with other facets of phonetic and phonological development (unstable phoneme categories, order of consonant emergence) and is a crucial component to a model of phonetic development.

1.3 Illustrating the theory: Child coarticulation

Stability and coarticulation as indices of mature speech

Though I have outlined an emergentist theory of child phonology, I have so far remained neutral regarding the level of abstraction in children's representations. It is hardly a secret that this is a contentious issue. Are children's phonological representations highly abstracted, feature-based categories (Bernhardt and Stemberger 1998; Fikkert and Levelt 2008; Hale and Reiss 1998), episodic words and word-like traces (Ferguson and Farwell 1975; Vihman and Croft 2007; Vihman and Keren-Portnoy 2013), or some combination of the two (Fikkert and Levelt 2008; Swingley and Aslin 2002; 2007)?

Recall that children have highly unreliable articulatory patterns. This results in, messy, unreliable phonological representations - lots of scratches upon the metaphorical landscape described in section 1.2, but few deep, memorable crevices. Now imagine that instead of adult-like abstract segments, children have a word-level representation. It is somewhat abstract (i.e. entrenched), particularly if the word is frequent in the child's ambient or spoken language. However, the word is not so devoid of context so as to entirely mimic the segment-level phenomena that characterize adult phonology.

Consequently, when a child produces a word, they grasp for their phonological representation, which *is an entire word*. This whole word includes the coarticulation present when the child first heard the word in their ambient environment. Adults have sufficient experience with language that they abstract away from word types and can string individual speech segments together; children do not have this experience. Adults and older children also have the benefit of literacy and any resulting phonemic awareness, since early reading facilitates the development of segmental phonology (e.g. McBride-Chang et al. 2008; Metsala and Walley 1998; Perfetti et al. 1987; Stanovich et al. 1986; Wagner et al. 1994). As a result, children appear to "coarticulate" more than adults.

However, if children's underdeveloped phonological representations cause coarticulation, then we are no longer referring to the planned, efficiency-driven coarticulation of adult speech (Bradlow 2002; Whalen 1990). Instead, the tendency for children to overlap their speech gestures more than adults might be more appropriately termed CHILD INDISCRIMINATION. Children are not masters of coarticulation – they are simply unable to discern the internal structure of words because their segments are not yet abstract enough. The hypothesis here is that as children gain phonemic awareness and organize their speech into smaller units, they will coarticulate less between adjacent segments.

Of the many studies on child coarticulation, only two remark upon the crucial difference between child indiscrimination and adult coarticulation: "phenomena commonly lumped together under the heading of 'coarticulation' may have diverse origins...some forms...are an indication of advanced speech production skills whereas others may be a sign of articulatory immaturity" (Repp 1986:1618; see also Whiteside and Hodgson 2000).

Like acoustic category instability, child indiscrimination is a sign of immature speech. Like acoustic category stability, *real* coarticulation is a sign of mature speech. Note that studies concluding that children distinguish between segments less than adults simultaneously conclude that children show more inter- and intra-speaker variability than adults (Zharkova et al. 2011).

Some immediate concerns arise with this interpretation:

- *What about the finding that child indiscrimination occurs even in nonce words and sequences (e.g. Nittrouer 1989; 1996; Zharkova et al. 2011)?*

  Children analogize from their whole-word representations and coarticulate according to those traces. This is less surprising if you incorporate internal feedback to phonological representations (Grossberg 1978; Tilsen 2016; see section 1.2).

- *A poverty-of-the-stimulus argument: if children's coarticulation patterns are based on their own exposure to language, how can they coarticulate more than what they witness in the ambient language?*

  More evidence from naturalistic speech is needed to confirm that children coarticulate more than adults. The differences may not be as stark as findings from lab speech would lead us to believe. In their study on child coarticulation, Whiteside and Hodgson (2000) used a naturalistic speech sample and a picture elicitation instrument, but only found differences between adults and children in one of the five coarticulation measurements taken. The reason for this may be that adults adopt a more formal register for lab recordings, eliminating some of their habitual coarticulation. Children have not mastered diglossic speech registers. The fact that adults, but not children, are able to do this further supports the idea of increased abstraction in adult phonology.

- *What about long distance coarticulation?*

  Most studies focus on sequential coarticulation. But as already mentioned, those that test long-distance coarticulatory planning tend to find that adults coarticulate more than children (Barbier et al. 2013; 2015; Goffman et al. 2008; Repp 1986 cf. Rubertus et al. 2013). This is because these studies measured actual coarticulation patterns and not child indiscrimination.

- *Does child indiscrimination also apply to perseverative coarticulation?*

  This is unclear. Anticipatory coarticulation in children is more studied, and consequently understood, than perseverative coarticulation. But, according to the interpretation under Whole-Word Phonology, child indiscrimination should apply in both directions. Still, Fricke and Johnson (2012) report that only adults, not children, show perseverative coarticulatory effects in fricative+V sequences. However, there are several confounding factors. Fricke and Johnson studied children who were younger (1;1-3;1) than those in most other anticipatory studies. More importantly, they used naturalistic data, not lab speech. More research is needed in this domain. The theory outlined here would have difficulty accommodating a finding that children exhibit anticipatory, but not perseverative, “coarticulation."

This dissertation

This chapter has outlined a unified theory of speech development in children. Speech development patterns depend heavily upon factors that are universal to all children, such as the child’s distinct anatomy and the routine that children are able to establish given their transient physiology. Likewise, this theory has outlined how culture- and language-specific factors, such as the quantity and quality of input that a child receives as well as the structure of the ambient language, are highly deterministic.

This rest of this dissertation will focus on two components of speech development that were introduced in this chapter - language structure and exposure - and how these factors manifest in bilingual children acquiring South Bolivian Quechua and Spanish. Chapters 2 and 3 focus on language structure. Chapter 2 discusses the role of phonological inventory and asks if children can attain adult-like acoustic variability at an earlier age in a language with less vowel contrasts. In doing so, two vowel normalization techniques are compared to test if the child’s stage of anatomical development should be factored into acoustic analysis of child speech. Then, chapter 3 turns to morphological structure and asks if children’s speech production varies systematically by word structure. This chapter also validates two relatively novel acoustic measures of coarticulation on a dataset of English-speaking children. Finally, chapter 4 focuses upon the role of language exposure for children’s speech development. Children’s use of Quechua and Spanish is estimated through the use of day-long audio recordings. This language dominance is then used to predict the child speech patterns measured in chapters 2 and 3.

Chapter 2

A phonological factor: The roles of phoneme inventory and vocal tract morphology on speech variation

2.1 Introduction

The pervasive acoustic variability of adult speech is largely attributed to phonetic reduction and coarticulation. Variability in child speech production, however, is often the result of anatomical development: children's transient anatomy or underdeveloped motor routines may explain why their speech patterns vary from one production to the next (Gerosa et al. 2006; Lee et al. 1999; Vorperian and Kent 2007).

It is this topic, the development of vocalic variation in spoken language, that the current chapter addresses. Mastering the production of vowel categories clearly poses a challenge throughout childhood. What factors - anatomical and linguistic - predict children's vowel variability? Vocalic development may be an essential component of phonological development. Listeners expect large amounts of phonetic reduction and coarticulation in spontaneous adult speech. As a result, speech which lacks this normal type of variability has reduced intelligibility (Aylett and Turk 2004; Ménard et al. 2007). The acoustic vowel space is also used as a metric of speech development in typically-developing children and children with hearing impairment after hearing device implantation (Schenk et al. 2003; Vorperian and Kent 2007).

To that end, this chapter examines two factors that may predict the development of normal patterns of vowel variability in children: phoneme inventory and vocal tract anatomy. Most studies that find that children exhibit high variability in vowel production have been drawn from languages with relatively large vowel inventories such as English (Lee et al. 1999) and French (Ménard et al. 2007). But vowel inventory size and intra- category variability – how dispersed each phoneme category's productions are from the category mean – may be negatively correlated in adults (Recasens and Espinosa 2006). Do children learning a language with fewer vowel contrasts achieve adult-like levels of vowel category stability earlier? Additionally, children first master consonants that are most frequent in the ambient language (Edwards and Beckman 2008). Consequently, one may expect that in systems with fewer vowel contrasts, where each vowel is more frequent, children may be able to master vowel contrasts sooner. Children are expected to be equally, if not more, variable with dispersed phoneme categories than adults when acquiring a language with few vowel contrasts. This hypothesis is tested in a cross-sectional sample of child and adult speakers of South Bolivian Quechua, henceforth "Quechua," a language with three phonemic vowel contrasts /a, i, u/ and two allophonic vowel contrasts [e, o].

The other potentially predictive factor for children's vowel production addressed in this chapter is vocal tract anatomy. It is well-known that the child vocal tract is not simply a miniature version of the adult vocal tract: the ratio between supraglottal cavities changes throughout development in a non-linear fashion (Fitch and Giedd 1999; Vorperian et al. 2005; Vorperian and Wang 2009). To evaluate this potential role of vocal tract anatomy on children's vowel production, two formant frequency scaling techniques, one age-independent and one age-dependent, are applied to the Quechua vowel data. Should the normalization results from these two scaling techniques differ, this would suggest that vowel normalization technique may be contingent upon a child's age/developmental stage. The results would also suggest that morphological differences in cavity size are not negligible sources of variation in child speech (cf. Ménard et al. 2007; Turner et al. 2009).

2.2 Background

Vocalic development

Production patterns such as cluster reduction and place assimilation that characterize early child phonology rapidly subside as typically developing children attune their motor planning skills (McAllister Byun and Tessier 2016). Yet children will still undergo years of highly variable phonetic production (Hazan and Barrett 2000; Lee et al. 1999; Pettinato et al. 2016). Throughout this dissertation, this process of attaining adult-like levels of within-category variability throughout childhood is referred to as ACOUSTICAL NARROWING, following the well-known phenomenon of perceptual narrowing in infant phonology.

Acoustical variability in the production of vowels is ubiquitous, even in adults (Hillenbrand et al. 1995; Peterson and Barney 1952). Variability in child speech is even more widespread, but the sources of the variation may differ. Between and within-speaker variation in children could result from underdeveloped coarticulatory planning and gestural movement (Nittrouer 1993; Nittrouer et al. 1996). However, this acoustic variability in children - and even between adult women and men - can also be attributed to anatomical differences (Denny and McGowan 2012a; Denny and McGowan 2012b). For example, children's vocal tracts are shorter than adults' and adult females' vocal tracts are shorter than adult males'. However, both children's and, to a lesser extent, adult females' vocal tracts, also exhibit a larger overall ratio of palate to pharynx length. In some cases, these anatomical differences have undeniable acoustic consequences: shorter vocal tracts translate into higher resonant frequencies during speech production. In other cases, such as differences in the relationship between supraglottal cavity lengths, the acoustic effect of anatomical difference between children and adults is unclear or unattested (Ménard et al. 2007; Turner et al. 2009).

Acoustic studies of vowel production in English have long noted that within speakers, children's vowels were more dispersed and variable than adults'. Comparing English vowel production in adults and a sample of 84 children (3;0-13;0), Eguchi and Hirsh (1969) concluded that the children showed more intra-subject acoustic variability along F1 and F2 than the adults. In a large, cross-sectional study of American English vowels measured in 436 children aged 5;0-17;0, and 56 adults, Lee et al. (1999) also found that intra-subject and intra-age group variability of f0, the first three formant frequencies, and vowel duration decreased with age. Children attained adult-like levels of formant frequency and duration variability around 12;0.

A meta-analysis of 14 studies on vocalic development came to similar conclusions concerning formant variability in English-speaking children (Vorperian and Kent 2007). Quantifying variability as the size of the vowel space, Vorperian and Kent (2007) concluded that the F1-F2 quadrilateral and F1-F2-F3 space decrease in size as a function of age in a sample of children aged 4;0-18;0 and adults. Most recently, work on older English-speaking children, aged 9;0-14;0, found that the size of a triangular vowel space area ([i], [æ], [ɔ]) and formant frequency ranges also decreased with age (Pettinato et al. 2016).

The consistent pattern of acoustical narrowing in vowel production is most often attributed to anatomical maturation (e.g. Lee et al. 1999). If acoustical differences between children and adults have anatomical origins, vowel development should not differ greatly cross-linguistically. The results of Ménard et al. (2007) support this claim. The authors examined the first three formant frequencies in vowels produced by French-speaking adults and children (3;7-4;2, 7;9-8;3) and found that younger speakers showed more intra-speaker spectral variability, with more dispersed vowel categories. Non-rhotic vowels are some of the earliest segments that young children and infants produce. However, taken together, results from these studies suggest that children do not control acoustic variability in formant frequency production at adult-like levels until early puberty.

Other studies have found exceptions to acoustical narrowing patterns. Nittrouer (1993) elicited the sequence /ə/-consonant-/ɑ, i, or u/ in English- speaking adults and children (3;0, 5;0, 7;0). Child variability in the first two formants of /a, i, u/ did not unilaterally reduce with the children's age. In fact, as early as 3;0, the children demonstrated minimal, almost adult-like levels of F1 variability. F2 variability, however, continued to decrease well after this age. Nittrouer explained this as a function of non-uniform motor development: children may master vertical jaw movement early, but other gestures (e.g. tongue dorsum fronting) require further maturation.

McGowan et al. (2014) examined naturalistic, adult-directed speech of six American English children (1;6-4;0). While children's within-subject formant frequencies were variable, with highly-dispersed phoneme categories, front vowels were more stable than back.

Furthermore, the children's formant variability did not appear to change over development suggesting that spectral variability may not always correlate negatively with child age, at least in younger children.

With the exception of some by-vowel effects, and McGowan et al. (2014)'s finding from naturalistic speech samples, acoustical narrowing appears to be a common developmental trend. This is perfectly cogent: child variability stems from immature motor development and the transient articulatory-acoustics mapping that children must continuously update as their anatomy changes. However, there are other factors that likely influence phonetic production. This chapter examines how phonological structure, a language-internal factor, could also mitigate acoustical narrowing in children's speech development.

Formant frequency scaling in child acoustics

The speech signal reflects vocal tract shape and configuration: vocal tract morphology, particularly length, varies by speaker gender and age. Children, for example, have relatively short vocal tracts. The average vocal tract length length for five- to six-year-olds is 9-11cm compared to 15cm and 18cm for female and male adults, respectively (Fitch and Giedd 1999; Vorperian et al. 2005). Children's shorter vocal tracts result in their characteristic higher frequency cavity resonances. Elsewhere, the disproportionately greater growth of the pharyngeal cavity relative to the oral cavity in young boys may result in a sexual dimorphism of formant frequencies by age 4;0, though imaging studies suggest that overall sexual dimorphism by *length* does not emerge until 6;9 or later (Vorperian et al. 2005).

To contend with this inter-speaker anatomic variability, phoneticians have long employed vocal tract length normalization techniques (Fant 1975; Johnson 1988; Lobanov 1971; Nearey 1977; Nordstrom and Lindblom 1975). By employing normalization, acoustic measures, which would otherwise be modulated by the filter of vocal tract length, can be isolated. This allows researchers to remove vocal tract length, one of the primary sources of between-speaker speech variability, from the speech signal.[1]

A thorough comparison of vocal tract length normalization techniques is beyond the scope or objectives of this chapter. However, vocal tract morphology may need to be factored into reports of formant frequencies in a cross-sectional study of children's vocalic development for two reasons. First, most obviously, the vocal tract will lengthen with age. Normalization must be employed to factor out differences between children that are due to anatomical development, and isolate those that are due to, for example, psycholinguistic maturation and speech planning.

However, if length were the only difference between adult and child vocal tract morphology, researchers could employ the same normalization techniques to factor out age differences between children as they do to factor out gender differences between adults. A second dif-

[1]Besides vocal tract length, only phone identity accounts for more variability in the speech signal (Turner et al. 2009).

ference between adult and child vocal tract morphology suggests that vowel normalization by age may not be so straightforward.

As alluded to previously, not only is the child's vocal tract relatively short in comparison to adult models, but the ratio between the oral and pharyngeal cavities also differs as a function of age (primarily) and gender (Goldstein 1980; Vorperian et al. 2011). In adult males, for example, the lowered position of the larynx means that males have a disproportionately longer pharyngeal cavity than adult females, even if the length of the overall vocal tract is held equal (Johnson and Sjerps 2018). The ratio between these anatomical cavities also differs between children and adults, and crucially for the purposes of this chapter, exhibits NON-UNIFORM GROWTH meaning that at times one anatomical cavity grows faster than another (Vorperian and Kent 2007; Vorperian and Wang 2009).[2] Young infants start out with a disproportionately large oral cavity, in comparison to the pharyngeal, because the infant's head is large relative to the neck. This accommodates the infant's larger tongue and facilitates sucking in the first year of life (Crelin 1987). Non-uniform vocal tract growth in child development then ensues as the pharyngeal cavity lengthens (Figure 2.1), caused in part by a shift of the glottal opening from the upper to lower larynx (Lieberman and Crelin 1972).

Results from comparisons of vowel normalization techniques suggest that changes in cavity ratio are linear, and so ratio differences by gender may not have a large effect upon acoustic output. The two techniques for scaling[3] formant frequency measurements are used to illustrate this argument: UNIFORM SCALING TECHNIQUES, or techniques that employ only one scaling factor (generally vocal tract length or a correlate) (Nearey 1977; Nordstrom and Lindblom 1975), and NON-UNIFORM SCALING TECHNIQUES, or techniques that employ multiple scaling factors by vowel or formant (Fant 1966, 1975; ; Nearey 1977; Umesh et al. 2002).

In adult speakers, several studies have concluded that uniform scaling techniques factor out anatomical differences by gender more or less as well as non-uniform scaling techniques (Johnson and Sjerps 2018; Nordstrom and Lindblom 1975; Turner et al. 2009), with Adank et al. (2004) concluding that the best normalization techniques are those that employ information extrinsic to the vowel in question (i.e. additional vowels), but intrinsic to the formant (i.e. isolating calculation of the scaling factor to a single formant). Consequently, normalization techniques traditionally disregard cavity ratio discrepancies between men and women, and focus exclusively on vocal tract length.

Differences in anatomical cavity size in children, however, have long been proposed to account for speech variability between children of different ages/sizes and between children and adults (Fant 1966; 1975). It is unclear if non-uniform scaling in child acoustics is necessary, or if more straightforward uniform scaling normalization suffices.

[2]Analysis here is limited to the relationship between resonances in oral and pharyngeal cavities, not nasal. It is, however, important to note that the child's nasal cavity is likewise elongated relative to adult models, again owing to the child's larger head (Lieberman and Crelin 1972).

[3]Here I limit discussion to scaling techniques, not vowel-extrinsic or -intrinsic approaches to normalization.

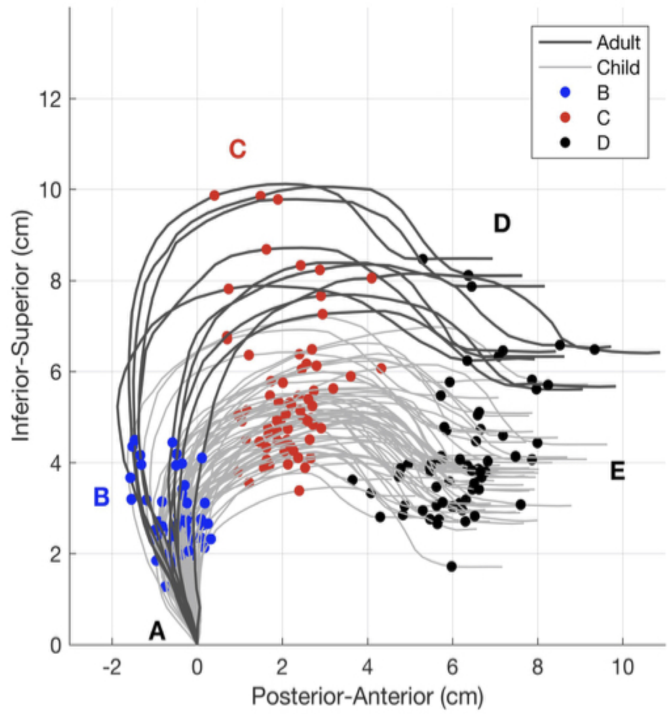


Figure 2.1: Vocal tract midline traces in a cross-sectional sample of 58 male children aged birth-12;0 (light gray) and 10 adult males (dark gray). The back, pharyngeal cavity lengthens with age, eventually overtaking the oral cavity in length. Reprinted from Story et al. (2018).

Evaluating the question of uniform versus non-uniform scaling, Turner et al. (2009) used the developmental vowel data in English from Peterson and Barney (1952) to account for various factors that could explain formant variability. The authors found that phone identity and vocal tract length were the primary contributors to speech variability, accounting for approximately 80% and 18% of formant frequency variability in the dataset, respectively. The remaining variability within the model was attributable not to speaker-specific articulatory strategies, but rather to formant tracking errors. Furthermore, the cavities' growth functions were linear, and the functions did not differ across children, male adults, or female adults. Turner et al. (2009) thus conclude that developmental changes, such as oral to pharyngeal cavity ratio, are statistically irrelevant for contending with variability between children and adults (and adult males and females) and sufficient variability can be accounted for with information on vowel identity and vocal tract length. The authors also replicate these results using the vocalic database from Lee et al. (1999).

Similar conclusions were drawn in the articulatory simulations and acoustic analysis of natural vowels in Ménard et al. (2007). Fifteen subjects, n=5 aged 4;0, n=5 8;0, and n=5 adults, produced 10 French vowels as isolated tokens in carrier phrases. If non-uniform growth played a large role on acoustic output, then differences between formants affiliated with the back cavity - which again is disproportionately larger in adults than children - should be non-linearly lower than formants affiliated with the front cavity (Ménard et al. 2007:3). However, the researchers did not find evidence of differences between the natural vowel formant values based on cavity affiliations. This analysis is extended in a simulation of peripheral vowels /i/, u/, and /a/. Two productions were synthesized: one production as if the model four- and eight-year-old were aiming for *acoustic* targets (attempting to mimic adult-like acoustic patterns) and another production as if the model was aiming for *articulatory* targets (attempting to mimic adult-like gestures). The authors found that the synthesized acoustic targets match the natural vowel data better than the synthesized articulatory targets and thus conclude that children are adapting their articulatory strategies to compensate for their vocal tract morphologies (see also Ménard and Boë 2000).

These works thus suggest that while the pharyngeal cavity does indeed grow faster than the oral cavity, young speakers may learn to compensate for morphological differences in anatomical cavities by adjusting tongue position or constriction location in the vocal tract. In the case of a young child producing [u], this would mean that the labial and tongue gestures required to create the Helmholtz resonators affiliated with F1 and F2 at age 4;0 are different than the gestures required to create the same resonators at age 8;0, when the pharyngeal cavity has grown longer (Ménard et al. 2008).

The current study will contrast two methods of formant frequency scaling, one uniform and one non-uniform, to further evaluate the hypothesis that children compensate for their vocal tract morphologies .[4] If similar normalization results are derived from both approaches, this would suggest that non-uniform scaling may not be necessary, and vocal tract length alone is sufficient to factor out between speaker anatomical difference in children.

The uniform scaling method employed here, referred to as ΔF, builds on the vocal tract length estimation procedure outlined in Lammert and Narayanan (2015). ΔF uses formant frequencies $f1$ through fn, where n refers to the highest integer formant measured (Lammert and Narayanan 2015), to compute vocal tract length. This length is, in turn, the sole scaling factor employed during normalization; neither vowel nor formant-pair identity are factored into the formant scaling. To calculate ΔF, the interval between the formants is calculated (Equation 2.1). Then, vocal tract length is estimated on the basis of the change between formants (ΔF), which is assumed to be constant (Equation 2.2).

Initially:

$$\Delta F = \frac{\frac{\mu_{F1}}{0.5} + \frac{\mu_{F2}}{1.5} + \frac{\mu_{F3}}{2.5} + \cdots \frac{\mu_{Fn}}{n-.5}}{n} \tag{2.1}$$

Followed by:

$$L = \frac{34000}{2 \cdot \Delta F} \tag{2.2}$$

[4]See Adank et al. (2004) and Johnson and Sjerps (2018) for a more complete description and comparison of scaling and normalization methods.

where L is the estimated vocal tract length.

The non-uniform scaling method employed is the Lobanov normalization method (Lobanov 1971). The Lobanov approach transforms frequency measurements via z-score normalization. To scale a particular formant n, the mean formant frequency of n, μ_{vn}, is averaged over all of the speaker's vowels (v), and subtracted from the formant measurement to be normalized (F_n). This difference is then divided by the standard deviation of the n formant measurements (σ_{vn}) (Adank et al. 2004) and, crucially, the process is performed separately for each formant (and F0) (Equation 2.3).

$$F_{n'} = \frac{F_n - \mu_{vn}}{\sigma_{vn}} \tag{2.3}$$

Section 2.2 outlines further predictions for these scaling techniques.

Current study

The primary objective of this study is to view how a different phonological system may mitigate developmental trends in children's vowel production. This is tested in Quechua, a language with three phonemic vowels, /i, a, u/, and two allophonic vowels, [e, o]. In Quechua, the allophonic vowels are derived in uvular environments (See Gallagher (2016) for further details). The Quechuan variety studied here, South Bolivian Quechua, is a Quechua-II/C language with over 1.6 million speakers in southwest Bolivia and northwest Argentina (Torero 1964).

The first research questions asks if a language's phonological inventory can mitigate vowel development trends in children:

1. Do the first two formant frequencies undergo acoustical narrowing from ages 4;0-10;0 in a language with a relatively small vowel inventory? Or have Quechua-speaking children already acquired adult-like acoustic variability by age 4;0?

I predict that, within-subjects, Quechua-speaking children will acquire adult-like levels of within-category vowel variability earlier than children learning languages with large vowel inventories (English, French). Languages with larger vowel inventories seem to show that phoneme categories are less dispersed in acoustical space (Manuel and Krakow 1984; Recasens and Espinosa 2006) or the difference in inventory size must be extreme to affect variability (Recasens and Espinosa 2009b). However, others found that intra-vowel category acoustic variability does not vary by inventory size – languages with large and small inventories show similar acoustic dispersion (Bradlow 1995). These studies reported on adult speech, but vowel inventory size may impact child speech variability as well.

The secondary objective of this study is to contrast a uniform scaling method (ΔF [Johnson and Sjerps 2018]) with a non-uniform scaling method (Lobanov [Lobanov 1971]) for the normalization of formant frequencies between children of different ages and between children and adults.

2. Will the Lobanov and ΔF techniques produce similar formant frequency patterns across the same age groups?

If the results of the two scaling techniques are the same, this study can conclude that uniform scaling is sufficient to factor out anatomical difference between children of different ages and between children and adults. This conclusion would suggest that children adapt their articulatory gestures to compensate for their vocal tract morphologies, as previous work has suggested (Ménard et al. 2007; Turner et al. 2009). If the results of the two scaling techniques result in different formant frequency patterns, then assuming uniform scaling by formant and vowel identity over the course of development may be premature. This result would suggest that the scaling factor employed to normalize formant frequencies in child speech may require adjustment based on chronological age, formant, and/or vowel. Here difference between the scaling techniques is quantified as the difference in category dispersion between Lobanov-scaled vowels and ΔF-scaled vowels, where more variable categories in Lobanov-scaled vowels indicate that there is increased variability to be accounted for after factoring out vocal tract length.

2.3 Methods

Participants

86 children aged 4;0-10;11 and 10 female adults (adult μ_{age}=23, σ=5.46, three did not report) participated in this study. Children's age distribution was as follows: 10 four-year-olds (μ=4;6, σ=0;4, one did not report exact birth date)[5], 11 five-year-olds (μ=5;7, σ=0;5, one did not report), 13 six-year-olds (μ=6;5, σ=0;3), 21 seven-year-olds (μ=7;8, σ=0;4, five did not report), 13 eight-year-olds (μ=8;7, σ=0;4, one did not report), 8 nine-year-olds (μ=9;6, σ=0;3, three did not report), and 10 ten-year-olds (μ=10;6, σ=0;5, three did not report). All participants were bilingual Quechua-Spanish speakers and were living in or around a mid-size town in southern Bolivia at the time of data collection. The child participants were either recruited at a local school where I was volunteering (n=18) or through personal contacts in the surrounding communities (n=68). The adult participants were recruited through local contacts.

Most children had typical speech and hearing development, per parental or teacher self-report. The caregivers of 3 children (2 seven-year-olds, 1 five-year-old) stated that their child was late to begin talking.[6] Note that these communities are medically under-served so some language delays/impairments may go unreported. Additionally, 3 children had lost

[5]When I say that the age was not reported, this means that the caregiver was able to determine the child's age, but the exact birth date was not available.

[6]Late talker status was not collected from the participants recruited from the school.

one or more of their front teeth (top or bottom) at the time of recording.[7] I attempted to complete a hearing test with the children, however it became clear after attempting with a few of the children that I was collecting false positives during the test as the children were nervous about making a mistake. Consequently, I cannot say with absolute confidence that all children would have passed a standard hearing screening. The adult participants did not report any speech or language disorders.

Socioeconomic status (SES), usually implemented as mother's level of education in child development research, is an important predictor of child language development in the United States (Hoff 2003; Pace et al. 2017). However, it is not clear that SES is predictive of language outcomes in all cultural or linguistic contexts. Specifically, it is unknown if SES predicts language outcomes in Bolivia as a whole, in these speech communities specifically, or for children learning Quechua. Still, I attempted to collect information on SES as it is an important predictor in many other cultural contexts.

I was able to collect information about the central caregiver's education level (usually the central caregiver was the mother, but occasionally it was the grandmother) from most of the families recruited from the surrounding community, but not those recruited at the school. There is no a priori reason to believe that the distribution of socioeconomic strata of the children recruited at the school would differ from those who were recruited from elsewhere in the community. That is to say, the children from the surrounding communities attended a similar school, just in a different location from where the school children were recruited and tested. There were 13 sibling pairs and 2 three-sibling pairs (no twins), in the child sample resulting in 69 unique caregivers. For the 35 caregivers of the children recruited from the surrounding community that SES information was obtained from, the caregivers' education levels were: 18 of the caregivers from the community had completed some primary school (less than six years of education), 5 had completed primary school (6 years of education), 4 had completed the equivalent of a middle school (10 years of education), 1 had completed secondary/high school (13 years of education), 3 had not received any formal schooling, and 4 did not report (Table 2.1).

[7] I report this because the presence of front teeth could have notable consequences for speech acoustics (e.g. anterior fricatives). This information is not typically reported in speech development research, but arguably should be.

Table 2.1: Maternal education distribution of study participants

Mat. Ed. in Years	N
0	3 (8.57%)
<6	18 (51.43%)
6	5 (14.29%)
10	4 (11.43%)
13	1 (2.56%)

An additional indicator of socioeconomic status in these indigenous communities in Bolivia may be the central caregiver's familiarity with Spanish. This is generally correlated with mother's education level as only women who have had the opportunity to attend school learn to speak or read in Spanish. A coarse estimation of the central caregiver's level of Spanish-Quechua bilingualism was collected from 33 of the 35 unique caregivers recruited from the surrounding communities: 8 of those caregivers were monolingual Quechua speakers, 5 were Quechua-dominant but spoke or understood some Spanish, and 20 were bilingual Quechua-Spanish speakers (Table 2.2). Again, there is no a priori reason to assume that the distribution of SES or maternal education would differ in this subset of the overall sample.

Table 2.2: Maternal Quechua-Spanish language experience

Lang. Experience	N
Monolingual Quechua	8 (22.86%)
Quechua dominant	5 (14.29%)
Bilingual Quechua-Spanish	20 (57.14%)

Tasks

Children aged 5;0 and up completed four tasks, all prompted with pictures, in the following order: 1) real word repetition, that included a morphological extension component, 2) Quechua nonword repetition, 3) Spanish nonword repetition, and 4) additional real word repetition with morphological extension. The children aged 4;0-4;11 completed only the first three tasks. For the word repetition tasks, children repeated the real words or nonwords after a model speaker (explained below). Results from the morphological extension tasks are reported in Chapter 3. Nonword repetition tasks are not discussed in this dissertation. The adult participants only completed the two real word repetition tasks as even ten-year-olds approached ceiling on the nonword tasks.

The 4;0-4;11 children did not participate in the morphological extension task. This decision was made to reduce the amount of time that the youngest children had to sit still for the tasks. Having the children complete the tasks on separate days, to lessen the time commitment each day, was not feasible because it was often difficult to contact the children, find them at home, or find time during the school day to complete the tasks. The intent was to keep the testing as uniform as possible between children of different ages and allowing partial completion of tasks would have introduced a large amount of variation.

The entire testing procedure was completed in one sitting and took approximately 30-40 minutes per child (testing time was equal for the younger children although they completed fewer tasks because they required more time to complete the tasks). The adults completed the repetition tasks in approximately 20 minutes.

For their participation, all children could choose an item from a toy bag. Children at the school additionally received academic assistance including lessons on English and Spanish language and American culture from me when I was volunteering. I also donated school supplies and materials to the school. The adult participants and caregivers of children from the surrounding communities who did not attend the school instead received a small monetary sum.

The order of the real word and nonword tasks was not counterbalanced between children. This decision was made because several children were nervous, especially at the beginning of testing. Completing the real word task first was a way to familiarize the children with the procedure of hearing words and then repeating them into a microphone before advancing on to the less-familiar nonword task.

Stimuli

The real word repetition tasks consisted of 56 high-frequency Quechua nouns (plus 6 training trials for 62 total lexical items) that are familiar to children learning Spanish and Quechua in southern Bolivia (full stimuli listed in Table 2.18 in the Appendices). There is no equivalent to the *Macarthur Bates Communicative Development Inventory* (Fenson et al. 2007), which reports stages of age-normed vocabulary development, for any Quechuan language or Bolivian Spanish. Nor is there a large, transcribed child-directed speech corpus for these languages to infer vocabulary development. For these reasons, I confirmed children's knowledge of the test items via a pre-test that demonstrated that children as young as 3;0 could spontaneously name all items. Female caregivers also confirmed that children as young as 3;0 should recognize the items in Quechua.

The real word stimuli came from recordings of an adult female bilingual Quechua-Spanish speaker. These recordings were digitized at a sampling frequency of 44.1 kHz using a portable Zoom H1 Handy Recorder. Stimuli were normed for amplitude between words, but not duration, since some words had ejectives, fricatives, etc. that are temporally longer. The real word picture stimuli were color photographs of the objects.

Children in these communities have limited exposure to technology (some mothers have flip phones but many of the children are unfamiliar with larger computing devices). Consequently, instead of presenting each picture stimulus on a screen, which could have been culturally inappropriate, pictures were presented on individual pages clipped into an 11 x 12.4" plastic binder. For this reason, the words were not entirely randomized for each participant. Instead, two different randomized lists were created and were counterbalanced between participants with half of the children and half of the adults receiving the first list and half of the children and adults receiving the second list. Repetitions of the same stimulus were always separated by at least two different stimuli and were presented with a novel photo of the item each time.

Data collection

For the experimental phase, participants were seated on the ground or on a stool, side-by-side with the experimenter. Audio stimuli were were played for the experimenter and participant from an iTunes playlist run on an iPhone 6. Each participant wore AKG K240 binaural studio headphones and the experimenter wore Apple earpods to follow along with the experiment; both headphones were connected to the iPhone with a Belkin headphone splitter.

For data collection, the participant first heard the audio stimulus (a bare noun) and was simultaneously presented with the accompanying photo in the binder. Then, each participant was instructed to repeat the word after the model speaker. For the second production, the participant was to inflect the target word with a given suffix. For the children, this inflection was elicited by placing a large plastic toy insect on top of the picture stimulus and prompting the child, "Where is the bug?" to which the child produced the word with the correct suffixal carrier e.g. *llama-pi* (llama-LOC, "on the llama") (The morpheme varied; see Chapter 3 for details). Thus, each of the target words was elicited twice per trial. Participants' responses were always repeated after the model speaker, and were not spontaneous. Ideally participants would not have had to repeat after a model speaker. However, in an earlier version of this task, I found that the youngest children sometimes could not follow the task when they were not prompted to repeat the word (Cychosz 2019). Elicited imitation is a common technique in studies of children's vowel development (e.g. Lee et al. 1999), so the methodological decision to have children repeat the prompts follows previous work on this topic.

The adult participants were instructed to name the item in the photo in a carrier phrase: *Noqa nini _____-pi iskay kutita* ("I say in the _____ two times."). Then the experimenter would manually advance to the next stimulus item. Participant responses were recorded with a portable Zoom H1 Handy Recorder at a 44.1 kHz sampling rate. Children were rewarded with stickers throughout the task and many additionally chose to help the experimenter flip through the pages of the binder.

Data analysis

The vowel data analyzed here come are a subset (n=24) of the words repeated in the real word tasks (Table 2.3). These words were selected because the target vowels fell in stressed and, where possible, word-medial position. Taking vowels from word-medial position avoids the effect of word-final devoicing and loss of spectral energy. Additionally, words were selected to avoid flanking consonants that would exert the strongest coarticulatory effects on the vowels (glides and laterals). Finally, note that the mid-vowels /e/ and /o/ are derived only in uvular environments (see Gallagher (2016) for further detail), so the flanking consonant in the words to elicit /e/ and /o/ was almost always uvular. Since the 4;0 children did not complete the morphological extension task for time and maturity reasons, tokens in the inflected form were not collected for that age group.

Table 2.3: Vowels analyzed in current study (in bold) and their lexical context

Vowel	Syllabified lexical item*	Translation
[a]	ˈ**a**.pi	‘corn/citrus drink’
[a]	ˈp**a**m.pa	‘prairie’
[a]	ˈp**a**.pa	‘potato’
[a]	ˈth**a**.pa	‘nest’
[a]	ˈm**a**.ma	‘mom’
[a]	hatunˈm**a**.ma	‘grandmother’
[i]	ˈch**i**.ta	‘sheep’
[i]	ˈt’**i**.ka	‘flower’
[i]	ham.ˈp**i**.ri	‘healer’
[i]	ham.pi.ˈr**i**-pi	‘healer-LOC’
[i]	a.ˈp**i**-pi	‘corn/citrus drink-LOC’
[i]	q’e.ˈp**i**-pi	‘bundle-LOC’
[u]	ˈp**u**n.ku	‘door’
[u]	ˈp**u**n.chu	‘poncho’
[u]	ju.k’ˈ**u**.cha	‘mouse’
[u]	ˈr**u**n.tu	‘egg’
[u]	ˈs**u**n.kha	‘beard’
[u]	u.hˈ**u**.t’a	‘sandal’
[e]	ˈp’**e**.sqo	‘bird’
[e]	ˈq’**e**.pi	‘bundle’
[e]	qol.ˈq**e**-pi	‘money-LOC’
[o]	ˈq**o**l.qe	‘money’
[o]	al.ˈq**o**-pi	‘dog-LOC’
[o]	p’e.ˈsq**o**-pi	‘bird-LOC’

* ˈ indicates stress, ’ indicates ejective, ’.’ indicates syllable boundary, ’-’ indicates morpheme boundary

The decision to elicit the vowels in real words instead of nonce items was made for a couple of reasons. Coarticulation between vowels and neighboring sounds is a real concern for a study of vowel variability. However, recall that all of the vowel stimuli came from the same words (thus the vowel’s environment and coarticulatory influences should be relatively constant between children). Second, the objective was to elicit *Quechua* vowels, not Spanish. But since the two languages’ vowel categories completely overlap (both languages have five vowels /i, a, u, e, o/, though the mid-vowels are allophonic in Quechua), it could be difficult to determine which language system the participants were using. If the participants repeated context-neutral vowels (e.g. say [æ] like “cat”), there was concern that the children would default to Spanish vowels, instead of Quechua. This was especially relevant since many of the children were tested in an environment where they are used to speaking Spanish (school) by someone who looks more likely to be a Spanish speaker than Quechua speaker (the white researcher). By eliciting the vowels within Quechua words, there was little doubt that the children were producing Quechua vowels, not Spanish.

Alignment

Each participant's audio file was first manually aligned to the word level in Praat (Boersma and Weenik 2019). To align to the phone level, a Quechua forced aligner was trained on all of the participants' data using the Montreal Forced Aligner (McAuliffe et al. 2017). Finally, the phone-level alignment was hand-corrected by one of two trained phoneticians. Alignment was conducted auditorily and by reviewing the associated acoustic waveform and spectrogram in Praat. Alignment was conducted auditorily and by reviewing the associated acoustic waveform and broadband spectrogram in Praat.

Acoustic measures of vowels can be sensitive to alignment decisions, so a number of parameters were set prior to alignment to ensure reliability. Word-initial plosive, affricate, and ejective onset corresponded to the burst. The start of vowels corresponded to the onset of periodicity and formant structure in the waveform and spectrogram. Nasals were identified by the presence of anti-formants in the spectrogram and dampened amplitude. Glide-vowel sequences were delimited visually, or when this was not possible, half of the vowel-glide sequence was attributed to the vowel and half to the glide. There is some variability in the realization of mid-vowels in Quechua speakers: vowels were transcribed phonemically.

To evaluate agreement between the phoneticians conducting the alignment, both phoneticians aligned two randomly-selected word lists, one from a child aged 5;9 and another from a child aged 7;4. For the 5;9 child's list, the difference between the aligners' average consonant duration was 4ms and the average difference in vowel duration was 2ms. Pearson correlations between the aligners for the 5;9 child's list were significant for consonants: $r=0.86$ $p<.001$, 95% CI=[0.83, 0.89] and vowels: $r=0.94$ $p<.001$, 95% CI=[0.93, 0.96]. For the 7;4 child's list, the difference between the aligners' average consonant duration was 2ms and the average difference in vowel duration was 2ms. Pearson correlations between the aligners for the 7;4 child's list were significant for consonants: $r=0.98$ $p<.001$, 95% CI=[0.97, 0.98] and vowels: $r=0.95$ $p<.001$, 95% CI=[0.94, 0.96]. The high levels of agreement between aligners suggest high fidelity to the alignment protocol.

Acoustic measurements

The first three to four formant frequencies were automatically extracted from each vowel at three evenly-spaced points. The spectral analysis of child speech, and formant frequency tracking in particular, can be challenging: in addition to a propensity for breathiness, the high fundamental frequencies of child voices mean harmonics are widely dispersed and the spectral shape can be undersampled. Formant measures derived from trackers employing linear predictive coding (LPC) analysis, where measurements can be influenced by adjacent harmonics in the spectrum (Atal and Schroeder 1974; Ménard et al. 2007). There have been different approaches to contend with the formant tracking problem in child speech including automatically tracking formants and excluding outliers beyond two standard deviations (Lee et al. 1999) or excluding problematic tokens from hand-verified LPC spectra (Sussman et al. 1996).

For the current analysis, a series of custom Python notebooks were written to run three different formant trackers: Inverse Filter Control Formant (Watanabe 2001), Entropic Signal Processing Systems (ESPS)'s *covariance*, and ESPS's *autocorrelation*. These notebooks and all other scripts used to generate the results in this chapter are available open-source in the Github project associated with this chapter (`https://github.com/megseekosh/vocal_tract_vowel`). The covariance and autocorrelation formant tracking methods employ LPC. Inverse Filter Control employs inverse filters that are modulated by frequency distributions such that only the spectral shape determines the estimation of the frequencies.

Inverse Filter Control does not permit specification of the filter order but does include a three-level parameter to specify speaker gender/age: male, female, or child. In the current study, the 'child' parameter was specified for the child participants and 'female' for the adults. For the ESPS formant trackers, an LPC filter order of 10 was specified for the children and an order of 12 was specified for the adults. Given that the number of formants tracked for a given filter order in ESPS is $(FilterOrder - 4)/2$, the ESPS formant tracking functions could only track three formants for the children. Consequently, the fourth formant was only tracked in the adults.[8]

The triple formant tracker script was run over each speaker's audio file. Formant measurements were recorded at 25%, 50%, and 75% of each vowel. This resulted in three measurements (one from each tracker) for each formant, at each of the three time points. In this way, anomalous measurements from any single tracker did not have an outsize influence. Only the vowel midpoint is analyzed in this work.

After the median measurement from the three trackers was computed, an additional cleaning procedure was conducted on the vowels. This procedure was designed to remove measurements where all three trackers may have erroneously tracked the wrong formant or reported a measurement that did not seem likely given a speaker's median formants. The cleaning procedure was conducted as follows: the median absolute deviation (MAD) was first measured for the midpoint formants (at 50% of the vowel token) of every speaker's tokens of a given vowel.[9] Then, an upper and lower MAD boundary was computed (plus or minus three MADs from the median). All tokens falling above or below three MADs from each speaker's midpoint vowel median were then removed. This resulted in the loss of approximately 4-12% of the data by age group, as Tables 2.4 and 2.5 demonstrate. No large

[8]The youngest children's formants (under age 7;0) were additionally tracked with an LPC filter order of 8, but I found that the formant tracker performed markedly worse - F3 was frequently mistracked as F2 - so I confidently continued with the filter order of 10 for all children and 12 for adults.

[9]The median absolute deviation (MAD) was used, instead of standard deviation of the mean, to further avoid the influence of outliers which this procedure was designed to remove.

differences in data removal were noted by age group, vowel, or individual formant.[10]

Table 2.4: Token counts and percentage of each formant removed, by age, using three MADs from median criterion

Age	F1 (MAD)	F2	F3	F4
4	30 (12.24 %)	19 (7.76 %)	10 (4.08 %)	NA
5	32 (7.96 %)	21 (5.22 %)	26 (6.47 %)	NA
6	49 (10.91 %)	33 (7.35 %)	23 (5.12 %)	NA
7	42 (6.11 %)	53 (7.71 %)	56 (8.15 %)	NA
8	41 (9.72 %)	29 (6.87 %)	34 (8.06 %)	NA
9	22 (7.53 %)	25 (8.56 %)	13 (4.45 %)	NA
10	30 (7.85 %)	43 (11.26 %)	33 (8.64 %)	NA
adult	29 (6.25 %)	32 (6.9 %)	31 (6.68 %)	20 (4.31 %)
Average	8.57 %	7.70 %	6.46 %	NA

Table 2.5: Token counts and percentage of each vowel removed, by age, using three MADs from median criterion

Age	[a]	[e]	[i]	[o]	[u]
4	20 (4.85 %)	13 (10.16 %)	13 (5.51 %)	0 (0 %)	13 (7.93 %)
5	26 (4.09 %)	13 (7.07 %)	22 (4.74 %)	4 (3.23 %)	14 (7 %)
6	35 (5.79 %)	11 (5.5 %)	26 (4.71 %)	19 (8.96 %)	14 (6.14 %)
7	45 (5.04 %)	24 (6.9 %)	40 (5.13 %)	22 (6.63 %)	20 (5.05 %)
8	38 (6.93 %)	14 (6.48 %)	20 (4 %)	14 (7 %)	18 (8.04 %)
9	24 (6.59 %)	7 (6.73 %)	15 (4.21 %)	9 (6.08 %)	5 (2.55 %)
10	35 (8.33 %)	18 (9.18 %)	18 (4.09 %)	14 (5.93 %)	21 (8.9 %)
adult	35 (5.87 %)	16 (6.9 %)	24 (4.69 %)	20 (7.14 %)	17 (7.2 %)
Average	5.94 %	7.36 %	4.64 %	5.62 %	6.60 %

One final cleaning procedure was conducted before continuing with the formant analysis. After removing measurements that fell out of the pre-determined range, I additionally

[10]No measurements from the 5;0 or adult group required removal.

inspected each speaker's vowel space for extreme outliers. After identifying the source of the outlier, I returned to the spectrogram to compare the automated formant measurement (median) and the actual formant measurement. When the automated measurement and hand measurement differed by more than approximately 400Hz (F1) or 800Hz (F2), the token was removed from analysis. In some cases one or other of the formants was also not visible in the spectrogram and so the token was also removed. This resulted in the removal of 35 additional tokens (See Table 2.6 for the distribution by age and phone.) On the basis of the triple formant tracker, removal of measurements outside of three MADs, and the visual comparison with the spectrogram, I confidently proceeded with the clean formant measurements.

Table 2.6: Hand-removed tokens by vowel and age group

Age	Phone	n
4	u	3
6	o	1
6	u	6
7	o	2
7	u	10
8	o	2
8	u	3
9	o	2
9	u	5
10	u	1

Following the formant cleaning procedure, formant frequency measurements were normalized between speakers via the two scaling techniques outlined in section 2.2 - Lobanov (Lobanov 1971) and ΔF (Johnson and Sjerps 2018) - using an additional custom Python notebook, also included in this project's Github repository (`https://github.com/megseekosh/vocal_tract_vowel`). The results of these calculations are presented in the results.

2.4 Results

The primary research question in this study asks if children's vowel variation decreases over the course of development in a language with a three-vowel contrast. The results for the first experiment begin with descriptive statistics of the participants' formant patterns by age, prior to speaker normalization, for the three phonemic vowels /a, i, u/. Then, the formant measures for these phonemic vowels are normalized using the ΔF scaling techniques, described in the background literature. Using these normalized data, a series of models are fit to predict the degree of participants' vowel dispersion and determine if vowel variability decreases over the course of development, as previous work would predict (Eguchi and Hirsh 1969; Lee et al. 1999; Ménard et al. 2007). Then, the second half of the results section is devoted to comparing the uniform (ΔF) and non-uniform (Lobanov) formant frequency scaling techniques to determine if non-uniform scaling techniques are necessary to normalize between children of different ages and between children and adults.

All analyses were conducted in the RStudio computing environment (version: 1.2.5033; RStudio Team 2020). Data visualizations were created with `ggplot2` (Wickham 2016). Modeling was conducted using the `glmmTMB` (Brooks et al. 2017) package and model summaries were presented with `papaja` (Aust and Barth 2018). The significance of potential model parameters was determined using a combination of log-likelihood comparisons between models, AIC estimations, and p-values procured from model summaries. In all models, continuous predictors were mean-centered to facilitate model interpretation.

Vowel category dispersion

Descriptive statistics of unnormalized data

The first objective in this study is to test how intra-subject and intra-age group vowel variability changes over the course of development from age 4;0 to 10;11. Because I intended to compute vowel variability on an individual speaker and group-level basis, a precautionary data cleaning step was taken before proceeding with the analyses. Any speaker vowel categories that had less than four F1 observations or four F2 observations were removed from analysis (e.g. less than four observations of the F1 of [i] from a given speaker). Since children's voices may be more prone to formant tracking errors due to their higher f0, and thus data removal, it was important to ensure that any differences between higher voices and lower voices was not due to a data scarcity or abundance in any particular age group. This cleaning procedure helped to standardize the measurements across ages. The reason for differing amounts of tokens per vowel category between speakers was due to the data cleaning procedures and occasional wind interference in the recording, as explained in the methods section.

The removal of vowel categories with less than four observations resulted in the removal of 48 vowel categories (see Table 2.7 for distribution by age group and vowel). All of the adults had at least four clean F1 and F2 measurements for each of their vowel categories so no adult data were removed.

To further ensure that accurate comparisons were being made between age groups - since younger children might be more likely to have data removed for tracking reasons than the adults - a random subset of 10 observations for those speaker vowel categories with more than 10 observations were also selected. In this way, no individual speaker contributed more than 10 or less than 4 data points for a given vowel. Unless noted otherwise, all analyses were conducted on these vowel categories that contained 4-10 observations.[11]

[11]Note that this cleaning procedure was not conducted on the mid-vowels [e] and [o] due to data scarcity.

Table 2.7: Number of vowel sets removed by age and phone to standardize measurements across age groups. No categories were removed from the adult speakers.

Age	a	i	u
4	NA	1	7
5	NA	NA	4
6	2	1	5
7	4	2	8
8	5	NA	4
9	NA	1	1
10	NA	NA	3

Summary statistics of acoustic vowel measurements by age (in Hz) are presented in Tables 2.8 and 2.9. The F2-F1 space with the three phonemic vowels, by age, is displayed in Figure 2.2. (See appendices for individual vowel plots by participant and a plot containing median formant values for phonemic and allophonic vowel categories.) Here the median and median absolute deviation (MAD) of each formant are reported, instead of the mean and standard deviation (SD), to provide a non-parametric estimate of the variation. The MAD of a distribution is calculated by first computing the median values of the distribution, subtracting this median from each point in the distribution, and finally computing the median of the computed absolute differences. As such, the MAD is relatively less susceptible to the effect of outliers than other measures of dispersion such as SD.

Table 2.8: Median absolute deviation formant measurements in Hertz for children and adults

Age	F1 (MAD)	F2 (MAD)	F3 (MAD)	F4 (MAD)	n
4	551.38 (302)	2080.49 (932)	3959.64 (423)	NA (NA)	205
5	491.6 (187)	2258.95 (1009)	3887.07 (482)	NA (NA)	288
6	483.44 (209)	2152.89 (1114)	3711.44 (415)	NA (NA)	313
7	484.25 (143)	2338.16 (1053)	3811.39 (396)	NA (NA)	488
8	500.69 (217)	2229.63 (1073)	3660.5 (390)	NA (NA)	293
9	471.67 (179)	1916.46 (1094)	3668.35 (556)	NA (NA)	193
10	485.78 (159)	1948.64 (1022)	3501.85 (421)	NA (NA)	303
adult	460.91 (123)	1603.28 (931)	2952.27 (289)	4060.2 (375)	328

Some mid-vowel categories had less than four observations from a given speaker. As a result, while the descriptive statistics report on mid-vowel data, statistical analyses were conducted exclusively on the peripheral vowels.

Table 2.9: Range (min-max) of formant measurements in Hertz for children and adults

Age	F1 range	F2 range	F3 range	F4 range
4	266-1474	774-3782	2938-4659	NA-NA
5	284-1451	797-3767	2343-4652	NA-NA
6	145-1290	783-3767	2607-4528	NA-NA
7	273-1163	794-3734	2501-4606	NA-NA
8	245-1216	808-3635	2526-4422	NA-NA
9	248-1219	847-3672	2237-4597	NA-NA
10	249-1076	796-3541	2630-4447	NA-NA
adult	262-1091	701-2929	2231-3668	2832-4726

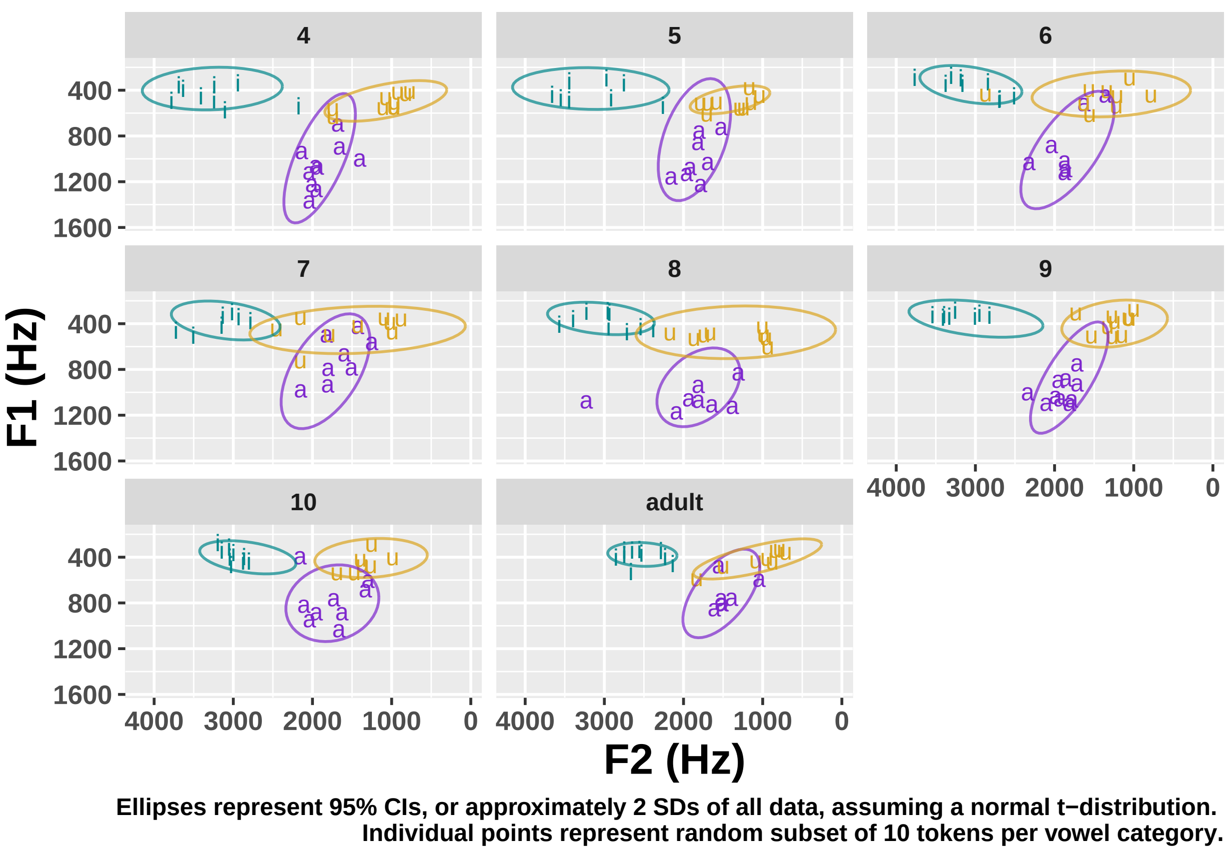


Figure 2.2: Vowel category development by age, in Hertz: adults and children

Table 2.8 demonstrates that overall, as anticipated, the median F1, F2, and F3 values, in Hertz, decrease with age as the vocal tract lengthens. Within the children, the median F2 appears to increase slightly in the 7;0 and 8;0 groups, likely due to the concentration of [u] at higher F2 frequencies. The adult women still exhibit a much lower median F2 than the any of the child age groups. Also as anticipated, the median formant value increases from F1 to F3/F4 across all participants (again F4 was not tracked for the children).

Between-speaker variability, quantified as the MAD, also decreases with age from roughly 300Hz for F1 in the 4;0 group to less than 130 Hz (for F1) in the adults. Notably, the variability even decreases between the older children: for example, for F1-F3, the MAD decreases between the 9;0 group and the 10;0 group, and again decreases between the 10;0 group and the adults. This pattern of variability reduction with age was not always apparent in the higher formants, however. The higher variability in F1 could be due to harmonic spacing in the lower frequencies. In fact, all age groups from 6;0-9;0 were more variable than the 4;0 and 5;0 groups along the F2 dimension. This is somewhat surprising since young children typically master jaw control (correlated with F1) earlier than horizontal lingual control (correlated with F2). But again the higher median F2 variability simply could be reflecting a shifted vowel space. For F3, variability does not appear to decrease notably by age until the 10;0 and adult group.

These results appear to confirm previous work on vowel development in English (Lee et al., 1999) and French (Ménard et al., 2007): younger children are more variable than adults. However, I wanted to ensure that the acoustic variability in the children's speech was due to articulatory instability and differences between acoustic-articulatory mappings in children, and not to the higher formant frequency ranges that the children speak in. In other words, in unnormalized data, child speech could simply *appear* to be more variable because a given amount of acoustic and/or articulatory slop at higher frequencies would result in less auditory perturbation than the same acoustic slop at lower frequencies.

Descriptive statistics of ΔF-normalized data

To ensure that variability in the children's vowel production was not simply due to the frequency ranges of the children's voices, the vowel data were normalized and the variation of each vowel category was computed. The vowels were normalized with the ΔF formant frequency scaling measure reported in the methods. As the calculation of ΔF requires estimation of vocal tract length, the following section begins with a description of the distribution of vocal tract lengths computed from F1-F3 for the children and F1-F4 for the adults, and the ratio between formants (ΔF) by age in this population. (Vocal tract length and formant ratios were computed on all vowels /a, i, e, o, u/). Then, descriptive statistics of formant measurements resulting from the ΔF normalization are presented and within-category dispersion of the phonemic vowels is again evaluated over the course of development.

Table 2.10 summarizes vocal tract length and the ratio between formant frequencies (ΔF) by age (also see Figure 2.3). Unsurprisingly, the estimated average vocal tract length increases with age from roughly 12 cm in the four-year-olds to between 13 and 15 cm in the ten-year-olds and adults. Due to the lengthening of the vocal folds and vocal tract, the average ratio between formant frequencies (ΔF) also decreases with age as the average

Table 2.10: Average vocal tract length and ratio between formant frequencies in Hertz (DeltaF) by age

Age	Vocal tract length (SD)	DeltaF (SD)
4	11.68 (0.84)	1461.73 (97.34)
5	12.13 (0.71)	1405.64 (82.37)
6	12.37 (0.7)	1378.49 (77.51)
7	12.49 (0.74)	1365.79 (80.52)
8	12.48 (0.62)	1365.18 (64.92)
9	13.03 (1.35)	1315.59 (119.91)
10	13.29 (0.48)	1280.58 (45.92)
adult	14.96 (0.68)	1138.81 (51.75)

formant frequencies lower. The vocal tract lengths computed acoustically here resemble the measurements taken from magnetic resonance images of vocal tract development in North American children (Vorperian et al., 2005). These acoustically-derived vocal tract length measures are slightly longer than those measured from articulatory imaging; acoustically-derived measures overreport vocal tract lengths since the effect of the end of the tube is just outside of the lips for those measures.

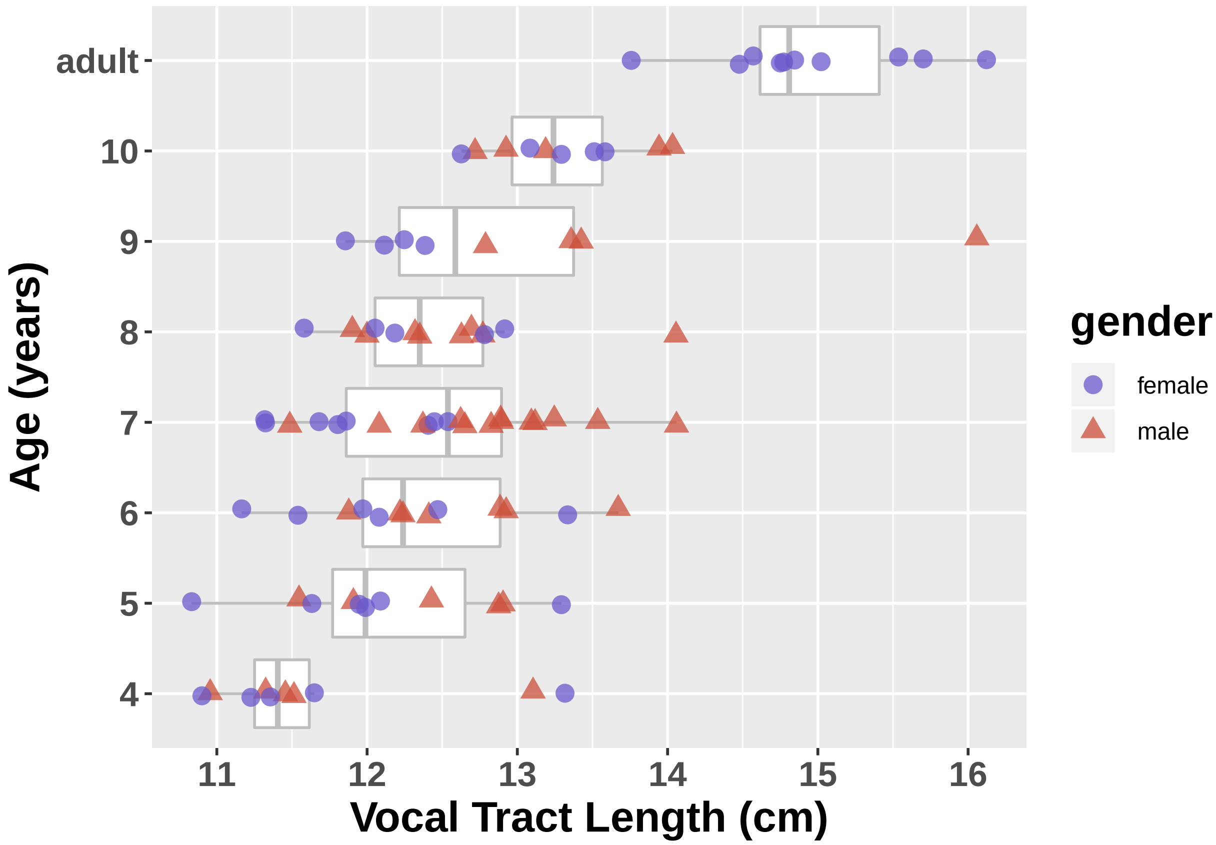


Figure 2.3: Average vocal tract length by age: adults and children

Measuring within-category variability

The comparison of within-category variability across age groups was conducted on the ΔF-normalized formant values. Figure 2.4 plots the median ΔF-normalized formant values. (See Tables 2.19 - 2.21 in the appendices for descriptive statistics of ΔF-normalized formant values.) As the vowel plots demonstrate, speakers tend to have larger within-category dispersion for [a] and [u] than [i]. However, the pattern by age is less identifiable. With speaker-intrinsic information (ratio between formant frequencies on a by-speaker basis) factored out, which is what ΔF does, adults appear to have somewhat tighter, more compact acoustic vowel categories for some vowels, particularly [i], than even the eldest children. Other vowels show little difference by age, or, in the case of [u], do not seem to follow a strict linear pattern of decreased variation with age.

To better ascertain the developmental pattern of vowel variability, the acoustic dispersion of each vowel category was computed across the age groups. To do so, the the average Euclidian distance in F1/F2 space from the vowel mean location was computed, which

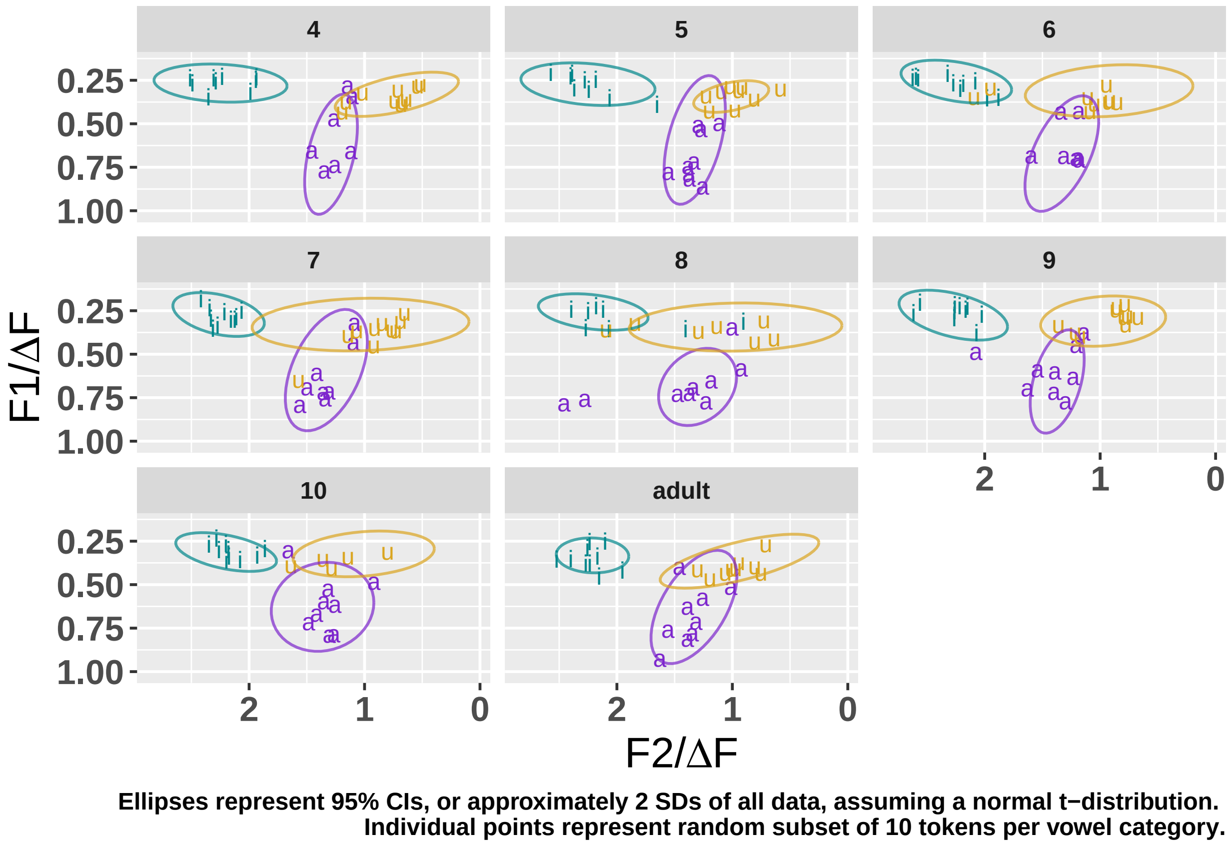


Figure 2.4: DeltaF-normalized vowels by age: adults and children

resulted in a single coefficient per vowel category.[12] This category dispersion coefficient reflects both the *mean* value of each vowel category and its *variability* along the F1 and F2 dimensions. To calculate the category dispersion coefficient, the following steps were taken.

1. First, the mean value of F1 and F2, for each age group's vowel categories, was measured (e.g. the mean F1 and F2 of [a] for the five-year-olds). This step estimated the position of the vowel category in acoustic space.

[12]Previous work has used the Coefficient of Variation (CoV) to measure vowel category dispersion in children (e.g. Lee et al. (1999)). The CoV is the ratio of the standard deviation of the mean to the mean of each phoneme category (Bradlow, 1995; Eguchi & Hirsch, 1965; Lee et al., 1999). One disadvantage of the CoV is that, unlike the category dispersion technique, separate coefficients must be computed for each acoustic dimension so the result is the CoV of F1, CoV of F2, etc.. Nevertheless, to facilitate comparison of this study with previous work, the CoV for each vowel category was also computed (for F1 and F2). In brief, the dispersion results measured via the CoV replicated the dispersion results measured via the category dispersion coefficient: the only reliable differences by age when measuring dispersion via CoV were for the F2 of [i]. Full dispersion results using the CoV, including descriptive statistics and statistical modeling for the CoV for each formant, are included in the appendices.

2. Next, the difference between each individual formant measurement and its vowel category mean was computed and the difference was squared. This was done for F1 and F2. The resulting F1 and F2 coefficients were summed.

3. Finally, the square root of the sum was taken.

These steps were repeated for each vowel production. The mean value of all productions of a given vowel from each age group was then computed (e.g. five-year-olds' [a] productions). This mean value reflected the dispersion of the category in space. The result was a single variability coefficient for each vowel category, for each child age group, that factored in both formants.

The average category dispersion coefficient by phone and age group is listed in Table 2.11 and Figure 2.5. Recall that these measurements were made over vowel categories containing 4-10 observations per speaker.[13] The results demonstrate that the category dispersion coefficient 1) varies by phone and 2) does not always decrease linearly with age. For [a], the category dispersion coefficient does not appear to decrease by age at all - the average category dispersion coefficient for [a] for the adults was 0.18, similar to the values for almost all the other age groups. The category dispersion for [u] is highest for the 7;0 and 8;0 groups, again not showing clear evidence of a linear decrease in variability. The category dispersion for [i] showed more change with age and was smallest in the adults (.13) and highest in the 4;0 and 5;0 groups (0.22 and 0.25, respectively).

Table 2.11: Vowel category dispersion by age group and phone

Age	[a] mean (SD) range	[i]	[u]
4	0.17 (0.11) 0.02 - 0.57	0.22 (0.2) 0.03 - 0.95	0.19 (0.13) 0.03 - 0.44
5	0.19 (0.1) 0.02 - 0.46	0.25 (0.25) 0.03 - 1.36	0.13 (0.08) 0.03 - 0.42
6	0.19 (0.11) 0 - 0.53	0.2 (0.17) 0 - 0.95	0.3 (0.25) 0.05 - 1.08
7	0.2 (0.1) 0.04 - 0.53	0.16 (0.12) 0.02 - 0.67	0.38 (0.28) 0 - 1.38
8	0.19 (0.21) 0.01 - 1.15	0.19 (0.2) 0.01 - 1.28	0.36 (0.22) 0.1 - 1.07
9	0.16 (0.13) 0.02 - 0.71	0.19 (0.15) 0.02 - 0.65	0.22 (0.14) 0.03 - 0.73
10	0.21 (0.15) 0.01 - 0.86	0.17 (0.12) 0.01 - 0.49	0.25 (0.27) 0.03 - 1.51
adult	0.2 (0.12) 0.02 - 0.59	0.13 (0.07) 0.01 - 0.32	0.29 (0.2) 0.03 - 1.1

Note:
4-10 observations per speaker

[13] Category dispersion coefficient measurements computed over *all* data points, including those categories with less than 4 data points, are included in Table 2.22 in the appendices. Overall there were no large differences observed between the category dispersion coefficient values computed over 4-10 tokens per vowel category and those computed over the entire dataset. However, standardizing the number of tokens per category avoids skewing in the event of data loss.

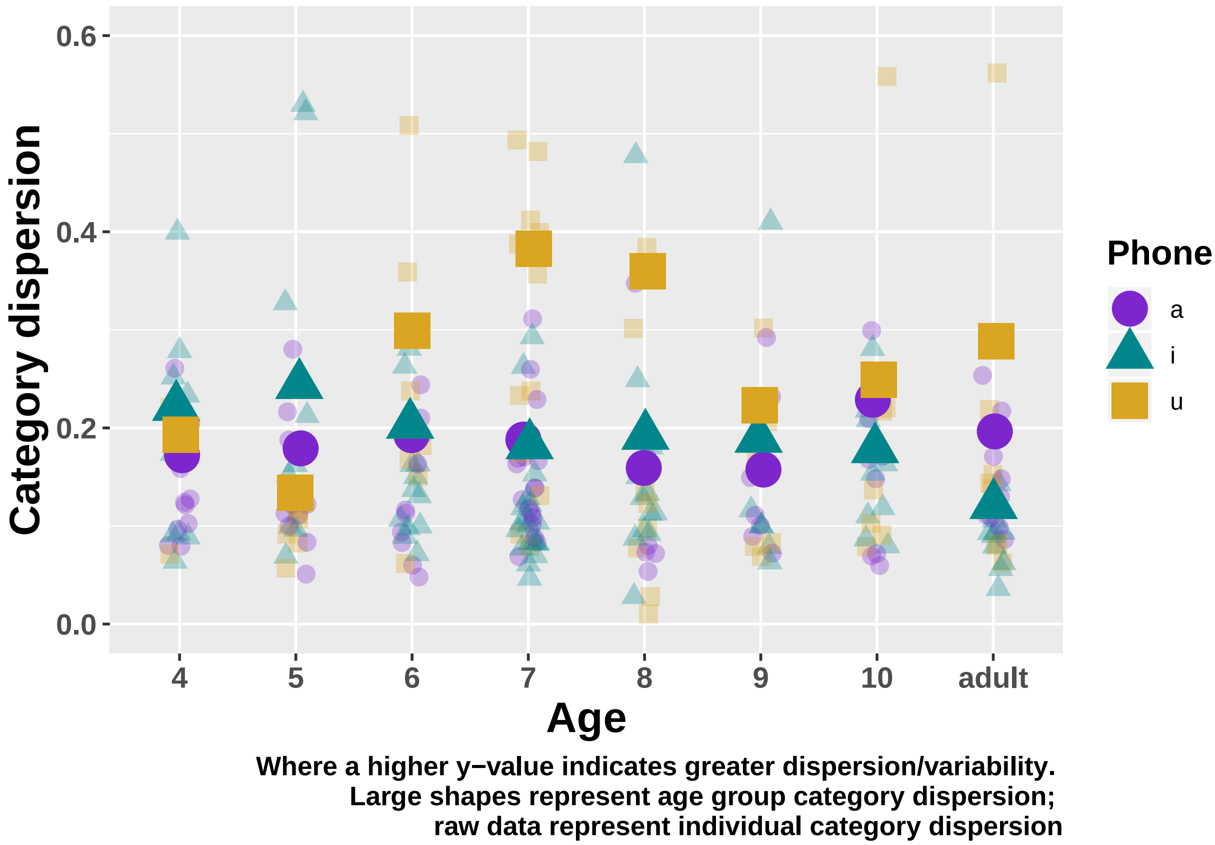


Figure 2.5: Group-level category dispersion by age and phone

The category dispersion of each vowel category was additionally computed on a by-speaker basis. This resulted in a single category dispersion coefficient for each vowel from each speaker. Computing the category dispersion on an individual speaker level generated a datapoint for each speaker's three vowel categories which were used to fit a series of models predicting vowel category dispersion.

Fitting models to predict within-category variability

For the category dispersion model fitting, generalized linear mixed effects models (GLMMs) were fit because the outcome variable - the individual speaker level category dispersion - was necessarily non-negative and left-skewed. Gamma GLMMs were fit using a log linking function to appropriately model the skewed, non-Gaussian distribution of the residual. Three models were fit, one for each phone (/a, i, u/). The model fitting procedure followed the same procedure for all models: first the baseline model, with just the random effect of **Speaker**, was fit. Then, the parameter **Age Group** (4, 5, 6, 7, 8, 9, 10, adult) was added.

The parameter **Age Group** was not significant for the model predicting the category dispersion of [a].[14] For the [u] model, only one child age group, the 7;0 group, differed significantly from the adults: the 7;0 had a significantly larger [u] category than the adults. The parameter **Age Group** did improve upon the model predicting the category dispersion of [i]. The model summaries for each phone model, with the **Age Group** parameter added, are presented in Tables 2.12-2.14.

Table 2.12: Model predicting category dispersion for [a]

term	estimate	S.E.	z.statistic	p.value	95% CI
Intercept	-2.11	0.15	-14.19	0.00	-1.82,-2.4
4	0.07	0.21	0.33	0.74	0.48,-0.34
5	0.14	0.21	0.67	0.50	0.54,-0.26
6	0.01	0.21	0.03	0.97	0.41,-0.4
7	0.12	0.19	0.64	0.52	0.49,-0.25
8	-0.10	0.22	-0.45	0.65	0.34,-0.54
9	0.00	0.22	-0.01	0.99	0.44,-0.44
10	0.19	0.21	0.88	0.38	0.6,-0.23

Table 2.13: Model predicting category dispersion for [i]

term	estimate	S.E.	z.statistic	p.value	95% CI
Intercept	-2.49	0.17	-15.06	0.00	-2.17,-2.82
4	0.62	0.24	2.58	0.01	1.09,0.15
5	0.79	0.23	3.46	0.00	1.24,0.34
6	0.45	0.22	2.03	0.04	0.89,0.01
7	0.24	0.20	1.18	0.24	0.64,-0.16
8	0.42	0.22	1.91	0.06	0.85,-0.01
9	0.27	0.26	1.06	0.29	0.78,-0.23
10	0.54	0.23	2.31	0.02	1,0.08

As the beta coefficients in the [i] model summary demonstrates, there were some differences by age group in category dispersion for the [i] phoneme. Specifically, the adults had significantly tighter [i] categories than all children except for the 7;0 and 9;0 groups

[14]The baseline models for [a] and [i] did not converge, but the models with the addition of **Age Group** did. Consequently, potential differences between age groups were determined solely from model summary outputs for those two models.

Table 2.14: Model predicting category dispersion for [u]

term	estimate	S.E.	z.statistic	p.value	95% CI
Intercept	-1.77	0.23	-7.59	0.00	-1.31,-2.23
4	-0.05	0.43	-0.10	0.92	0.81,-0.9
5	-0.45	0.34	-1.33	0.18	0.21,-1.12
6	0.31	0.33	0.94	0.35	0.95,-0.33
7	0.59	0.28	2.10	0.04	1.15,0.04
8	-0.27	0.33	-0.81	0.42	0.38,-0.91
9	-0.12	0.33	-0.38	0.71	0.52,-0.77
10	0.11	0.33	0.32	0.75	0.75,-0.53

(the difference between the adults and 7;0 only approached significance). Furthermore, as the beta coefficients demonstrate, the children's category dispersion coefficient values did not necessarily decrease linearly with age - the 10;0 group, for example, actually had significantly more dispersed categories than the 9;0 group.

Overall, this analysis shows that there are not significant differences in category dispersion between adults and children for [a]. For [u], only the 7;0 group had a significantly larger category than the adults. There are significant differences in category dispersion for [i]: the adults have tighter categories than many of the child age groups. However, the expected trend of reduced dispersion with age is not observed.

Compensation for vocal tract morphology

Having addressed the first objective of this study, the secondary research question asks if uniform formant frequency scaling adequately factors out anatomical differences between children and adults. If a comparison of uniform and non-uniform formant frequency scaling techniques results in similar formant measurements, this may demonstrate that children compensate articulatorily for their vocal tract morphology during vowel production. If, however, a comparison of the two scaling techniques shows large differences between child and adult within-category variability, this suggests that there are additional sources that explain the differences between adults and children, beyond phone identity and anatomical difference.

To evaluate this question, the vowel data were normalized using two formant scaling techniques - one uniform (ΔF) and one non-uniform (Lobanov), as previously described. For this analysis, the allophonic vowels [e, o] are included, in addition to the phonemic vowels /a, i, u/. The difference in category variability between the scaling techniques is then compared: if there is additional, unexplained variability present in the Lobanov-normalized vowels - meaning there is additional variability after scaling the formants uniformly (accounting for

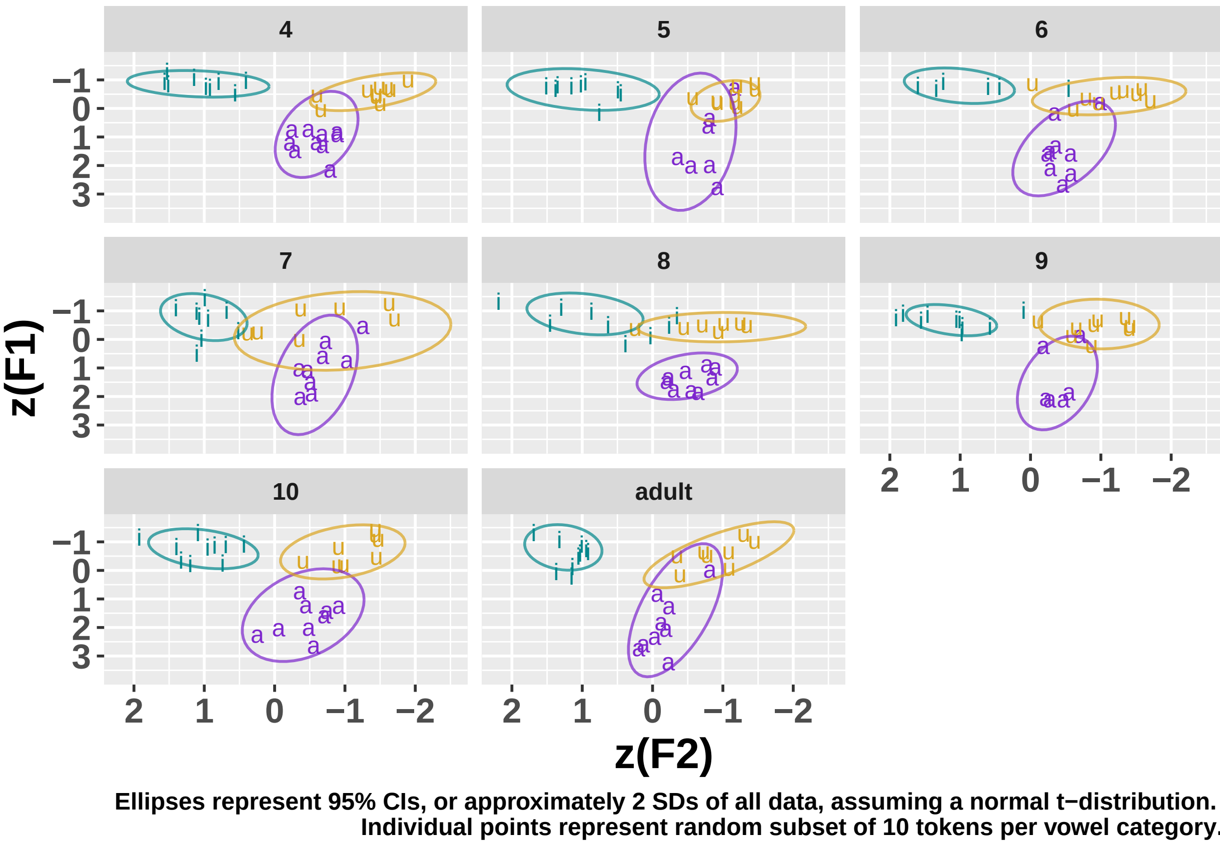


Figure 2.6: Lobanov-normalized vowels by age: adults and children

vocal tract length) and factoring in phone identity - then these results would suggest that another factor, such as articulatory configuration, may differ between the speakers. Such a result would suggest the children may not always compensate for the ratio between their supraglottal cavities during vowel production.

The vowel data were first normalized using the Lobanov scaling technique. Figure 2.6 plots the median Lobanov-normalized formant values (for phonemes) by phone and age group and Tables 2.26-2.28 in the appendices list the descriptive values for the phonemes /a, i, u/. See Tables 2.19-2.21 in the appendices, and Figure 2.4 in the previous section, for the ΔF-normalized formant values.

After scaling the vowels using the two scaling methods, the vowel variability between the children and adults for both sets of normalized vowels (Lobanov and ΔF) was measured. Again, for the Lobanov normalization technique separate normalization factors are calculated for F1 and F2 for each speaker. The ΔF technique, however, uses a single normalization coefficient for each speaker. The variability of each vowel category was measured for each age group (so three categories per age group) by computing the category dispersion coefficient, as described in the previous section. This measurement was made separately to the ΔF-normalized vowels and the Lobanov-normalized vowels.

The category dispersion coefficient, made using the Euclidean distance equation described in the previous section, was again used to calculate the difference in dispersion between adults and children. As before, the mean value of each formant, for each vowel category and age group, was computed. Then, the difference between the mean value for the adults and a given child age group was measured. For example, the mean F1 of [a] was measured over all of the 5;0 children and the mean F1 of [a] over all of the adults and then the difference between these values was taken. As in the previous section, this step estimated the position of the vowel category in acoustic space. The resultant F1 and F2 measurements were both squared, and then summed. Finally, the square root of the sum was taken.

Figure 2.7 displays the difference between the adult and child vowel categories for the two scaling techniques. The left panel of the figure shows the difference between adult and child ΔF-normalized categories, or normalization that incorporates just vocal tract length. The right panel shows the difference between adult and child Lobanov-normalized categories, which incorporate individual formants into the normalization procedure. The visual clearly demonstrates that there are larger differences between normalized adult and child vowel categories that take into account individual formants than vocal tract alone. In other words, beyond vocal tract length and even phone identity, there is still additional variability between adults and children to be accounted for, as the larger values in the right panel demonstrate. These differences between Lobanov-normalized categories and ΔF-normalized categories will be quantified below. It is important to note that the differences between adult and child vowel categories did nevertheless differ by age and phone. For example, there was a large amount of variation between the adult and 4;0 [e] category, but less between the adult and 10;0 [o] category.

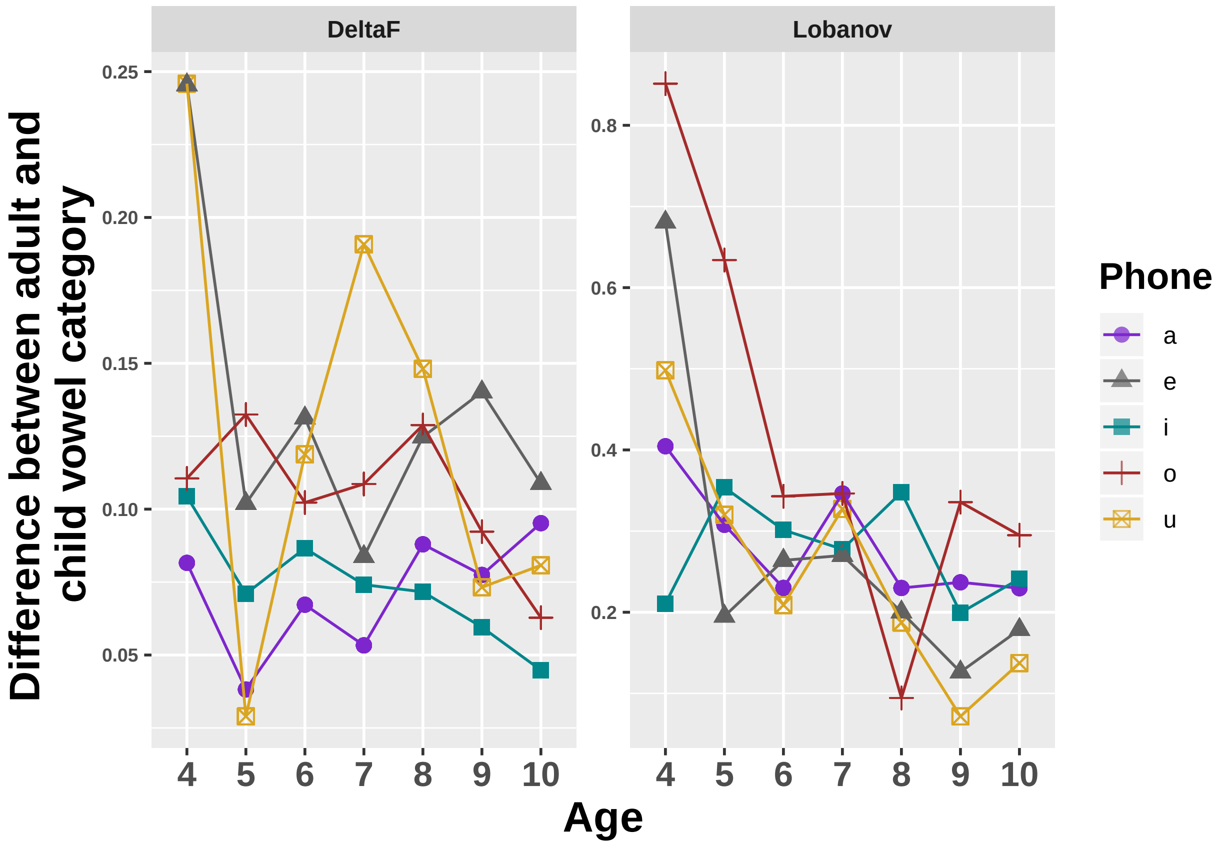


Figure 2.7: Difference between adult and child vowel categories: category dispersion coefficient method. Note the free scales on the y-axis.

Thus far, though a different methodology was employed, these results corroborate Turner et al. (2009): there is additional, unexplained variability in formant frequencies even after removing the effects of phone identity and vocal tract length. However, Turner et al. (2009) concluded that this outstanding variability was due to measurement error in formant tracking in spectrograms and the use of LPC. The authors concluded that any remaining variability beyond that was likely to be statistically meaningless.

On the basis of this observed difference between the two scaling techniques, I hypothesized that the increased variability in the children's productions *could* be due to their articulatory configurations, and not simple measurement error. Specifically, if some children were not compensating for the ratios between their front and back cavities by adjusting their lingual positioning to approximate adult-like formant frequency ratios between cavities, then we might see this exact difference between the uniform and non-uniform scaling techniques. However, this preliminary conclusion requires additional evidence.

Several steps were taken to evaluate the idea that the outstanding variability might reflect children's lack of compensation for their vocal tract morphologies.

1. The children were classified into two groups on the basis of their ratios between back-affiliated and front-affiliated formants. This was done using a technique that did not require us to make assumptions about which formants originated in which cavity (to be explained in detail in the following section). This split was designed to classify the children into more "adult-like" articulators (those children who have either approximated adult-like cavity ratios or who adjust their lingual articulations to approximate adult-like acoustics). The rest of the children were classified into more "child-like" articulators (those children who may not compensate articulatorily as much for their vocal tract morphologies).

2. A series of statistical models were fit to the children's un-normalized formant frequency data. In doing so, it was possible to evaluate in a stepwise manner the effect of adding parameters that are known to influence formant frequency values, such phone identity or vocal tract length, as well as parameters that I hypothesized *might* influence formant values, such as the ratio between front- and back-cavity affiliated formants.

The following two sections outline the results of these analyses.

Classification into adult-like and child-like articulators

The first way that I evaluated if children were compensating for their vocal tract morphology was to divide the children into two groups: adult-like articulators and child-like articulators. This classification was made on the basis of the children's ratios between the first and second formants for [a] and ratios between the second and third formants for [a]. These criteria for the classification were chosen for several reasons. One, if children do not alter their articulatory strategies to compensate for their relatively longer oral cavity, then they would be expected to show different ratios between these formants when compared to adults. Specifically, the longer oral cavity relative to pharyngeal - because the pharyngeal cavity grows disproportionately fast in childhood for boys and girls - results in a heightened F2 compared to an adult model (when assuming adult-like formant-cavity affiliations). Thus, those formants deriving from the back cavity in children would be lower relative to the formants affiliated to a front cavity.

The decision was made to calculate formant frequency *ratios* because doing so does not require knowledge or assumptions of formant-cavity affiliations. Formant-cavity affiliations may be problematic assumptions to draw acoustically because they can become highly unpredictable in the event of some articulatory modifications (e.g. undershoot). Formant-cavity affiliations can also potentially vary due to idiosyncratic vocal tract characteristics (e.g. a heightened palate). Formant ratios, such as the measures proposed here, merely reflect the relationship between the cavity sizes, independent of specific affiliations, allowing us to skirt entirely the issue of affiliation (Apostol et al., 2004).

Finally, F1 to F2 and F2 to F3 ratios were chosen for [a] in particular because it has been predicted that the ratio between F1 and F2 for [a] is larger in children than adults and that the ratio between F2 and F3 for [a] is smaller in children than adults (Ménard et al. 2007).

I additionally opted to calculate formant ratios for [a], instead of [i] or [u], as Helmholtz resonators are the source of (some of) the formants of [i] and [u]. Since Helmholtz resonators reflect constriction length and area, in addition to cavity length and volume, these were less ideal options to reflect the ratio between cavity sizes in the children.

Before splitting the children into groups on the basis of these formant ratios, an additional check was made on our assumptions concerning the ratios between back- and front-affiliated formants. To do that, the hypotheses concerning the formant ratios were evaluated on two data sets containing formant measurements from adult speakers. Like the child vocal tract, though to a lesser extent, the female adult vocal tract has a smaller pharyngeal cavity relative to oral. Consequently, to validate the formant ratio approach, the mean F1, F2, and F3 was calculated for [a] over all of the adult speakers in the classic Petersen & Barney (1952) vowel dataset and also the Hillenbrand et al. (1995) vowel dataset using the phonTools R library (Barreda, 2015). The Petersen & Barney dataset contained n=28 adult women and n=33 adult men and the Hillenbrand et al. dataset contained n=48 adult women and n=45 adult men. Overall, the relationship between F1 and F2 for [a] was confirmed: in the Petersen & Barney dataset, women had a higher average F1:F2 ratio (female mean=0.70, sd=0.06) than men (male mean=0.66, sd=0.04). This was also the case for F1:F2 in the Hillenbrand dataset: F1:F2 for women (mean=0.61, sd=0.07) and F1:F2 for men (mean=0.58, sd=0.06).

The ratio between F2 and F3 in the datasets was less straightforward. As anticipated, women had a smaller F2:F3 ratio in the Petersen & Barney dataset, though this difference was slight (female: mean=0.44, sd=0.05, male: mean=0.45, sd=0.06). However, women had a larger average F2:F3 ratio in the Hillenbrand data (female: mean=0.54, sd=0.05, male: mean=0.52, sd=0.06). As a result of this, although both F1:F2 and F2:F3 are factored into the calculation of the median split by children, only the F1:F2 ratio is used for the statistical modeling procedure presented in the following section because that ratio appears to be the most reliable indicator of cavity size.

Having confirmed that the formant ratios do reflect differences in cavity size, I turned to the calculations of formant ratios for the children in the current study. To do this, the mean F1, F2, and F3 was calculated for [a] for each child. (Participant c64 only had one valid [a] token, so formant ratios for that participant were not computed and c64 is not included in the following analyses.) Then, the median F1:F2 ratio and F2:F3 ratio was calculated for [a] over all the children. Those children who had both a higher F1:F2 ratio *and* lower F2:F3 ratio for [a] were classified as "child-like" articulators. Children who did not satisfy both criteria were classified as "adult-like" articulators. So a child who only had a higher F1:F2 ratio or only a lower F2:F3 ratio was not classified as child-like. This method allowed us to maximally conservative in the classification of the children.

Table 2.15 displays the number of children by age group who were classified as adult-like articulators and child-like articulators. As anticipated, in the older age groups 9;0 and 10;0, there are fewer children who were classified as child-like articulators. Still, the fact that several children were classified as child-like articulators, even a few in the eldest age groups, demonstrates that there may be great variability in when children learn to approximate adult-like acoustic patterning and adult-like cavity sizes. This makes sense as children grow at different rates, have growth spurts at different times, and have different experience practicing language. These factors are further considered in the discussion section.

Table 2.15: Number of children classified as child-like and adult-like articulators, by age

Age	adult-like articulator	child-like articulator	total
4	6	4	10
5	10	1	11
6	6	7	13
7	16	4	20
8	8	5	13
9	8	NA	8
10	8	2	10

Figures 2.8 and 2.9 display the results of the comparison by median split. On the left, the adult-like articulators show a linear decrease in difference between Lobanov and ΔF vowel category size. This linear decrease with age is anticipated because as the adult-like children age, there should be fewer and fewer differences between adults and children outside of vocal tract length. This is presumably because as children age, they are more likely to have mastered vocal tract morphology compensation. (A linear decrease is not exactly present for [i].)

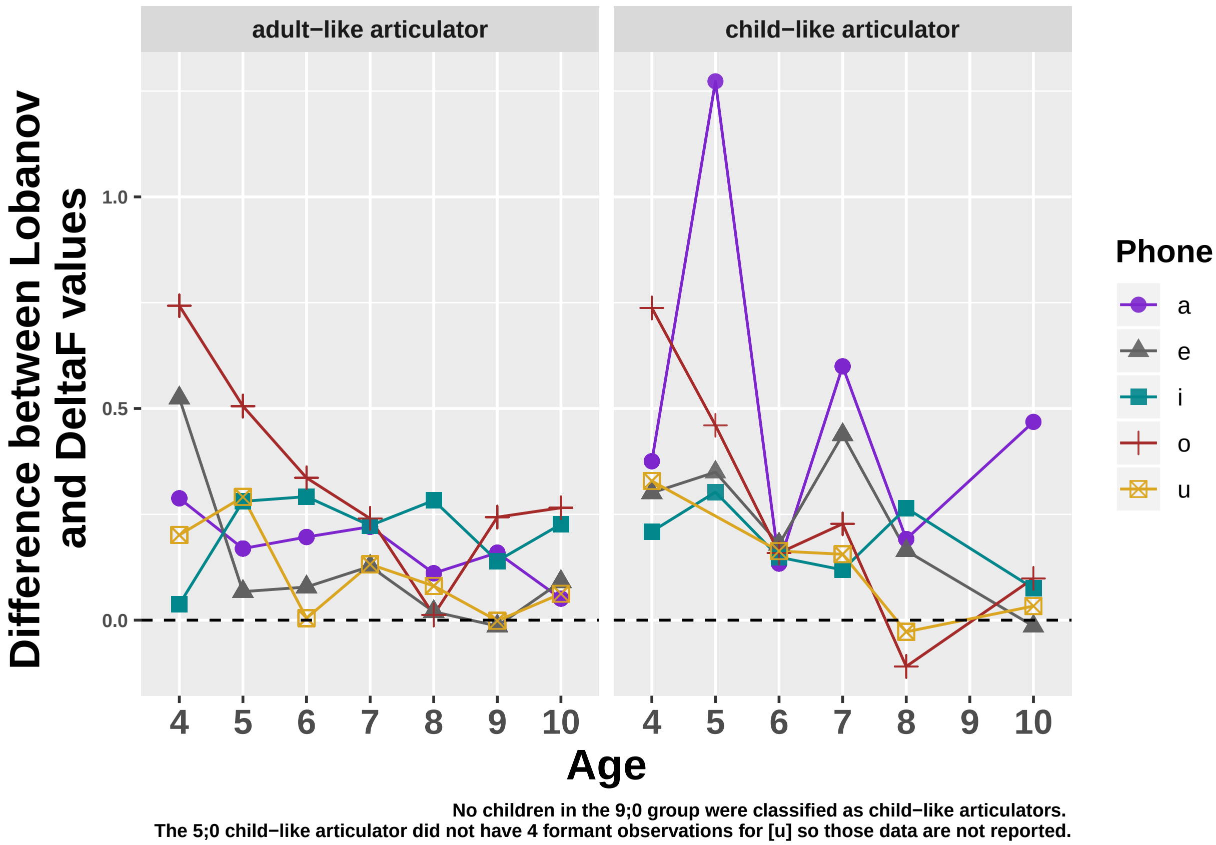


Figure 2.8: Difference between adult and child vowel categories by scaling method and articulatory status

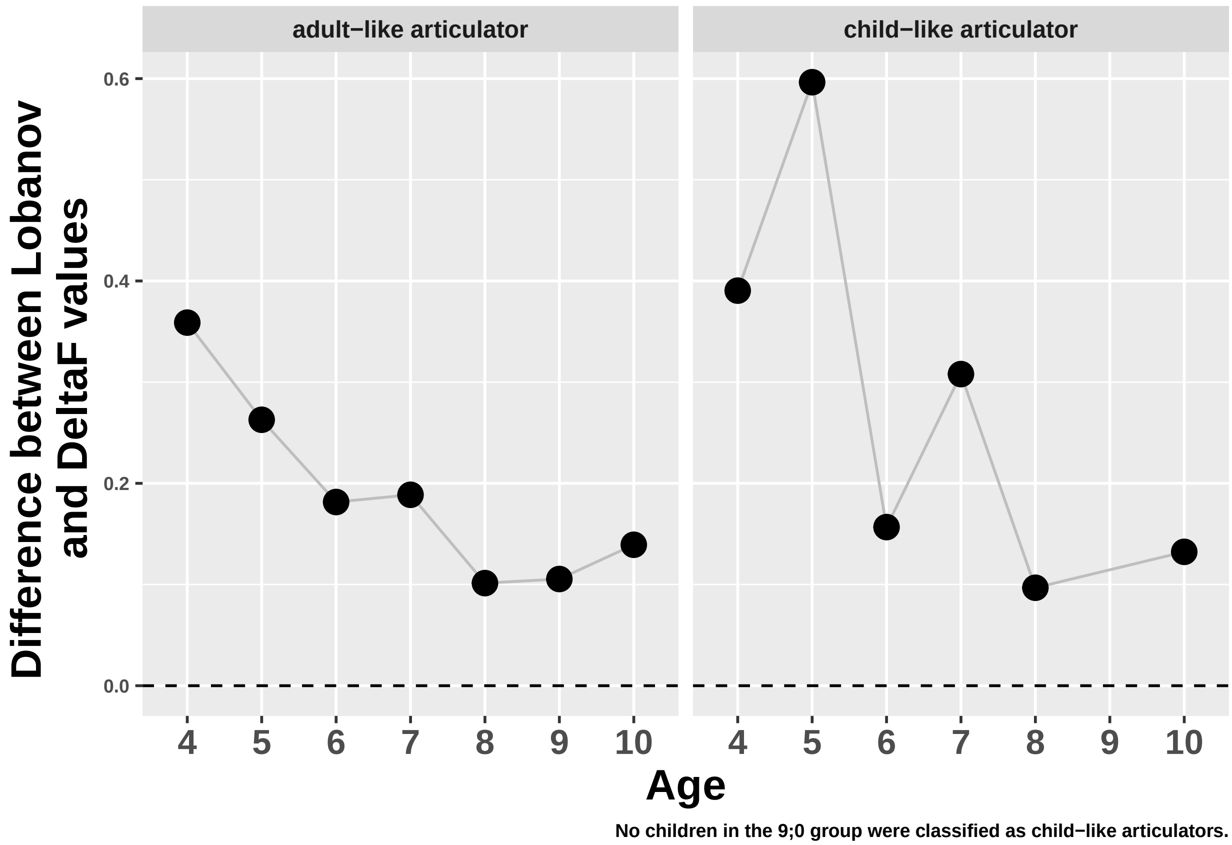


Figure 2.9: Average difference between adult and child vowel categories by scaling method and articulatory status.

Fitting models to un-normalized formant frequency data

As a final step to determine what accounts for the outstanding variation in the formant data, I built statistical models to predict the values of unnormalized formant frequency data from the child speakers. The goal of this modeling was to determine the predictors of formant frequencies in the children. The constructed models permitted control of factors known to influence formant frequencies, such as phone identity and vocal tract length, so that I could determine if the addition of cavity size (ratio of F1:F2 for [a]) explained any remaining formant frequency variance. Thus the dependent variable in the modeling was *un-normalized* formant frequency data (separate models for F1 and F2; details below) because the intent was to factor out anatomical differences via the addition of a **Vocal Tract Length** parameter in the models. Additionally, for this modeling, only the peripheral, phonemic vowels /a, i, u/ were incorporated because there were not sufficient data for the allophonic vowels to include them in the modeling.

Two generalized linear mixed effects models (GLMMs) were fit to predict the children's un-normalized formant frequency data. One model was fit to predict the F1 measures and another model was fit to predict the F2 measures. The choice to fit GLMMs was made due to the non-negative and right-skewed nature of the F1 and F2 outcome variables. Continuous variables were mean-centered and standardized to facilitate model interpretation. Both models were fit according to the following procedure: first, a baseline model with only the random intercepts for individual **Speaker** was fit. Then, model parameters were added in the following order: **Phone** (/a, i, u/), **Vocal Tract Length**, **F1:F2 ratio for [a]**, and **Gender**. Once again the models were only fit to predict formant variability within the child speakers.

For both the F1 and F2 models, **Phone**, **Vocal Tract Length**, and **F1:F2 ratio for [a]** improved upon baseline model fits. **Gender** did not improve upon the F1 or F2 models containing **Phone**, **Vocal Tract Length**, and **F1:F2 ratio for [a]**. The final model summary for F1 is listed in Table 2.16 and the model summary for F2 is in Table 2.17. This result for the models suggests that there is additional variance in the formant frequency measures beyond phone identity and vocal tract length - something that Turner et al. (2009) likewise concluded However, the addition of the **F1:F2 ratio for [a]** parameter in a model already controlling for extant variability (i.e. in the error term) suggests that the outstanding variance in the formant frequency measures can be accounted for with the size of the children's cavities.

Table 2.16: Model predicting F1 frequencies

term	estimate	S.E.	z.statistic	p.value	95% CI
Intercept	861.63	5.71	150.78	0.00	872.83,850.43
Phone:[i]	-479.88	7.76	-61.83	0.00	-464.67,-495.1
Phone:[u]	-403.58	9.87	-40.88	0.00	-384.23,-422.93
Vocal Tract Length	-47.40	5.16	-9.18	0.00	-37.28,-57.52
F1:F2 ratio for [a]	579.18	60.26	9.61	0.00	697.29,461.07

Overall, these results confirm previous research that vocal tract length and phone identity explain large amounts of between-speaker acoustic variation. However, via a step-by-step elimination of these known variables, and statistical model building, I have demonstrated that some school-aged children do not yet approximate adult-like articulatory configurations, as evidenced in the ratios of front and back cavity-affiliated formants. This lack of compensation for vocal tract morphology, which operates *somewhat* independently of vocal tract length, may explain some of the variability known to characterize child speech.

Table 2.17: Model predicting F2 frequencies

term	estimate	S.E.	z.statistic	p.value	95% CI
Intercept	1,818.95	13.32	136.57	0.00	1845.06,1792.85
Phone:[i]	1,220.24	16.17	75.47	0.00	1251.93,1188.55
Phone:[u]	-400.71	21.21	-18.89	0.00	-359.14,-442.28
Vocal Tract Length	-174.68	12.94	-13.50	0.00	-149.32,-200.05
F1:F2 ratio for [a]	-537.91	150.20	-3.58	0.00	-243.52,-832.29

2.5 Discussion

This chapter had two objectives. The first objective was to determine if the development of within-category vowel dispersion followed a different trajectory in children learning a language with three phonemic contrasts. The second objective was to determine if uniform formant frequency scaling techniques adequately factor out between-speaker speech variability in a large cohort of children.

Within-category vowel dispersion

Previous work evaluating the development of within-category vowel dispersion in English (Lee et al. 1999) and French (Ménard et al. 2007) has concluded that within-category vowel variability decreases with age, likely due to mastery of fine motor control (Eguchi and Hirsh 1969). However, according to approaches such as Adaptive Dispersion Theory, speakers may modify their acoustic targets in accordance with vowel inventory size (Recasens and Espinosa 2009b; cf. Bradlow 1995). Specifically, speakers of languages with relatively large vowel inventories, such as English and French, may have tighter, more compact acoustic vowel categories to avoid category overlap and compensate for a more crowded acoustic space. Accordingly, speakers of languages with less vowel contrasts, such as Quechua, may have less constraints upon their within-category dispersion due to the reduced functional need to accommodate multiple categories. As a result, child Quechua speakers may approximate adult-like within-category variability at a younger age and may not demonstrate the supposed universal linear decrease in acoustic vowel category size that accompanies aging in childhood.

The results on this topic showed some developmental changes in within-category vowel dispersion, but only the differences by age for the F2 of [i] reliably reached statistical significance. As evidenced from the individual speaker-level CoVs, the adult CoVs tended to be smaller, on average, than the children's CoVs. The vowel plots by age presented in the results section (Figure 2.10) also demonstrate that there is a reduction in variation between the children (as a whole) and the adults. The adult vowel categories for [i] and [a] in particular are more compact than any of the children's [i] and [a] categories. However, these trends were not significant.

Furthermore, I did not necessarily see a linear decrease in category dispersion over the course of development, a finding that runs counter to most previous work on this topic. Instead, the group-level CoVs either remained constant throughout development (in the case of F2 for [a]) or fluctuated non-linearly (e.g. for F2 of [u]).

A decrease in within-category vowel dispersion has been reported by previous work using robust samples (Lee et al. 1999; Vorperian and Kent 2007) on different languages (Lee et al. 1999; Ménard et al. 2007; Pettinato et al. 2016). Thus, this chapter does not argue that a general trend towards reduced within-category dispersion is an unfounded conclusion. Instead, this work suggests that the spoken production of vowels may progress differently based on the language of exposure. This is a relatively uncontroversial conclusion, but it does suggest that there may be exceptions to ubiquitous child speech variation - and that the topic of child speech variation almost certainly warrants further cross-linguistic examination. The Quechua data analyzed here suggest that children may hone in on their acoustic categories as they progress to adulthood, and that they may have mastered more of this variation at a younger age. This is potentially due to the phonological inventory of their language.

The analysis for the study of within-category dispersion in this chapter also presents an opportunity to discuss how vowel variability appears to be contingent upon vowel normalization technique. Consider the vowel plots by age for the unnormalized data (Figure 2.2) and ΔF-normalized data (Figure 2.10) in the results section. The difference between the children and adults is more apparent in the unnormalized data and less apparent in the vocal-tract length normalized data (ΔF-normalized).

This observation suggests that *some* of the supposedly inherent variation in children's speech may not simply be the result of children's unsteady articulatory executions. Instead, some variability could be the result of the high frequencies of children's speech (See also Chen et al. (2019) for findings concerning the increased difficulty of formant tracking in higher frequency ranges which may lead to erroneous reports of heightened variability). The logarithmic nature of speech frequencies means that speech produced with a higher f0 could permit more variability relative to speech produced with a lower f0. This effect of f0 may occur because articulatory perturbations at higher frequencies result in less auditory disruption relative to the same perturbation at a lower frequency.

Analyses of adult speech provide support for this idea that children's speech variation may be inherently tied to their higher f0s. Di Benedetto (1994) compared the formant frequencies of vowels produced by speakers with different f0s (two men, one woman). In a series of perceptual tasks and acoustic analyses, the author found that vowel identification was tied to f0, but in idiosyncratic ways that depended on vowel height (e.g. [ɪ] and [ɛ]). For example, in a vowel identification task, it was found that simulated f0 (125Hz versus 185Hz) affected the perception of vowel height in low vowels, but not high. Acoustic analyses corroborated these results. Despite the higher f0 of the female speaker in the study, the formant frequencies for the high vowel [ɪ] were nearly identical between the female and male speakers. But for the lower vowel [ae], the female speaker's F1 frequencies were higher relative to the male F1 frequencies. In other words, this study suggested that a higher f0 corresponds to a higher F1 *even after standard normalization*.[15]

The results of Di Benedetto (1994) are relevant for a discussion on how f0 affects vowel variability. Given the results of Di Benedetto, one can anticipate that accurate vowel height (F1) identification for low vowels will be more contingent upon f0 than identification of high vowels. Likewise, for the purposes of this paper, one might anticipate that vowel identification in the realm of children's higher formants could be more contingent on f0 relative to adults' lower formants. In practice, this would mean that the higher the f0, the more reliant the formant will be upon f0 to ensure accurate vowel identification. If we envision children employing a form of auditory feedback to fine-tune their acoustic output, it becomes more understandable why children could have more variable vowel categories.

As Di Benedetto demonstrated, at higher frequencies, listeners have to incorporate both F1 and f0 to adequately identify vowel height. Do children employ both of these acoustic cues during feedback and as they fine-tune their speech categories? Adults may not even employ both cues during perception. Consider that the low vowel [a] consistently exhibits more within-category variation relative to higher vowels in adult speech.

Whether children successfully employ both f0 and F1 during auditory feedback is an open question. If they do not, it could explain a heightened variability that, crucially, does not stem exclusively from lack of practice or immature motor control, but instead from an inability to simultaneously incorporate two acoustic cues during the construction of acoustic categories. As children age, and their f0s lower, reliance on f0 decreases (though it never disappears, as Di Benedetto demonstrated), potentially explaining some decrease in variation correlated with age.

For these reasons, as mentioned in the results section, it is imperative to measure formant variability on normalized data. (But as the following section highlights, even standard normalization procedures may not be sufficient if children do not compensate articulatorily.) Note, however, that Lee et al. (1999) and Eguchi and Hirsh (1969) - the studies on children's vowel variation that are most frequently cited - used unnormalized data to make conclusions concerning within-category dispersion over the course of development. Going forward, measurements of formant frequency variability should be made on normalized data, even when comparing just between children, but almost certainly when comparing between children and adults.

Uniform versus non-uniform formant frequency scaling

The second objective of this study was to evaluate if uniform formant scaling techniques adequately factor out between-speaker differences in a cohort of children. Not only are children's oral cavities larger relative to their pharyngeal cavities throughout development, but the vocal tract also grows non-uniformly, making it difficult for children to establish acoustic-

[15] Note that the fact that f0 becomes increasingly less relevant in the lower frequencies likewise entails a form of non-uniform normalization (Fant 1975).

articulatory mappings (Vorperian et al. 2005) (see Section 2.2 for more detail). Thus, if a uniform scaling technique (ΔF) performed as well as a non-uniform scaling technique (Lobanov) for vowel normalization, this result would suggest that children appropriately compensate for their vocal tract morphologies by modifying the constriction location during their production of vowels. Though the question of children's vocal tract morphology compensation has previously been evaluated (Ménard et al., 2007; Turner et al., 2009), the comparison of a uniform and non-uniform formant scaling techniques, and subsequent analysis and modeling, was a novel methodological approach to the question.

Overall, the results of the formant scaling procedures demonstrated that there was additional between-speaker formant frequency variability to be accounted for, after factoring out the known effects of vocal tract length and phone identity. The potential source of this variability was further examined by splitting the child participants into two groups: children who approximate more adult-like formant frequency ratios for [a] and children who maintain more child-like frequency ratios for [a] (those that would be expected given a larger oral cavity relative to pharyngeal). This analysis demonstrated that some children show evidence of adult-like formant frequency ratios derived from the oral and pharyngeal cavities, but others do not. In further support of this conclusion, was the fact that I found a progressively smaller percentage of children within each age group to be child-like in their formant frequency ratios. This is to be expected as more and more children should learn to compensate for their cavity ratios as they age and growth slows.

Additional support for the idea that some children are not compensating for their vocal tract morphology came from within the adult-like articulators (those children who approximated adult-like formant frequency ratios for [a]). In those children, the difference between the non-uniform scaling technique and uniform scaling technique categories *decreased* with age, suggesting that as children age, their articulatory configurations become less and and less relevant. This leaves only the primary sources of variability - vocal tract length and phone identity - as the sources of variation between adults and children. The statistical modeling also confirmed this result as the addition of the F1 to F2 ratio for [a] improved upon standard models predicting unnormalized F1 and F2 values with only phone identity and vocal tract length.

The finding that some children do not approximate adult-like formant frequency ratios suggests that there are *some* children who do not compensate for their vocal tract morphologies by modifying lingual placement. However, previous work on this topic has come to an alternative conclusion, and has found that all children appear to adjust their articulatory strategies over the course of development as cavities grow. The source of these different conclusions warrants some exploration. Comparing methodological differences between the current study and previous work is not meant to be a critique of previous work on this topic. Rather, the following discussion attempts to isolate how the current approach differed to explain how previous work came to the conclusion that children reliably compensate for their cavity ratios.

One reason why the current study found evidence that children did not uniformly compensate for their vocal tract morphologies could be due to the methodologies of previous work. Ménard et al. (2007) synthesized formant data to reflect a four-year-old child who was not compensating for their vocal tract (aiming for adult-like articulatory targets) and a child who was compensating for their vocal tract (aiming for adult-like acoustic targets). The authors then compared these synthesized productions to productions from actual four-year-old children. The data from the real children clearly matched the data synthesized to reflect a child who was aiming for adult-like acoustic targets by shifting the tongue forward in the oral cavity.

Turner et al. (2009) likewise evaluated the question of outstanding formant variability for the estimation of vocal tract length by constructing a mathematical model to predict wavelength (F1-F3) variability that controlled for the known parameters of phone identity and vocal tract length. The authors did conclude that even after including parameters for phone identity and vocal tract length, there was outstanding variability within their model. However, this was attributed to measurement error inherent to formant tracking in spectrograms and the use of LPC. Extant variability, beyond this measurement error, was statistically meaningless, suggesting that speakers of all ages modify their articulatory strategies. This model was verified both on the Peterson and Barney (1952) dataset as well as the much larger (n=436 participants) Lee et al. (1999) dataset.

The results of this study are not entirely at odds with Turner et al. (2009) and Ménard et al. (2007)'s conclusions, however. First, this work finds evidence that some children successfully modify their lingual placement to compensate for their longer oral cavities. The current conclusion concerns individual differences in the degree of compensation. This work finds that some children may simply be better compensators than others and that this development is not necessarily linear (i.e. predictable by vocal tract length). Consider that Ménard et al. (2007) measured the real vowel productions of 5 children aged 4;0 - perhaps those authors sampled five 4;0 children who were, on average, fairly successful compensators. A different sample of 4;0 children may have been less adept compensators. Though the current study does not have evidence of this, it seems plausible that factors such as a recent growth spurt, in addition to a child's age, could predict how well children compensate and explain some of this variability. Thus, overall, older children will be better at compensation but additional developmental factors almost certainly play a role which may explain why even some children aged 10;0 appear not to approximate adult-like ratios.

Of course sampling cannot entirely explain differences between this study's conclusions and previous work. Turner et al. (2009), for example, tested their model assumptions on the 15 children from Peterson and Barney (1952) but also the much larger sample of 436 participants from Lee et al. (1999). One difference between this study and Turner et al. (2009) could have to do with some modeling assumptions made during the analyses. Though the mathematical model proposed in Turner et al. (2009) is carefully laid out, the authors nevertheless do make an explicit assumption concerning formant-cavity affiliations (p.2380). However, as explained in the results, affiliations can be variable and are contingent upon several poorly-understood factors, many of which are outside of researcher control (e.g. degree of articulatory constriction or palate height). One advantage of the measure of morphological compensation proposed here is that it avoids specifying formant-cavity affiliations, instead focusing on the overall ratio of the formants. The results of Turner et al. (2009) rested upon cavity affiliation assumptions modeled on an adult male vocal tract (Fant 1975). Yet in a study of speech development, these assumptions may not be met, thereby jeopardizing the model results (see Denny and McGowan (2012a) for similar concerns regarding adult-based assumptions for modeling speech).

Formant-cavity affiliations, which are dependent upon cavity size, may, logically, shift over the course of development as the ratio between the cavities' lengths shifts (as suggested in Vorperian et al. (2011); see also analyses in Martland et al. (1996)). Some light has been shed on the topic of developmental changes in cavity affiliation. For example, Vorperian and colleagues have been conducting the important work of mapping age- and gender-specific cavity size changes throughout development (Story et al. 2018; Vorperian et al. 2005; Vorperian et al. 2011), as we may be able to predict affiliations based on the difference between oral and pharyngeal cavity lengths. Still, further modeling and articulatory imaging work is needed to fully understand the potential changes in cavity affiliation throughout childhood. Until then, a best practice is to avoid assumptions of affiliation and, ideally, skirt the issue entirely in the study of non-adult male vocal tracts.

Another reason behind the different conclusions that were reached in this study may be the methodological approach of comparing formant frequency scaling techniques. Specifically, the first method that I used to evaluate if children were compensating for their vocal tract morphologies was to compare the ΔF and Lobanov scaling techniques. However, these scaling methods - one formant-intrinsic and the other formant-extrinsic - could inadvertently highlight some biases in formant tracking.

The ΔF scaling technique computes a single scaling factor from as many formants as possible to factor out anatomical differences. However, the Lobanov technique employs z-score normalization and as such only employs data from individual formants - F1 is scaled according to F1 data, F2 scaled according to F2 data, etc. Thus, errors in the tracking of a single formant could exert a stronger influence over the Lobanov-scaled data than the ΔF-scaled data. Tracking errors from a single formant in ΔF-scaled data (i.e. F1 consistently tracked as F2 over many vowels) could thus be mitigated by other formants in a way that they could not in Lobanov-scaled data. In such a case, differences between ΔF-scaled data and Lobanov-scaled data would be due not to vocal tract configurations but instead to the effect of tracking errors upon scaling technique results. In such a case, F1 tracking errors *might* be expected to be most severe because the range of possible F1 values is much smaller (200-1200 Hz for adults and children) than the range of possible F2 values (800-3500 Hz for adults and children). What's more, bandwidths, and thus formant peaks, are wider at higher frequencies which could lead to easier tracking of higher formants compared to lower formants with a more narrow bandwidth and ensuing peak. Regardless of the cause of the error, a tracking error in F1 would nevertheless cover a larger proportion of the possible F1 measurements than a similar tracking error for F2 or F3.

These concerns on automated formant tracking aside, in the current study, every available step was taken to ensure that the formant measurements were clean. Furthermore, the conclusion concerning children's lack of consistent vocal tract morphology compensation had numerous sources beyond the comparison of the two scaling techniques. This study also demonstrated that the different ratio between between front- and back-cavity affiliated formants changed with age, suggesting more mature compensation strategies. The statistical modeling demonstrated that the ratio between front- and back-cavity affiliated formants was relevant to factor out formant frequency variability, above and beyond phone identity and vocal tract length. Thus, while the issues of automated formant tracking, even with extensive post-processing and cleaning, are well-known and are worse for child speech (Chen et al. 2019), and the possibility that Lobanov-normalized measurements were more sensitive to tracking errors cannot *definitively* be ruled out, multiple analyses led to the conclusion that children do not necessarily compensate for their vocal tract morphologies. Taking all of this evidence together, the results of the current work remain more conclusive.

Given the individual differences found in the vocal tract compensation component of this work, a clear next step is to evaluate what factors predict compensation in children. As previously mentioned, chronological age and height are expected to be highly predictive of compensation since older children should become better at adjusting for their distinct vocal tract shapes. However, there are numerous other potential predictors. I mentioned recent changes in the child's height (give the linearity of height and vocal tract length), but also linguistic effects such as the frequency and functional load of a given vowel within a language. Here the prediction would be that children may be more prone to compensate during the production of more frequent vowels for which they have a more stable articulatory target. Other linguistic effects may also interact with compensation - effects of lexical frequency or articulatory practice with a given sound sequence. These, in addition to the vocal tract imaging and modeling work already being carried out, are all possible avenues for future research on the topic of vocal tract morphology compensation in childhood.

2.6 Conclusion

This work compared within-category vowel dispersion in adults to a cross-sectional sample of children. This category dispersion was measured in Quechua, a language with a three-vowel /a, i, u/ contrast, to evaluate the effect of phonological inventory on acoustic variability. Overall, even the youngest children appeared to have mastered adult-like levels of acoustic variability, though this pattern varied somewhat by phone. The anticipated linear decrease in within-category dispersion as children aged was not apparent. This result suggests that language of exposure can mitigate some variability that has been reported to be inherent to children's speech production.

The formant frequency patterns evaluated for the first study were additionally used in a study to determine if non-uniform scaling is necessary to factor out between-speaker anatomical differences in children. To evaluate this, this work tested if children compensate for the different ratios of oral to pharyngeal cavities in their vocal tract morphologies during spoken language. The results support an individual difference account of vocal tract compensation: some children were better at compensating than others. Degree of compensation was not entirely predictable by chronological age or vocal tract length. Additional predictors of compensation differences between children, such as recency of a growth spurt, were proposed. Overall, however, a larger percentage of the eldest children (age 10) were compensating for their vocal tract morphologies, and the most relevant factors for formant frequency variability - phone identity and vocal tract length - progressively become the only relevant predictors of formant variation as the children aged.

2.7 Appendices

Table 2.18: Complete list of real word stimuli

Syllabified lexical item*	Translation	Trial type
ˈqha.ri	‘man’	*Training*
ˈwar.mi	‘woman’	*Training*
ˈwa.si	‘house’	*Training*
ˈal.qo	‘dog’	*Test*
ˈa.pi	‘corn/citrus drink’	*Test*
ˈchi.ta	‘sheep’	*Test*
ˈch’u.lu	‘hat’	*Test*
ˈcu.ca	‘coca leaves’	*Test*
ham.ˈpi.ri	‘healer’	*Test*
ha.tun.ˈma.ma	‘grandmother’	*Test*
i.ˈmi.lla	‘girl’	*Test*
ju.ˈk’u.cha	‘mouse’	*Test*
ˈlla.ma	‘llama’	*Test*
ˈlla.pa	‘lightening’	*Test*
ˈma.ma	‘mom’	*Test*
ˈpam.pa	‘prairie’	*Test*
ˈpa.pa	‘potato’	*Test*
ˈp’es.qo	‘bird’	*Test*
ˈpun.chu	‘poncho’	*Test*
ˈpun.ku	‘door’	*Test*
ˈq’a.pa	‘palm of hand’	*Test*
ˈq’e.pi	‘bundle’	*Test*
ˈqol.qe	‘money’	*Test*
ˈrun.tu	‘egg’	*Test*
ˈsun.kha	‘beard’	*Test*
ˈtha.pa	‘nest’	*Test*
ˈt’i.ka	‘flower’	*Test*
u.ˈhu.t’a	‘sandal’	*Test*
ˈwa.ka	‘cow’	*Test*
ˈwall.pa	‘chicken’	*Test*
ˈwa.wa	‘baby/child’	*Test*

* ˈ indicates stress, ’ indicates ejective, ’.’ indicates syllable boundary

Descriptive statistics of Delta-F-normalized formants

Table 2.19: DeltaF-normalized formant frequencies by age group: [a]

Age	F1 Median	MAD	F2 Median	MAD	F3 Median	MAD	F4 Median	MAD
4	0.70	0.11	1.28	0.10	2.63	0.26	NA	NA
5	0.56	0.20	1.31	0.12	2.70	0.31	NA	NA
6	0.69	0.12	1.33	0.13	2.72	0.30	NA	NA
7	0.61	0.16	1.34	0.16	2.67	0.28	NA	NA
8	0.69	0.11	1.29	0.11	2.58	0.38	NA	NA
9	0.64	0.14	1.38	0.09	2.59	0.29	NA	NA
10	0.61	0.11	1.35	0.19	2.60	0.34	NA	NA
adult	0.61	0.13	1.34	0.13	2.44	0.22	3.42	0.21

Table 2.20: DeltaF-normalized formant frequencies by age group: [i]

Age	F1 Median	MAD	F2 Median	MAD	F3 Median	MAD	F4 Median	MAD
4	0.25	0.04	2.29	0.20	2.81	0.27	NA	NA
5	0.27	0.05	2.24	0.22	2.81	0.23	NA	NA
6	0.25	0.06	2.25	0.18	2.79	0.19	NA	NA
7	0.26	0.06	2.27	0.18	2.83	0.21	NA	NA
8	0.25	0.04	2.19	0.16	2.77	0.26	NA	NA
9	0.26	0.06	2.22	0.22	2.79	0.29	NA	NA
10	0.31	0.05	2.21	0.23	2.79	0.27	NA	NA
adult	0.33	0.05	2.20	0.18	2.71	0.13	3.68	0.25

Table 2.21: DeltaF-normalized formant frequencies by age group: [u]

Age	F1 Median	MAD	F2 Median	MAD	F3 Median	MAD	F4 Median	MAD
4	0.34	0.05	0.68	0.09	2.50	0.49	NA	NA
5	0.35	0.06	1.00	0.13	2.65	0.43	NA	NA
6	0.31	0.06	0.92	0.27	2.63	0.23	NA	NA
7	0.33	0.05	1.03	0.41	2.73	0.27	NA	NA
8	0.36	0.04	0.83	0.24	2.63	0.29	NA	NA
9	0.30	0.08	0.98	0.27	2.81	0.37	NA	NA
10	0.32	0.06	1.04	0.22	2.92	0.39	NA	NA
adult	0.37	0.07	0.91	0.33	2.54	0.23	3.52	0.38

Category dispersion

Table 2.22: Vowel category dispersion by age group and phone: all data

Age	[a] mean (SD) range	[i]	[u]
4	0.17 (0.11) 0.02 - 0.57	0.22 (0.2) 0.03 - 0.94	0.15 (0.12) 0.03 - 0.42
5	0.2 (0.11) 0.01 - 0.56	0.23 (0.21) 0.03 - 1.36	0.13 (0.09) 0.02 - 0.42
6	0.19 (0.11) 0.02 - 0.55	0.19 (0.16) 0.01 - 0.97	0.32 (0.26) 0.04 - 1.08
7	0.19 (0.09) 0.01 - 0.54	0.18 (0.12) 0.02 - 0.67	0.38 (0.27) 0.01 - 1.37
8	0.21 (0.23) 0.01 - 1.12	0.16 (0.17) 0.01 - 1.28	0.36 (0.23) 0.1 - 1.07
9	0.16 (0.11) 0.02 - 0.72	0.17 (0.13) 0.02 - 0.64	0.22 (0.15) 0.02 - 0.75
10	0.19 (0.13) 0 - 0.89	0.16 (0.11) 0.02 - 0.5	0.21 (0.14) 0.02 - 0.65
adult	0.18 (0.12) 0.01 - 0.63	0.12 (0.07) 0.01 - 0.3	0.29 (0.21) 0.03 - 1.11

Dispersion measured via CoV

Table 2.23: Average CoV by formant and vowel for children and adults

	F1			F2		
Age	[a]	[i]	[u]	[a]	[i]	[u]
4	0.26	0.18	0.15	0.09	0.13	0.30
5	0.30	0.19	0.14	0.10	0.16	0.15
6	0.23	0.22	0.22	0.12	0.12	0.38
7	0.26	0.23	0.23	0.13	0.09	0.42
8	0.17	0.19	0.18	0.18	0.13	0.41
9	0.22	0.22	0.20	0.11	0.10	0.26
10	0.22	0.17	0.17	0.16	0.09	0.34
adult	0.27	0.15	0.22	0.13	0.07	0.34

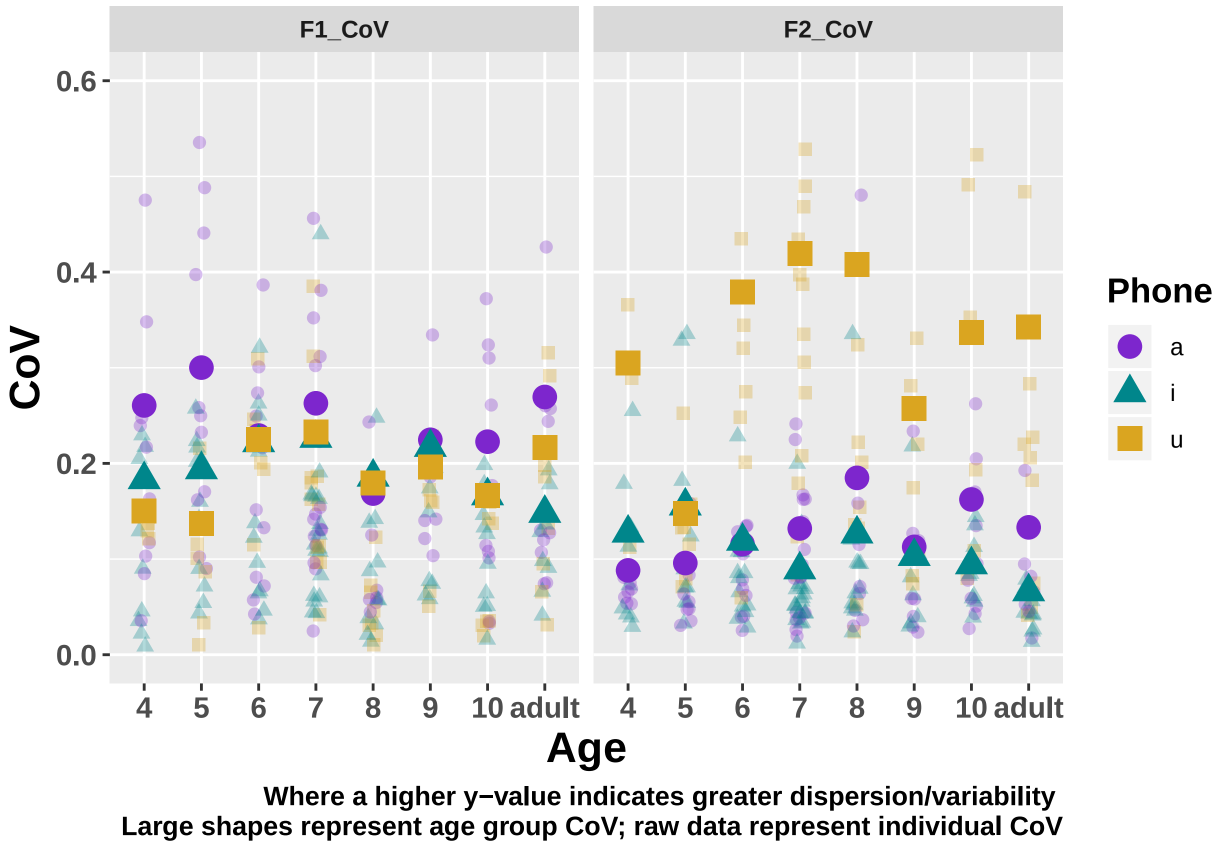


Figure 2.10: Group-level category dispersion by age and phone

Models predicting dispersion measured via CoV

Table 2.24: Model predicting CoV for F2 of [i]

term	estimate	S.E.	z.statistic	p.value	95% CI
Intercept	-3.23	0.19	-17.22	0.00	-2.87,-3.6
10	0.71	0.27	2.68	0.01	1.23,0.19
4	0.69	0.27	2.54	0.01	1.23,0.16
5	0.93	0.26	3.60	0.00	1.44,0.43
6	0.57	0.25	2.24	0.03	1.07,0.07
7	0.31	0.23	1.33	0.18	0.76,-0.15
8	0.69	0.25	2.76	0.01	1.18,0.2
9	0.49	0.29	1.66	0.10	1.06,-0.09

Table 2.25: Model predicting CoV for F1 of [u]

term	estimate	S.E.	z.statistic	p.value	95% CI
Intercept	-1.77	0.20	-8.63	0.00	-1.37,-2.17
10	-0.76	0.31	-2.46	0.01	-0.15,-1.37
4	-0.28	0.41	-0.68	0.50	0.52,-1.08
5	-0.54	0.31	-1.74	0.08	0.07,-1.15
6	0.10	0.31	0.34	0.74	0.71,-0.5
7	0.01	0.27	0.06	0.96	0.54,-0.51
8	-1.22	0.30	-4.10	0.00	-0.64,-1.81
9	-0.14	0.31	-0.46	0.64	0.46,-0.75

Descriptive statistics of Lobanov-normalized formants

Table 2.26: Lobanov-normalized formant frequencies by age group: [a]

Age	F1 Median	MAD	F2 Median	MAD	F3 Median	MAD	F4 Median	MAD
4	1.00	0.48	-0.64	0.28	-0.37	1.04	NA	NA
5	1.05	1.22	-0.56	0.35	-0.16	0.92	NA	NA
6	1.54	0.60	-0.48	0.32	0.10	1.20	NA	NA
7	1.30	1.03	-0.58	0.25	-0.48	1.08	NA	NA
8	1.29	0.33	-0.48	0.29	-0.37	1.17	NA	NA
9	1.56	0.77	-0.42	0.29	-0.52	0.78	NA	NA
10	1.55	0.67	-0.44	0.46	-0.48	0.96	NA	NA
adult	1.37	0.90	-0.28	0.31	-0.69	1.04	-0.41	0.84

Table 2.27: Lobanov-normalized formant frequencies by age group: [i]

Age	F1 Median	MAD	F2 Median	MAD	F3 Median	MAD	F4 Median	MAD
4	-0.88	0.18	1.14	0.40	0.33	1.15	NA	NA
5	-0.64	0.26	0.98	0.43	0.24	1.01	NA	NA
6	-0.82	0.22	0.99	0.37	0.21	0.71	NA	NA
7	-0.81	0.32	1.04	0.24	0.39	0.91	NA	NA
8	-0.94	0.28	0.97	0.34	0.41	0.95	NA	NA
9	-0.68	0.18	1.05	0.22	0.50	0.76	NA	NA
10	-0.78	0.30	0.97	0.37	0.02	1.08	NA	NA
adult	-0.79	0.42	1.30	0.28	0.52	0.77	0.63	0.77

Table 2.28: Lobanov-normalized formant frequencies by age group: [u]

Age	F1 Median	MAD	F2 Median	MAD	F3 Median	MAD	F4 Median	MAD
4	-0.55	0.27	-1.47	0.21	-0.79	0.67	NA	NA
5	-0.24	0.35	-1.05	0.21	-0.84	1.37	NA	NA
6	-0.42	0.30	-1.15	0.35	-0.51	1.34	NA	NA
7	-0.29	0.62	-1.09	0.53	-0.14	1.24	NA	NA
8	-0.46	0.23	-1.03	0.41	-0.20	1.14	NA	NA
9	-0.57	0.34	-0.93	0.36	0.17	1.21	NA	NA
10	-0.58	0.42	-0.90	0.30	0.67	1.25	NA	NA
adult	-0.48	0.54	-0.95	0.51	-0.29	0.86	0.07	1.24

Vowel plots

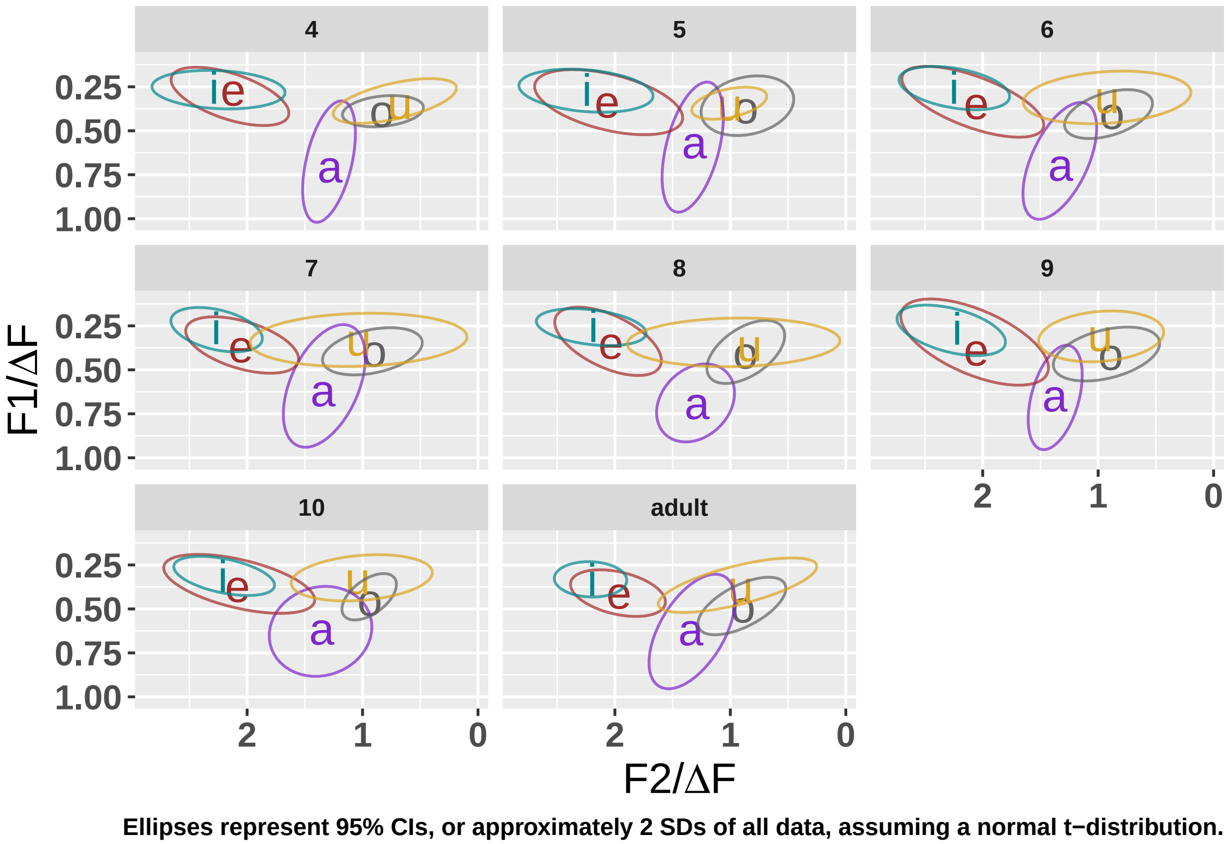


Figure 2.11: DeltaF-normalized vowels by age: adults and children

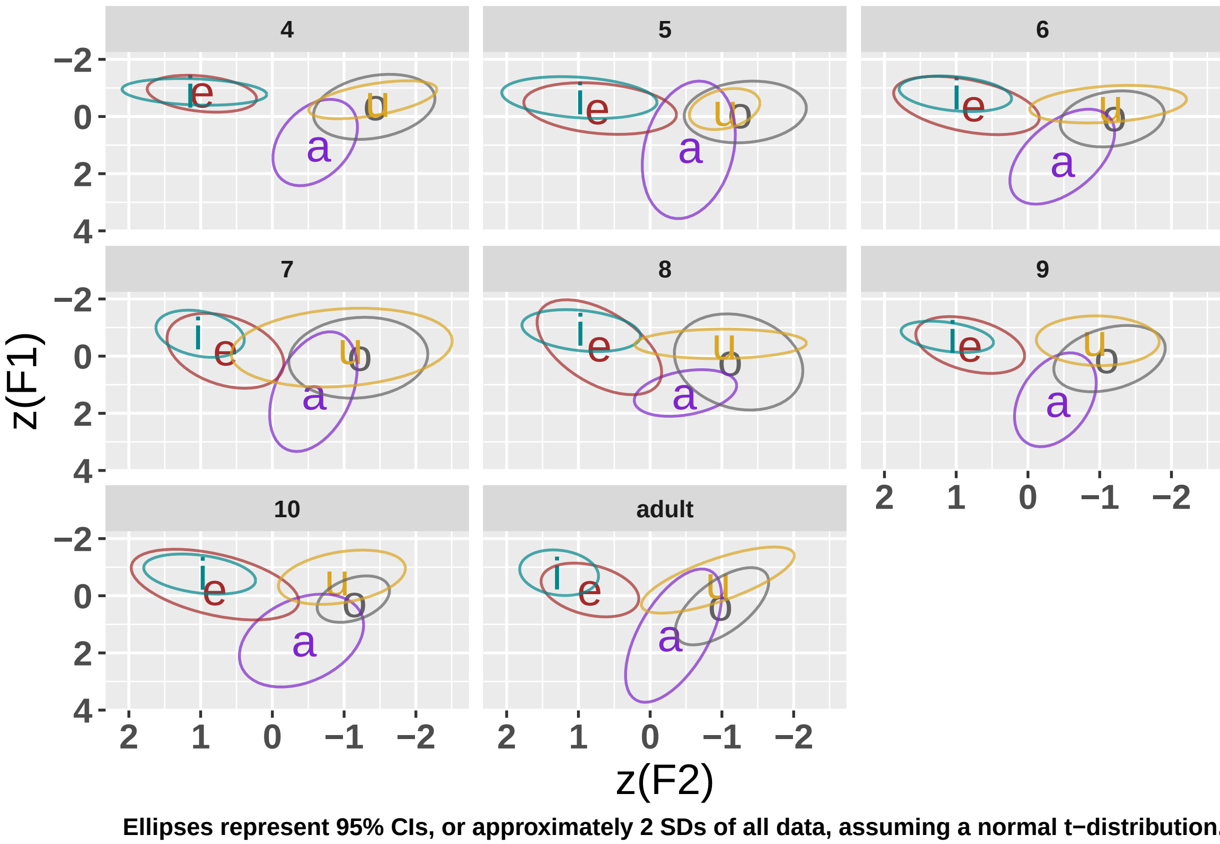
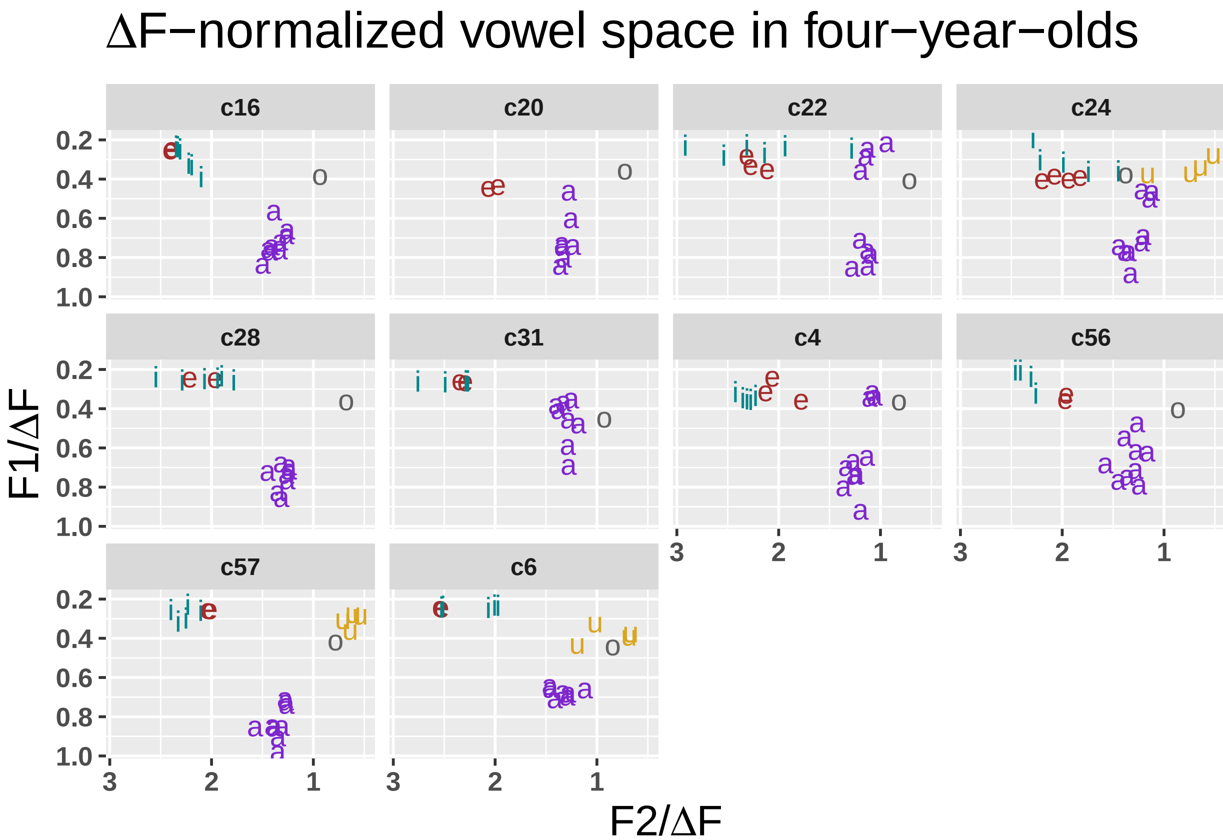
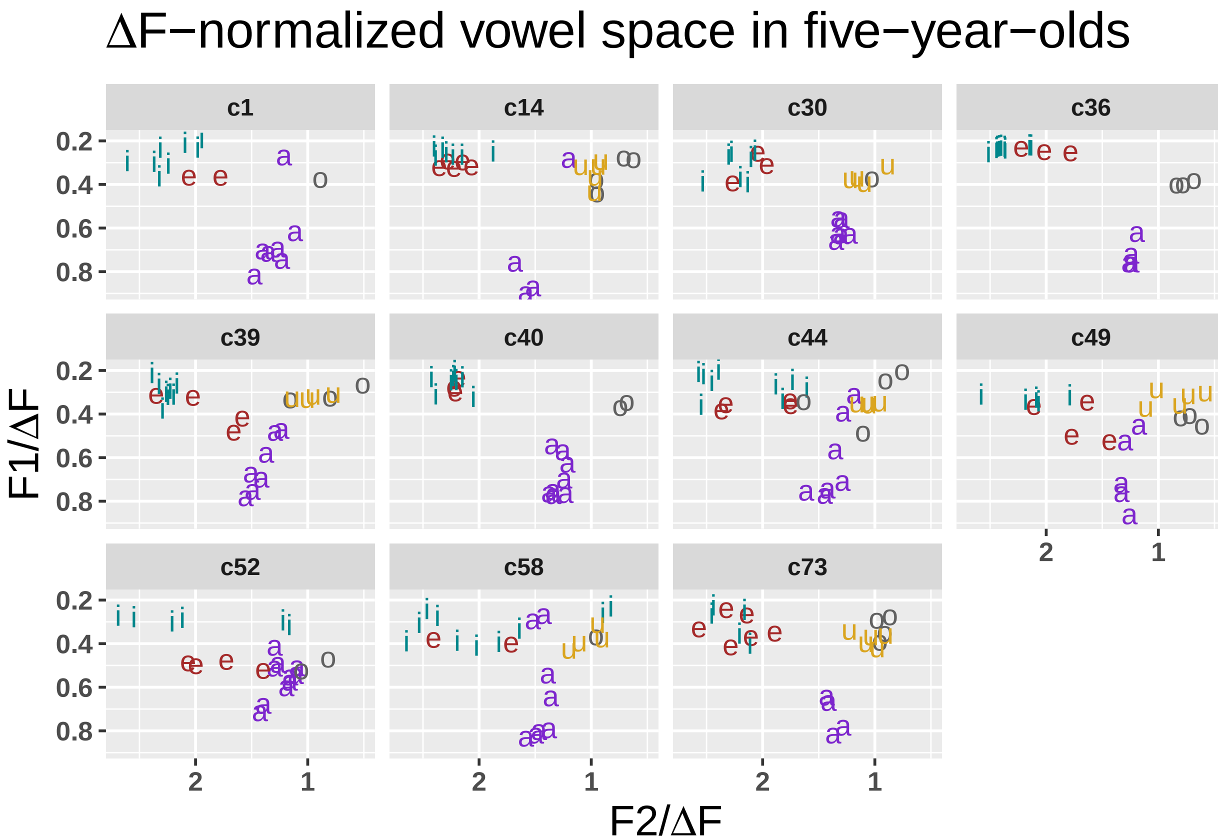
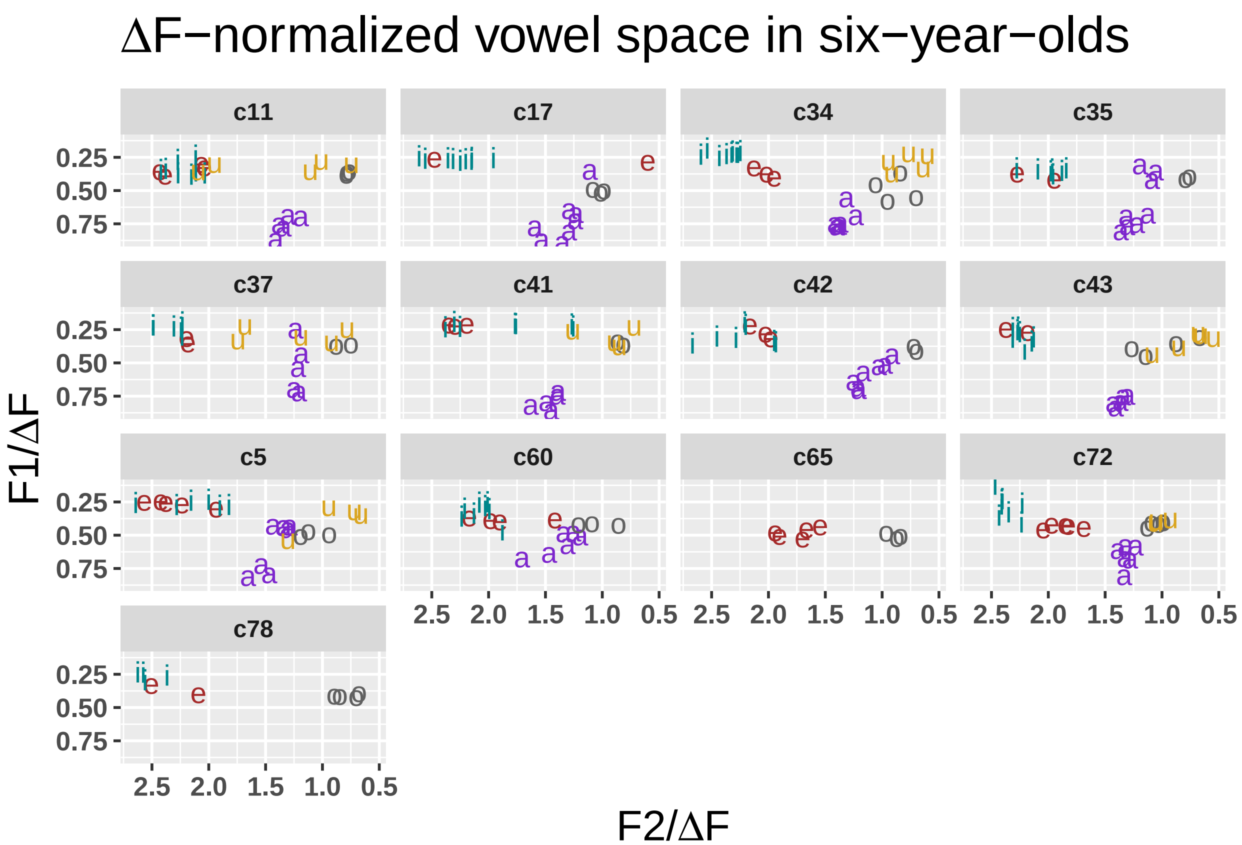
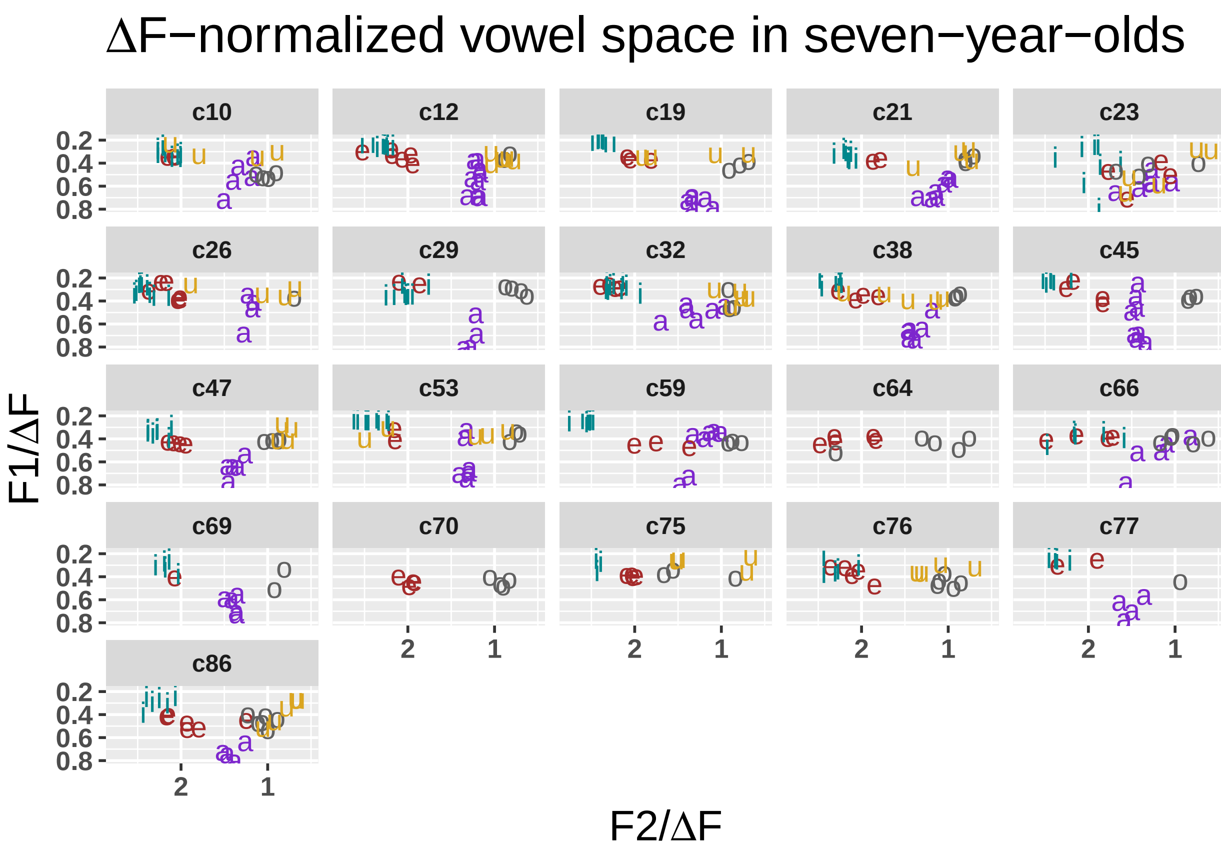


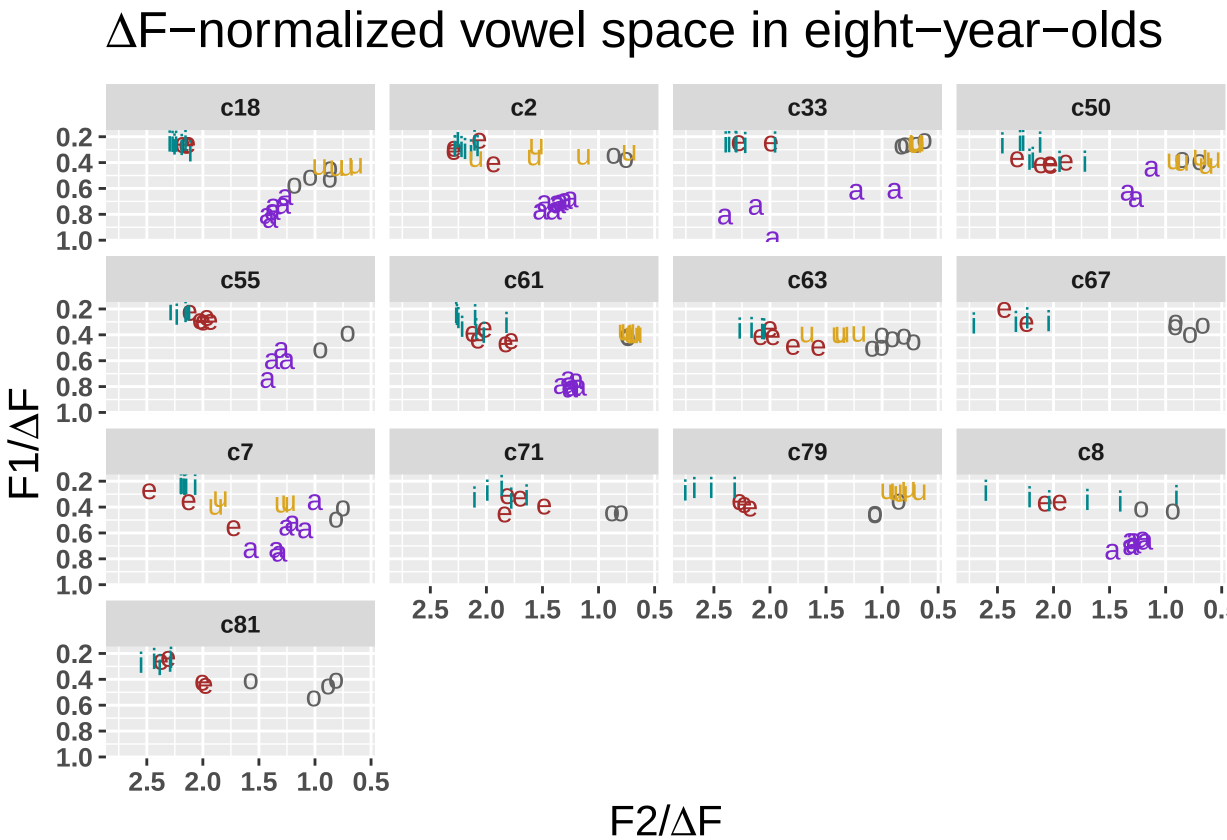
Figure 2.12: Lobanov-normalized vowels by age: adults and children

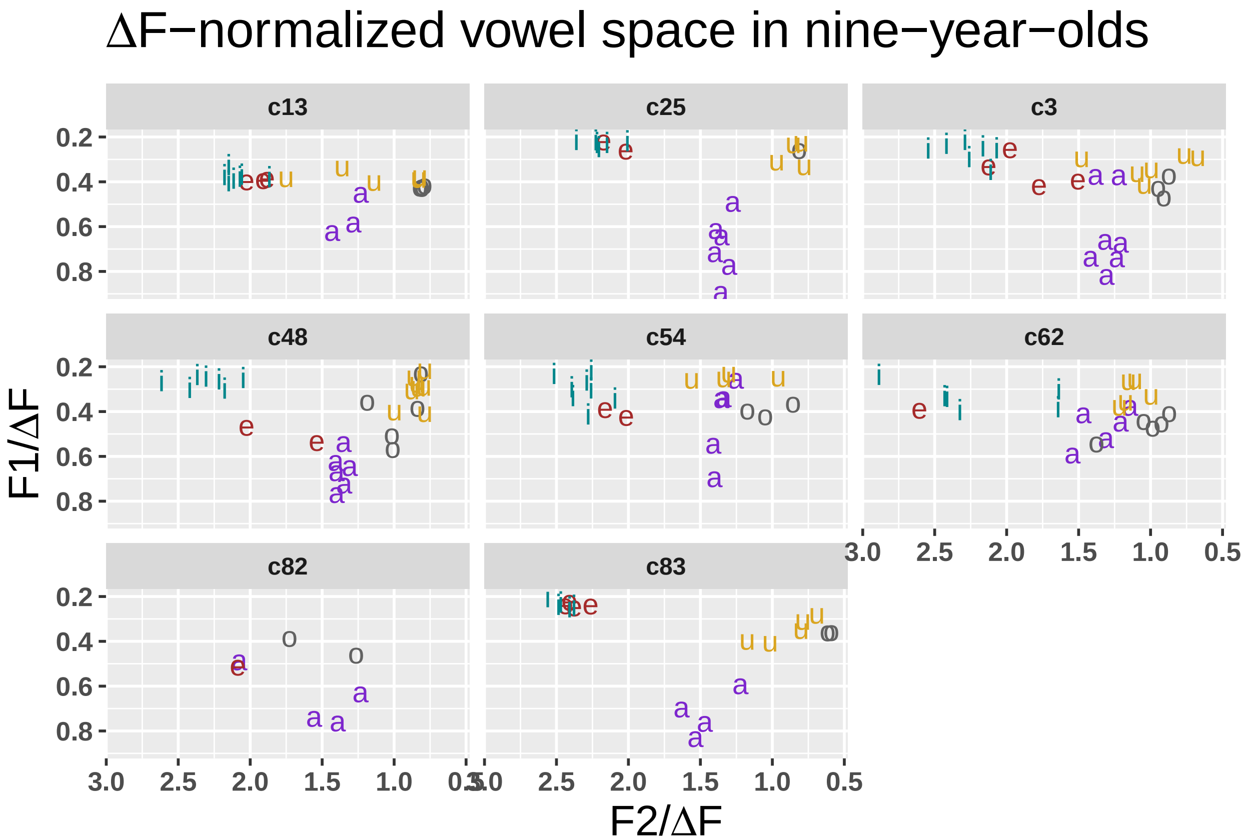

ΔF−normalized vowel space in four−year−olds
c16
c20
c22
c24
c28
c31
c4
c56
c57
c6
F1/ΔF
F2/ΔF
0.2
0.4
0.6
0.8
1.0
3
2
1

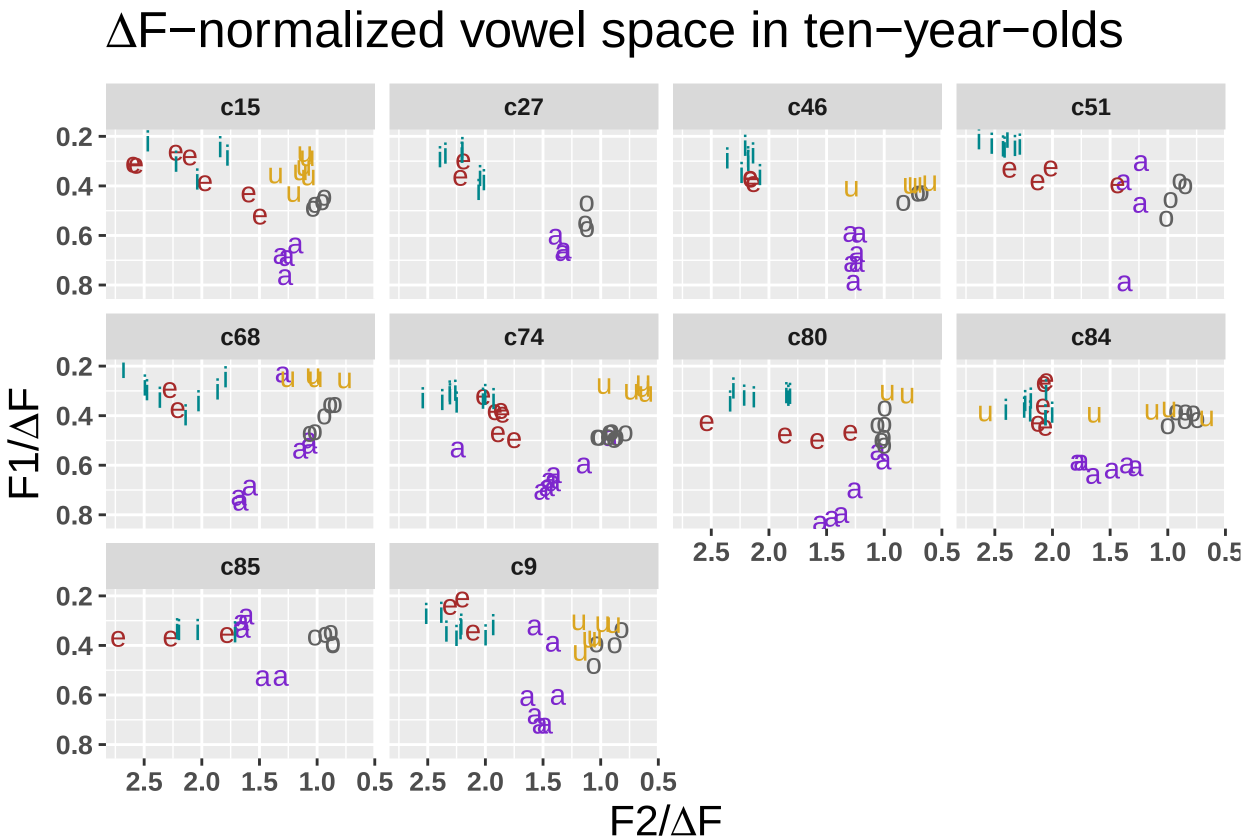

ΔF−normalized vowel space in five−year−olds
c1
c14
c30
c36
c39
c40
c44
c49
c52
c58
c73
F1/ΔF
F2/ΔF
0.2
0.4
0.6
0.8
2
1

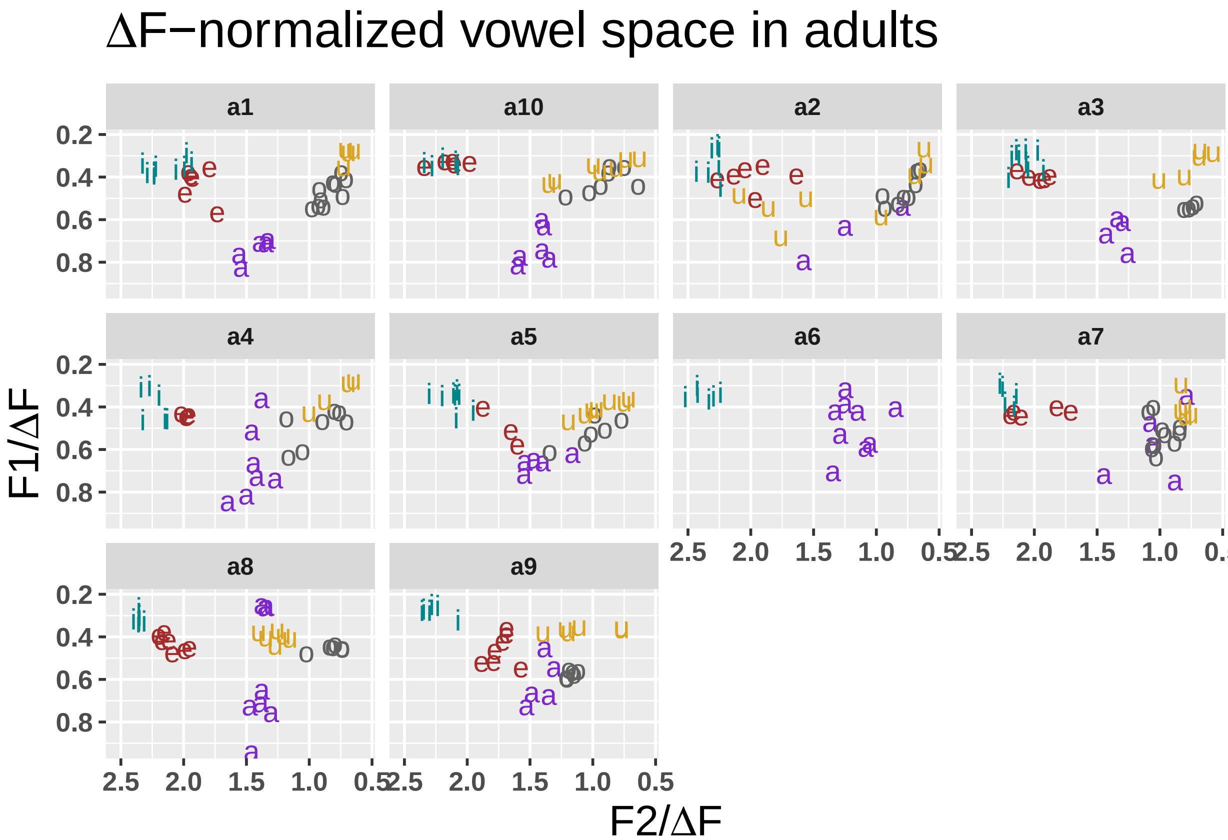

ΔF−normalized vowel space in six−year−olds
c11
c17
c34
c35
c37
c41
c42
c43
c5
c60
c65
c72
c78
F1/ΔF
F2/ΔF
0.25
0.50
0.75
2.5
2.0
1.5
1.0
0.5


ΔF−normalized vowel space in seven−year−olds
c10
c12
c19
c21
c23
c26
c29
c32
c38
c45
c47
c53
c59
c64
c66
c69
c70
c75
c76
c77
c86
F1/ΔF
F2/ΔF
0.2
0.4
0.6
0.8
2
1


ΔF−normalized vowel space in eight−year−olds
c18
c2
c33
c50
c55
c61
c63
c67
c7
c71
c79
c8
c81
F1/ΔF
F2/ΔF


ΔF−normalized vowel space in nine−year−olds
c13
c25
c3
c48
c54
c62
c82
c83
F1/ΔF
F2/ΔF
0.2
0.4
0.6
0.8
3.0
2.5
2.0
1.5
1.0
0.5


ΔF−normalized vowel space in ten−year−olds
c15
c27
c46
c51
c68
c74
c80
c84
c85
c9
F1/ΔF
F2/ΔF
0.2
0.4
0.6
0.8
2.5
2.0
1.5
1.0
0.5


ΔF−normalized vowel space in adults
a1
a10
a2
a3
a4
a5
a6
a7
a8
a9
F1/ΔF
F2/ΔF
0.2
0.4
0.6
0.8
2.5
2.0
1.5
1.0
0.5

Chapter 3

A morphological factor: Morphology and phonetics in early Quechua speech

3.1 Introduction

Adult speakers can readily compose novel, morphologically complex word forms like *daxy* or *unraxlike*. Evidence of u-shaped learning curves in children, and the results of Berko (1958)'s Wug Test, have led developmental researchers to recognize that young children must share in adults' morphological productivity. If not, children would not be able to extend morphophonological patterns, such as the correct plural allomorphs, to novel lexical environments.

There are factors, beyond the ability to flesh out a morphological paradigm with novel forms, that reveal whether children have adult-like production mechanisms for complex words. One window into the nature of lexical access and storage is to examine the effects of morphological structure on speech production - at least in adult speakers. Adults, and even children, have been shown to produce complex words differently, arguably as a result of planning differences, or maybe practice (Cho 2001; Hay 2003; Lee-Kim et al. 2013; Plag 2014; Song et al. 2013a; Song et al. 2013b; Strycharczuk 2019; Sugahara and Turk 2009; Tomaschek et al. *under review*). For example, in adult speech, morphologically complex words like *sighed* can be longer in duration than their simplex counterparts like *side* (Sugahara and Turk 2009). And children have been shown to coarticulate more between adjacent segments within mono-morphemic words like *box* than between otherwise identical segments that straddle a morpheme boundary as in *rocks* (Song et al. 2013b).

This chapter takes advantage of these known relationships between acoustics, specifically coarticulation and duration, and word structure. The complex relationship between speech production and morphological structure in children is still relatively under-explored (cf. Redford 2018; Song et al. 2013b). Yet measuring how morphological structure is reflected in children's speech could allow us to infer about the status of children's early complex word forms. Are words that, to an adult are transparently morphologically complex, treated as such by children? And when a child does recognize that a given word is morphologically complex, how does this recognition affect children's lexical storage, access, and production? To that end, the goal of this chapter is to evaluate how the composition of morphologically complex words is reflected in children's speech production patterns in Quechua, a highly agglutinating language with over 200 unique, productive verbal and nominal suffixes.

Given the focus on production and acoustics in child speech, the first part of this chapter is a validation study that employs two novel measures of coarticulation that are relatively immune to the measurement difficulties (i.e. formant tracking) inherent to children's high fundamental frequencies. The prediction is that these measures will replicate known lingual coarticulation findings from the literature, but in a large sample of four-year-old children whose high f0 and breathiness might prove problematic for other acoustic measures of coarticulation.

The second part of this chapter is an experimental study that extends the growing line of research evaluating relationships between morphology and speech production (e.g. Hay 2003; Plag 2014; Song et al. 2013a; Song et al. 2013b; Strycharczuk 2019; Tomaschek et al. *under review*). For this study, coarticulation is measured across and within morpheme boundaries in adults and a cross-sectional cohort of school-aged children to examine if, and how, Quechua speakers distinguish coarticulatorily between morphological environments. Here, it is anticipated that children, who have less experience with language and potentially less abstract linguistic categories, will coarticulate similarly between and within morpheme boundaries. This behavior would suggest that the children are not as likely to analyze the internal structure of complex word forms. Adults, however, will coarticulate differently between and within morpheme boundaries because adults analyze the internal structure of complex words and compose these words together from their component morphemic parts.

The choice to study these patterns by word environment in a highly agglutinating language like Quechua is an important one. Quechua may make an important typological contribution to this literature because, given its morphological complexity, there are few doubts as to children's productivity from a very young age (Courtney and Saville-Troike 2002). Furthermore, the sheer number of available inflectional forms in Quechua, compared to more analytic languages such as English, likely translates into less competition between forms accessed in their inflected form and those composed online from individual morphemes during word retrieval (Hay 2003; Pinker and Ullman 2002). In short, evidence that children are not analyzing the internal structure of complex words would be especially noteworthy in a language of this structure.

3.2 Background

Accessing complex words

Adult speakers' ability to compose novel, morphologically complex words (e.g. *twinkly*) appears to be highly flexible and abstract. Children's early overgeneralization errors (e.g. *I* *taked**) demonstrate how even very young speakers analyze the internal structure of words and apply morphemes in new, unheard environments. Consequently, morphological productivity is a defining characteristic of the language faculty after a certain point in development.

Still, there is some behavioral evidence that adult speakers do not uniformly decompose complex words (Taft and Forster 1976). Using lexical frequency statistics, some work suggests that adults may instead access complex words, especially high-frequency complex words, holistically in the mental lexicon, without decomposing them (Baayen 1992; Baayen et al. 2003). For example, Colé et al. (1997) tested this idea and demonstrated a processing advantage for high-frequency base stems (e.g. *fish*) versus derived words (e.g. *fisher*). Crucially, this advantage diminished when the derived word was more frequent than the base stem.

Hay (2003) likewise demonstrated relative frequency effects. Measuring speech production (contrastive pitch accent placement and phonetic reduction at morpheme boundaries), Hay showed that when derived words are more frequent than their corresponding base form (e.g. *disentangle* vs. *entangle*), these words may be accessed holistically (see Pluymaekers et al. 2010 for an alternative explanation). Similarly, greater root frequency relative to a suffix is usually predictive of decomposition, at least in more analytic languages (Smith et al. 2012). And less frequent words, or words that only occur with a given suffix, do not necessarily manifest complete morphological decomposition (Kemps et al. 2005). Hay (2003)'s findings in particular imply a dual-route model of complex word access, where two different lexical mechanisms - holistic storage and online complex word formation - compete during complex word production (Baayen 1992; Koenig and Jurafsky 1995; Pinker and Prince 1994; Pinker and Ullman 2002). The method of construction that is fastest, predicted by the relative frequency of base to derived form, wins.

Finally, a number of computational models have attempted to predict decomposition versus whole-word storage. In a model of English morphology, Plag and Baayen (2009) demonstrated that affixes that are highly parsable - affixes that speakers can more easily separate from their root (e.g. English -*ness*) - predict decomposition. Affixes that are highly fused to the stem - those that speakers are less able to parse from the stem (e.g. English -*th*) - predict whole-word access. And more recently, O'Donnell (2015) proposed a probabilistic model where speakers weigh the productiveness of rules versus the storage of idiosyncratic, complex words to predict decomposition.

Given children's u-shaped learning curves, and their performance on the Wug test (Berko 1958), it is often assumed that children as young as four years share adults' morphological productivity capacities. Consequently, the above findings that cast doubt on whether composition is always the online strategy that wins during lexical access for *adults* are highly relevant to the study of how children access morphologically complex words. Frequency ratios appear to predict word (de)composition in adults (e.g. Hay 2003). Thus, since frequency ratios change as children learn more words, stems, and suffixes, one can predict that the fastest manner for children to access a complex word - by composition or holistic access - will also change over the course of development. But lexical frequency ratios are just one method to evaluate complex word access. Another method may be to evaluate how children, particularly a cross-sectional cohort of children spanning several years of development, access complex words throughout development. This is the objective of the current study.

Morphological structure and speech production

One way to study how speakers analyze complex words, and access them in the mental lexicon, is to measure the effect of morphological structure on speech production. Morphophonetics, or how morphological structure interacts with phonetic variability, is increasingly studied in adult speech (see Plag (2014) and Strycharczuk (2019) for recent overviews). Researchers now know that morphological composition dictates some speech variability in adults (Cho 2001; Guy 1991; Hay 2003; Plag et al. 2017; Smith et al. 2012; Sugahara and Turk 2009; Tomaschek et al. *under review*; cf. Hanique and Ernestus 2012), but it is often unclear when, how much, or even why (Mousikou et al. 2015; Plag 2014).

Adult speech

One of the earliest studies to examine the relationship between phonetics and word structure was Losiewicz (1995) who found that frequency affected the phonetic realization of morphemes. Specifically, English past tense -*ed* was temporally longer on low-frequency verbs than high-frequency verbs.[1]. Elsewhere, Sugahara and Turk (2009) found that heteromorphemic words were longer than otherwise identical monomorphemic words (e.g. *sighed* versus *side*), while Smith et al. (2012) found that pseudo-prefixes (*mistake*) were shorter in duration than real prefixes (*misjudge*). Finally, Seyfarth et al. (2018) examined paradigmatic influences on segmental duration in English, for example predicting differences between *frees* and *freeze* based on membership in the *free* paradigm. After controlling for lexical frequency, prosodic structure, and orthography, the authors found that paradigm membership can sometimes predict segment duration, though this depends upon the segment studied (fricatives versus stops).

Articulatory methods have likewise been used to evaluate the relationship between morphology and phonetic variation in adult speech. Cho (2001) found that intergestural timing in Korean, measured with electromagnetic articulagraphy and electropalatography, was more stable within a word than across a morpheme boundary. The author found a similar effect for a lexicalized compound word versus a non-lexicalized compound word (noun phrase). In addition, English /l/ darkness has been found to vary by position at morphological boundaries (Lee-Kim et al. 2013). The phone /l/ is darker in stem-final position (*coolest*) than affix–initial (*coupless*), independent of duration, because *coolest* is analogizing to word-final /l/ in its base form *cool* (See also Strycharczuk and Scobbie 2016). Most recently, Tomaschek et al. (*under review*) measured anticipatory coarticulation patterns in English. The authors used electromagnetic articulagraphy to evaluate the degree of morphological opacity in inflected verbs. Results showed that English speakers exhibited greater anticipatory coarticu-

[1] Recent statistical advances have called into question the results of Losiewicz (1995)(see Seyfarth et al. (2018)).

lation of the vowel in a verb stem (e.g. *clean*) for those verb inflections with which they had increased experience and practice.

However, the relationship between speech production and morphology is not always straightforward. For example, Plag et al. (2017) found that, after controlling for a host of other factors such as number of syllables, speaking rate, and surrounding context, the duration of English /s/ and /z/ varied systematically by morphological status. But the morphemic plural and non-morphemic /s, z/ were longer in duration than morphemic /s, z/ used as third person markers. The cause of this difference is not clear, though some authors suggest that the spontaneous speech analyzed in Plag et al. (2017) may explain the findings (Seyfarth et al. 2018). Elsewhere, others have argued that it is not morphological structure but word information load (Hanique and Ernestus 2012; Pluymaekers et al. 2010) or contextual predictability (Cohen 2014) that explains these interactions between speech production and word structure.

A complete review on the consequences of morphological structure for adult speech production is beyond the scope of a paper on children's speech development. What is important, however, is that the previous two decades of morphophonetics research has converged somewhat on the fact that words with productive morphology are processed and produced differently than simplex words (Kemps et al. 2005; Plag 2014; Tomaschek et al. *under review*). For child speech, these findings present a new method to detect when and whether children are decomposing words. Furthermore, the current work extends the study of morphophonetics to Quechua. This could be an important typological contribution because Quechua's morphology is highly productive, more so than in other more analytic languages that tend to be the object of study. As a result, children's productivity from a young age is not in doubt (Courtney and Saville-Troike 2002). And, furthermore, Quechua's agglutinating structure means that there should be less relative frequency effects between forms composed online and those accessed in the inflected form (Baayen 1992; Hay 2003; Pinker and Ullman 2002).

The goal of this paper is twofold. First, it extends the acknowledged relationship between speech production and word structure from adults to a cross-sectional sample of children. Second, it examines production and word structure in a language with a vastly different morphological structure from the western European languages that have been the focus of previous research.

Child speech

While we understand some of the explanatory mechanisms behind morphological structure and phonetic variation in adult speech, studies have rarely examined this relationship in children (cf. Song et al. 2013a; Song et al. 2013b). As an uncharted area in speech development, morphophonetics has the potential to inform our understanding of the structure of the mental lexicon throughout development. For the current study, examining child morphophonetics may also contribute to a line of research interested in the status of children's early word forms: some theorists posit that children's lexical forms are fully adult-like from early toddlerhood (e.g. Wexler 1998), while others posit that they emerge over time with usage (e.g. Ambridge et al. 2015).

A handful of studies have used morphophonetics to address this question. Song et al. (2013b) studied morphophonetic development in English-learning children. They employed acoustic and ultrasound analyses to study adult and 2;0 (year;month) children's CC syllable (/ks/) coarticulation in the coda of bimorphemic (*rocks*) and monomorphemic (*box*) (These, along with *rock* and the nonword *das*, were the items tested.). The articulatory data showed that children reliably raised their tongue more for *rocks* than *box*, suggesting that they could distinguish between morphemic and non-morphemic coda consonant clusters. While there was no evidence of anticipatory coarticulation for *box* in the adults or children studied, there was a strong perseveratory coarticulation effect of /k/ on /s/ in *box* and anticipatory coarticulation effect of /s/ on /k/ in *rocks*. This finding led the authors to conclude that the primary articulatory target for monomorphemic *box* was the C_1 of the consonant cluster (/ks/), while the articulatory target for bimorphemic *rocks* was C_2 or the plural morpheme /s/. The articulatory target differs by morphological role, suggesting that semantic information may be encoded in articulatory gestures. This applied for both adults and children.

Elsewhere, coda position American English fricatives /s, z/ were studied in the naturalistic speech of children and their caretakers (Song et al. 2013a). Children's morphemic /z/ was longer in duration than non-morphemic /z/ in word-final position. This acoustic evidence suggests that, as early as two years of age, child speakers of American English reliably distinguish between otherwise identical morphemic and non-morphemic segments. The authors used this finding to argue that even at this young age children may not rote-memorize morphologically complex forms, but may instead compose the words online.

Finally, Redford (2018) studied the relationship between word structure and phonetic variation across word boundaries in prosodic words (e.g. *the bat*). The results showed that schwas in children's *the* productions were reliably louder and longer than in adults' productions. According to Redford, this result was strong evidence that children exhibit immature timing control, but the results were inconclusive concerning children's speech organization - whether children accessed the prosodic words holistically or in parts.

These developmental studies aside, the relationship between speech production and morphological structure has been almost entirely unexplored in children. This is to the detriment of the field because a parallel line of research on children's coarticulation has shed considerable light on early phonological representations. Conflicting evidence from this research suggests that children may store speech and word forms in larger chunks than adults, at the syllabic level or higher (e.g. Nittrouer et al. 1989; Zharkova et al. 2011; Noiray et al. 2019b). The following section summarizes this work on child coarticulation.

Children's phonological storage: Evidence from coarticulation

Numerous works have found that children coarticulate more than adults until early puberty, with the degree and variability of coarticulation decreasing with age (Goodell and

Studdert-Kennedy 1992; Nittrouer et al. 1989; Nittrouer et al. 1996; Noiray et al. 2019b; Zharkova et al. 2014; Zharkova et al. 2018 *inter alia*). Many authors have interpreted this relationship between coarticulation and age as evidence that young children represent speech in syllable-sized units,[2] which gradually individuate into phoneme-sized units over the course of development.

While one may expect children to coarticulate less than adults – children speak slower and with less coordinated movement (Smith and Goffman 1998) – many studies have concluded that children coarticulate *more* than adults (Goodell and Studdert-Kennedy 1992; Nittrouer et al. 1989; Nittrouer et al. 1996; Zharkova et al. 2011). Citing this evidence, some have argued that the phenomenon of "coarticulation" in child speech reflects children's phonological representations (this will be outlined in further detail in the following section). This approach thus argues that unlike adult coarticulation, child coarticulation does not necessarily reflect efficiency in speech planning or the clear anticipation of upcoming segments (Whalen 1990; Bradlow 2002), but instead reflects a larger representational unit (see Section 1.3).

Studying children's coarticulation, Nittrouer et al. (1989) concluded that children may have more holistic, syllabic representations. The authors measured coarticulation in fricative-vowel sequences within nonce words to test the anticipatory effect of vowels on /s/ and /ʃ/. Their results from children 3;0-8;0 showed that children coarticulated more than adults – vowels affected the children's fricative production more than the adults' production. The authors concluded that this pattern in the fricative-vowel sequences manifested in child speech because the children did not distinguish between /s/ and /ʃ/ as reliably as the adults. It was not, the authors argued, because children differed from adults in their degree of anticipatory lip rounding or even the lingual constriction shape (see also Nittrouer et al. 1996).

Nittrouer et al. (1989)'s findings partially supported conclusions from the two children studied in Repp (1986). There, the younger child (4;8) showed greater anticipatory lip rounding for /s/ before /u/ than the older child (9;5). More recently, the coarticulation patterns of children with and without apraxia of speech and adults without apraxia were studied (Nijland et al. 2002). Participants produced nonce sequences of ə-V-C where V was /a, ɪ, or u/ and C was /s, x, b, or d/. The typically-developing children once again showed more intra- and inter-syllabic anticipatory coarticulation than the adults.

Most recently, Zharkova and colleagues (Zharkova et al. 2011; Zharkova et al. 2014; Zharkova et al. 2018) and Noiray and colleagues (Rubertus and Noiray 2018; Noiray et al. 2019b; Noiray et al. 2018; Noiray et al. 2018) incorporated articulatory ultrasound data and have frequently observed that children coarticulate more than adults. The authors frequently infer that children's propensity to coarticulate reflects their larger representational units. In Zharkova et al. (2011), children aged (6;3-9;9) altered their productions of /ʃ/ based on the following vowel (/a, i, or u/) more than adults. However, Zharkova et al. (2014) conducted a

[2]The default assumption in these studies is to propose syllable-sized representations, but presumably the authors would also entertain other sub-lexical units such as morae or feet.

similar experiment with older children (10;0-12;4) and did not find any differences between child and adult coarticulation patterns. The authors suggest that children may approximate adult-like phonological representations by preadolescence.

Also using ultrasound imaging, Noiray et al. (2019b) measured vocalic anticipation in children (3;05-7;06) and adults. They found that in V#CV strings, children anticipated the upcoming vowel sooner than adults at age 3;0 and 5;0, but progressively less by 7;0. (See Noiray et al. 2018 for segment-specific explanations for coarticulation development.) Most recently, Noiray et al. (2019a) correlated children's (4;06-7;02) coarticulation, measured articulatorily with ultrasound imaging, with measures of expressive vocabulary and phonological awareness. The amount of children's intrasyllabic coarticulation and gestural individuation was negatively correlated with phonological awareness (a measure of speaker awareness of segmental units), which the authors suggest reflects a lack of segmental individuation in children.

Authors of the above acoustic and articulatory studies often argue that children's coarticulation patterns reflect holistic, syllable-sized representations. According to such an interpretation, representations progressively individuate and become more adult-like as children age. It is nevertheless important to note that results in the child coarticulation literature are mixed, and the methods for measuring children's coarticulation are varied. Numerous studies have found that adults coarticulate more than children. These works often argue that children's phonological representations are just as or more discretely organized into individual segments than adult phonology. As a result, children learn to coordinate and exhibit appropriate coarticulatory overlap as part of standard phonetic/phonological development (Barbier et al. 2013; Barbier et al. 2015; Katz et al. 1991; Kent 1983; Whiteside and Hodgson 2000). Other works have found no differences in coarticulatory patterns between adults and children (Flege 1988; Goffman et al. 2008; Noiray et al. 2013; Sereno and Lieberman 1987; Sereno et al. 1987).

Methodological differences between the studies summarized above may account for some of the different conclusions concerning children's coarticulation patterns. There are innumerable ways to measure coarticulation - even in child populations where data collection techniques are often more limited. There are acoustic techniques (e.g. Nittrouer et al. 1996) and articulatory approaches (e.g. Zharkova et al. 2014). But even within the acoustic realm there are different techniques. Most frequently, for fricatives, coarticulation is quantified as the change in centroid (average) based on vocalic context while coarticulation within vowels is typically quantified as the midpoint or steady-state formant value (Nittrouer et al. 1989; Nittrouer et al. 1996). Elsewhere, coarticulation has been measured acoustically as the spectral change between adjacent segments (Gerosa et al. 2006)

The current study employs two of these measures of spectral change, outlined in Gerosa et al. (2006) and validated in Cychosz et al. (2019). These measures take into account change occurring over the timecourse of segments, instead of static measurements. Children speak slower than adults, so taking measurements from the steady-state or midpoint of a vowel is not ideal. Children's slower speaking rate permits them greater opportunity to achieve a steady acoustic signature. Measuring coarticulation from a larger portion of the overall segment, which is what is proposed here, mitigates the effect of speaking rate to a certain degree.

Current study

The language

The current study measures coarticulation in morphologically complex words in South Bolivian Quechua, henceforth Quechua, a Quechua-II/C language with over 1.6 million speakers in southwest Bolivia and northwest Argentina (Torero 1964). This variety of Quechua is spoken in the Chuquisaca department of southern Bolivia. Like many Quechuan varieties, Quechua in southern Bolivia has been in intense contact with Spanish for hundreds of years resulting in large amounts of language mixing and borrowing (Muysken 2012a; Muysken 2012b). Furthermore, public schools in Bolivia are conducted primarily in Spanish. Consequently, almost all school-age children who speak Quechua in the home, such as the children in the present study, also speak Spanish. And because of Spanish schooling, children in Bolivia learn to read in Spanish. Though Quechua has an established writing system, children generally do not learn to read or write in Quechua and few Quechua speakers use the language in its written form.

Quechua is a highly-agglutinating language with over 200 productive nominal and verbal suffixes that encode argument structure and grammatical relations (for comparison, English has around 35 productive suffixes). The morphology is nevertheless highly regular and not subject to significant morphophonological processes or fusion. The phonological inventory includes three phonemic vowels /i, a, u/ and two allophonic vowels derived in uvular contexts [e, o] (Gallagher 2016). The consonantal inventory contrasts voiceless stops, aspirated stops, and ejectives along four places of articulation, /p, t, k, q/, as well as a three-way alveopalatal stop-aspirated stop-ejective distinction /ʧ, $ʧ^{h}$, ʧ'/. Nasals are contrasted along three places of articulation, /m, n, ɲ/ with an allophonic velar nasal. See Appendix A for a complete consonant inventory.

Research objectives

The primary objective of this chapter is to measure how Quechua word structure affects speakers' production of morphologically complex words. To accomplish this, two relatively novel acoustic measures of coarticulation are employed. In the first part of this chapter, these coarticulation measures are validated to ensure their applicability for young children's voices and a variety of consonants. The measures are tested on a large corpus of four-year-old English speakers by replicating broader anticipatory lingual coarticulation findings (Mooshammer et al. 2006; Recasens and Espinosa 2009b; Recasens et al. 1997). After validating the coarticulatory measurements, the second part of this chapter turns to coarticulation patterns in adult and child speakers of Quechua to evaluate if and how patterns differ between children and adults, and across development, in a cross-sectional sample of school-aged children.

Validation study background

Coarticulation has played numerous illustrative roles in phonetic and phonological theory. It motivated the development of Articulatory Phonology, a theory that takes individual, overlapping gestures as the basic units of speech (Browman and Goldstein 1989; 1992). Its ubiquity and relevance for accurate perception challenge more abstractionist models (Lahiri and Marslen-Wilson 1991). However, in this dissertation, these theoretical frameworks and debates are less relevant. Here coarticulation is quantified as the shared acoustic information between two adjacent phones (spectral) or the speed of transition between two phones (temporal). Of course articulatory gestures and acoustics are interrelated – articulatory modifications modulate the acoustic signal. And articulation is a significant factor in coarticulatory patterns and development (Iskarous et al. 2013). However, here the focus is upon the ensuing acoustics and not the underlying coarticulatory gestures.

The child speech apparatus creates multiple issues for the study of acoustic phonetics, spectral analyses in particular (Vorperian and Kent 2007). Small vocal tracts result in higher resonant frequencies and small vocal folds result in widely-spaced harmonics in the spectral envelope. This can render an undersampled spectral shape and obfuscate formant frequency peaks. Often the only remedy for such problems is to make arbitrary data cleaning decisions, such as excluding all measurements outside of a predetermined range (Lee et al. 1999).

Acknowledging the difficulties inherent to child acoustics, Gerosa et al. (2006) employed two novel measures of coarticulation to study consonant-vowel transitions in adult and child speech. The first, a spectral measure, calculates the distance between Mel-frequency cepstral coefficient (MFCC) vectors averaged over adjacent phones. The second measure, a temporal estimation, dynamically calculates the transition duration between phones in a given biphone sequence as a function of their spectral overlap (explained in further detail below). This measurement reflects what proportion of the CV sequence is spent in transition where a greater proportion of transition time indicates more coarticulation (i.e. a larger chunk of the biphone sequence is devoted to transitioning between steady-state targets).

The applicability of traditional coarticulation measures, such as formant transitions (Lehiste 1972; Öhman 1966) or the more dynamic Peak $\mathrm{ERB_N}$ (Reidy et al. 2017), is limited to speakers with longer vocal folds or to certain segments such as fricatives. However, the measures employed in Gerosa et al. (2006) are versatile and should be reliable for a broader range of speakers and all consonant manners.

Experimental study background

As an agglutinating language, Quechua provides unique insight into interactions of morphology and phonetics. Quechua speakers appear to have highly flexible inflectional and derivational lexicons: suffixes and roots are abstracted away from the original lexical contexts and are easily rearranged for novel stem+suffix pairings. This process is similar to how speakers of more analytic languages, such as English, effortlessly arrange novel noun-adjective pairings. It is this morphological structure that makes Quechua an interesting typological contribution to morphophonetics. For one thing, as noted in the previous sections, given the number of possible complex word forms, is difficult to doubt the the morphological productivity of young Quechua speakers. Quechua is understudied, both in child language research and linguistics more broadly. Still, much like well-studied, less synthetic languages, there is undeniable evidence of morphological productivity in children's Quechua (and other highly agglutinating languages). Child speakers of Cuzco Quechua appear to have a productive morphology by age 5;0, if not earlier (Courtney and Saville-Troike 2002).[3]

Another reason why Quechua makes an important typological contribution again concerns its morphological structure. In Quechua, frequency ratios between words that might be compiled online and those that might be accessed holistically are likely to be much smaller. This also means that the type frequency of each word is lower than in moderately synthetic languages, like English and Dutch, which are more commonly studied in morphophonetics. So concerns about a race between online compilation and holistic access may be less relevant.

One final reason why Quechua may be an interesting addition to the morphophonetics literature concerns its applied use in Bolivia. As described in section 3.2, very few Quechua speakers read or write in the language (speakers who attend school instead become literate in Spanish). Disengaging morphological effects from orthographic effects has been immensely difficult for morphophonetics, and the study of phonetics in general (Seyfarth et al. 2018; Warner et al. 2006). Quechua removes this problematic variable because speakers do not have strong orthographic representations.[4] Thus, while the primary typological interest of Quechua concerns its morphological productivity and what that means for the structure of the lexicon, there are additional, applied considerations that make Quechua a unique language to examine the effects of word structure on speech production.

Given these characteristics of Quechua, this experiment makes two predictions. First, relying on morphophonetic studies on adults and coarticulation studies on children, it predicts that children will produce complex words differently from adults. The hypothesis is tested in a tightly-controlled experimental paradigm where coarticulation between the phones [a] and [p], or [a] and [m], is measured. When [ap] or [am] straddle a morpheme boundary (e.g. sunkh**a-p**i 'beard-LOC'), adults, and potentially older children, will coarticulate less between the phones than when [ap] or [am] fall inside of a root morpheme (e.g. p**ap**a 'potato').

[3]Cuzco Quechua and South Bolivia Quechua are distinct varieties, but they are mutually-intelligible. There should be no reason to assume that children's morphological productivity would vary greatly between them.

[4]Concerning Quechua orthography, it is important to emphasize two things. First, there *is* a standardized, written form of South Bolivian Quechua. For example, students who study Quechua at the university level learn to read and write in the language and *some* textbooks for school-aged children contain some words written in Quechua. Second, in my fieldwork, I have found that speakers certainly *can* write in Quechua, because the orthography is relatively transparent and Quechua shares many phonological characteristics with Spanish (e.g. five vowel system). But I have found this to be uncommon in the communities where I work.

The second prediction of this study is that children will coarticulate more than adults at morpheme boundaries meaning that overall, adults will show a larger differential between within- and across-morpheme coarticulation than children.

If older speakers' phonetic realization of the exact same [ap] or [am] sequence differs within a morpheme versus across a boundary, then we can infer that the adults are breaking words down at constituent boundaries: they have learned to parse speech segmentally and store it morphophonemically. The relative lack of distinction in the children's speech, however, infers that children may be accessing complex words in larger, inflected chunks whose internal structure the children may not have analyzed.

The predicted outcomes of this experiment do not suggest that Quechua-speaking children are morphologically unproductive. Instead of suggesting a lack of productivity, the predictions in the experimental study reflect recent proposals in child speech and language research that children may initially analyze language differently than adults and, as a result, children may store language more holistically (Lieven et al. 1997; Redford 2019; Davis and Redford 2019). Such proposals argue that children do not always recognize complex forms as having an internal structure, and thus do not deconstruct words into morphemes, or morphemes into phones. Instead, children may be more prone to represent and access language in chunks, particularly frequently-occurring combinations or phonotactically-probable "subwords." You could imagine that chunks in English, like *Immagonna* or *lookatit*, could be stored in this way, in addition to their deconstructed forms (individual words, morphemes, and phonemes). Thus, child language may be redundant, with representation at multiple levels in the grammar.

This approach to speech and language development may surprise some audiences – after all the birth of modern linguistics marked a decided turn away from memorized, holistic chunks (Chomsky 1959; Skinner 1957). This is because Behaviorism made untenable demands on psycholinguistic representation and instigated logical fallacies such as the Poverty of the Stimulus. However, abstraction is not undone by the "chunking" approach to language acquisition outlined here. In a chunking approach, children learn language chunks such as words and syllables which then individuate into smaller linguistic units with time. This chunking approach could postulate that abstraction is postponed until later in development: abstraction of morphemes and phonemes could emerge piecemeal via diffusion in the child's lexicon and grammar. Alternatively, a chunking approach could argue that children (and adults) have redundant representations at multiple levels (lexical, syllabic, phonemic) (Arnon and Christiansen 2017; Arnon and Cohen Priva 2013; Arnon 2010; Bannard and Matthews 2008). Some recent exemplar-based models of adult phonology propose exactly this organization, where both speaker-specific and abstracted categories co-exist with the latter emerging out of the former (Pierrehumbert 2016).

So the proposal that children do not reliably analyze the internal structure of word forms, thus storing some forms more holistically, does not claim that children are not morphologically productive language users. It also does not claim that speakers do not develop abstract language. Finally, a concept of more holistic storage in childhood does not argue that speakers must store every experienced multisyllabic/multiword sequence in perpetuity.

More holistic storage *does* argue that linguistic representations are probabilistic, emerge with experience, and exist at multiple levels in the grammar. If we ignore biases like literacy and linguistic concepts like words and phonemes, it actually makes sense that young children, who are not exposed to written language and have limited experience with spoken language, might analyze and store words differently than adults. This chapter empirically tests this theory.

3.3 Validation study

The primary objective of the validation study is to test two relatively novel acoustic measures of coarticulation. This study has been published in its final form in Cychosz et al. (2019).

Calculations

In the current study, coarticulation was quantified using two automatically-extracted acoustic measures, one spectral and one temporal, that were first reported in Gerosa et al. (2006). In that study, coarticulation was measured in a cohort of children (N=40), aged 5, 7, 9, 11, 13, 15, and 17 years. The current study attempts to validate the measures proposed in Gerosa et al. (2006) in a single cohort of four-year-old children. In the current study, both measures were made using custom Python scripts running Librosa functions (McFee et al. 2015).

The spectral coarticulation measure is the difference between the averaged Mel-frequency log magnitude spectra from two adjacent phones. To compute this, the acoustic signal was first downsampled to 12 kHz. Then, each phone was segmented into 25.6ms frames, with a 10ms step. All the Mel-frequency spectral vectors from a given phone were then averaged. Finally, the Euclidean distance between the averaged Mel spectral vector for the phones was measured in the relevant biphone sequences for each word as displayed in Eq. 3.1a:

$$d_{sa} = \sqrt{\Sigma(\bar{x}_s - \bar{x}_a)^2} \tag{3.1a}$$

where d_{sa} is the Euclidean distance between segments /s/ and /a/ in the biphone sequence /sa/ (for example), and $\bar{x}_s$ and $\bar{x}_a$ are the averaged Mel spectral vectors of each segment. Unlike Gerosa et al. (2006), who computed the averaged MFCC vector from each adjacent phone, a discrete cosine transformation was not applied to the Mel-frequency spectra to compute MFCCs because the compression of Mel spectra to MFCC can result in the loss of acoustic information.

Gerosa et al.'s temporal coarticulation measure was also implemented. This measure reflects the duration of the transition between adjacent phones. The acoustic signal was again downsampled to 12 kHz with each phone was segmented into 25.6ms frames with a 10ms step. The region of the transition duration was determined dynamically, based on acoustic difference between a given Mel-frequency spectral frame and the average spectrum of each phone. As in Gerosa et al. (2006), this first required that computing a function for the distance between each sampled spectrum and the average Mel-frequency spectrum for each phone in the sequence as shown in Eq. 3.2a:

$$f_{sa}(i) = d(\bar{x}_s, x_i) - d(\bar{x}_a, x_i) \tag{3.2a}$$

where $\overline{\mathrm{x}}_{\mathrm{s}}$ is the average Mel spectral vector for /s/, and $\overline{\mathrm{x}}_{\mathrm{a}}$ is the same for /a/. i is the spectral vector to be compared to the average spectrum (iteratively sampled over the phone), and d denotes the distance between the single spectral vector and the averaged spectral vector for that phone. The function *f(i)* centers around zero and is negative over the first segment and positive over the second segment in the biphone sequence.

The number of frames where *f(i)* is between an upper and lower bound is n and $n \cdot t$ is the duration of the transition in milliseconds, with step size t=10ms. The transition region, determined by the upper and lower bounds, is set as the portion of *f(i)* that spanned the middle 80% of the range *f(i)*. Transition duration was then scaled by the duration of the CV sequence dur_{sa} to compute the relative transition duration between phones in the CV sequence as shown in Eq. 3.3a.

$$\frac{n \cdot t}{dur_{sa}} \tag{3.3a}$$

Hypotheses

Two important predictions regarding coarticulation in CV sequences are made:

1. Place of vowel articulation: In fricative-vowel sequences, fricative segments consistently show assimilatory effects to the following vowel. For example, in anticipation of the lip rounding required for [u], peak fricative frequencies are lower in the sequences [su] and [ʃu] than [si] and [ʃi] (Soli 1981), reflecting anticipation of the upcoming round vowel. Furthermore, larger distances traveled along the palate during the articulation of a CV sequence result in increased coarticulatory influence of one phone on another when compared to segments that are articulated in the same region. For two biphone sequences of equal duration, speakers may be more capable of differentiating the fricative and vowel in [sæ] than in [su] due to the time constraints of articulating both segments in a given window.

   Anticipatory coarticulation in fricative-vowel sequences is one of the most well-documented cases of coarticulatory influence in the literature: fricatives articulated at or behind the alveolar ridge consistently demonstrate anticipatory coarticulation effects that vary by vocalic context, particularly before front and round vowels, in a variety of languages (Hoole et al. 1993; Mann and Repp 1980; Soli 1981; Whalen 1981).

Fricatives articulated at the alveolar ridge show more evidence of the upcoming vowel when that vowel is both front and round than when the vowel is not front and round.

For the current measurements, I predict a smaller Euclidean distance between adjacent phones in [su] than [sæ], reflecting the greater influence of [u] on [s] than [æ] on [s]. In addition, I predict that sequences requiring a lingual transition from the palatal ridge to the velar region, such as [su], will have a longer transition duration than segments such as [sæ], reflecting the increased movement required to articulate [s] and [u].

2. Manner of articulation: Consonant manner is a predictor of coarticulatory patterning with some manners demonstrating more COARTICULATORY RESISTANCE, or restraint from the coarticulatory influence of an adjacent segment, than others (Recasens et al. 1997; Recasens and Espinosa 2009b). Coarticulatory resistance decreases with lingual contact. Supra-glottal fricatives, for example, have a smaller surface contact area at the palate than glides which explains why anterior fricatives resist the influence of adjacent segments better than labiovelars or vowel-like rhotics (Recasens 1985).

   The relationship between coarticulatory resistance and lingual contact also interacts by speech articulator with segments realized with more sluggish articulators, such as the tongue dorsum, unable to resist coarticulatory influence as well as consonants articulated with the tongue blade (Mooshammer et al. 2006; Recasens and Espinosa 2009b).

   I attempt to replicate these patterns of coarticulatory resistance in a hierarchy of sounds with different amounts of lingual contact and tongue dorsum involvement: alveolar fricatives > alveopalatal affricates > labiovelar glides. In this hierarchy, alveolar fricatives should show maximal coarticulatory resistance because articulation 1) involves the tongue tip (minimal palatal contact and tongue dorsum uninvolved) and 2) is highly constrained (to generate noisy turbulence). Alveopalatal affricates should exhibit relatively less resistance because tongue position is more flexible and lingual contact more fleeting (i.e. could be articulated at several points along the horizontal dimension to similar acoustic effect).

   Finally, labiovelar glides should show the least resistance because of a large area of lingual contact and articulation with a sluggish articulator (dorsum). This order by manner of articulation should translate to a smaller Euclidean distance between glide-vowel sequences than affricate-vowel sequences and smaller distance between affricate-vowels than fricative-vowels. For the temporal measure, I anticipate that glide-vowel sequences will show longer transition duration than affricate-vowel and fricative-vowel, in that order.

To validate the novel temporal and spectral coarticulatory measures, I replicated these well-known coarticulatory patterns in a corpus of four-year-old children's speech recordings.

Methods

The corpus

Data for this study come from 103 four-year-old children (56 girls, 47 boys; range=3;3-4;4 [years;months], μ=3;5, σ= 0;3). All children were monolingual speakers of English. Children were participating in a longitudinal study of lexical and phonological development (see Mahr and Edwards 2018 for further detail). I report on data collected at the second of three timepoints. Each child passed a hearing screening in at least one ear at 25dB for 1000, 2000, and 4000Hz. Ninety (87.4%) of the children had normal speech and hearing development, per parental self-report. The 13 remaining children were identified as late talkers by their caregivers. However, the late talkers' scores on the series of language assessment tasks administered did not differ significantly from the remaining children. Consequently, data from all children were used in the current analysis.

For the data collection phase, each child completed a word repetition task where the participant repeated words after a model speaker. Children repeated a total of 94 words (including 4 training/practice items). All words contained a consonant-vowel (CV) sequence in word-initial position and were bisyllabic with penultimate stress. To ensure that children of this age would recognize words used in the task, words were chosen from the MacArthur Bates Communicative Development Inventory (Fenson et al. 2007), the Peabody Picture Vocabulary Test-4 (Dunn and Dunn 2007), and other sources (e.g. Morrison et al. 1997).

Here subset of 5 of the original test items is analyzed (Table 3.1). The first group of words, *sandwich* and *suitcase*, evaluates the place of articulation hypothesis by measuring the anticipatory coarticulation of [s] before [ae] versus [u]. The second group of words, *sister*, *chicken*, and *window*, tests manner of articulation by measuring the coarticulation between CV segments where the manner of consonant articulation varies but the vowel is constant.

Table 3.1: Stimuli used in validation experiments.

Word	Transcription	CV Sequence	Hypothesis	Correctly Produced
sandwich	[sændwɪʧ]	[sæ]	Place of articulation	N=73 (70.87%)
suitcase	[sutkes]	[su]	Place of articulation	74 (71.84)
sister	[sɪstɚ]	[sɪ]	Manner of articulation	86 (83.50)
chicken	[ʧɪkən]	[ʧɪ]	Manner of articulation	74 (71.84)
window	[wɪndo]	[wɪ]	Manner of articulation	89 (86.41)

A young female speaker of American English provided the recordings for the word stimuli. Recording prompts were digitized at a frequency of 44,100 Hz using a Marantz PMD671 solid-state recorder. Amplitude was normalized between words.

Each participant was guided through the repetition task by at least two experimenters. First, the child was seated in front of a computer screen and presented with a photo while the accompanying word played over external speakers. The child was then instructed to repeat the word. After each trial, the experimenter manually advanced to the subsequent trial. Stimuli were presented randomly with E-prime software (Schneider et al. 2012). The task lasted approximately fifteen minutes.

Segmentation

The production accuracy of each CV sequence was scored. Accuracy scoring was conducted offline in a feature-based system by a trained phonetician who is a native speaker of American English. Child participants had to produce the correct consonant voicing, manner of phone articulation, and place of articulation. Children additionally had to produce the correct height, length, and backness for the vowel and repeat the word's prosodic structure correctly (number of syllables, consonant in correct position, and vowel in correct position). Scoring was conducted auditorily and by reviewing the acoustic waveform. To ensure scoring accuracy, a second rater, also a trained phonetician and native speaker of American English, scored a 10% subset of the original words. An intraclass correlation (ICC) statistic assessed inter-rater agreement. The intraclass correlation between raters was 0.881, which was significantly greater than chance ($F(374,375)=15.9$, $p<.001$, 95% CI=0.86, 0.90). In the event of disagreement between the annotators, the word annotation made by the primary researcher was used. Only CV sequences that were produced correctly underwent acoustic analysis. Acoustic analysis and accuracy scoring were conducted on separate occasions for different research programs. The number of tokens for each word used in the current study is listed in Table 3.1.

The words that were repeated correctly then underwent acoustic analysis. Each correct CV sequence was manually transcribed in a Praat TextGrid (Boersma and Weenik 2019) by a native speaker of American English who is a trained phonetician. The audio files were aligned using the visual representation from the waveform and spectrogram in addition to auditory analysis. As coarticulation measures are highly dependent upon segmentation decisions, a number of steps were taken to standardize alignment. Alignment conventions for each CV sequence were established. The start of affricate/fricative-vowels corresponded to the onset of high-frequency energy in the spectrogram. For affricate/fricative-vowel sequences, the start of the vowel corresponded to the onset of periodicity in the waveform and formant structure in the spectrogram. These criteria were sufficient to demarcate all affricate/fricatives from vowels. Delimiting glide-vowel sequences was more gradient: a steady state formant delimited glide offset and vowel onset. Transcribers were encouraged not to rely on auditory analysis for glide-vowel segmentation decisions, which is notoriously problematic (McAllister Byun et al. 2016a). In the rare event that a steady-state formant could not be identified, 50% of the sequence was assigned to the consonant and 50% to the vowel. A second transcriber, blind to the validation experiment objectives, independently aligned a 10% subset of the words. The difference between the coders' average consonant duration was 2 ms and the average difference in vowel duration was 10 ms. Pearson correlations between the coders were significant for consonants: $r=0.96$ $p<.001$, 95% CI=[0.95, 0.96] and vowels: $r=0.87$ $p<.001$, 95% CI=[0.85, 0.89], suggesting high fidelity to the alignment procedure. Despite these efforts, it is important to note that hand-segmentation is often highly subjective.

Results

There are two hypotheses for this validation study. First, it is hypothesized that the two coarticulation measures being validated here replicate place of vowel articulation effects in fricative-vowel sequences. Specifically, the children will coarticulate more (smaller Euclidean distance) between adjacent phones in [su] sequences than [sæ] sequences. The second hypothesis predicts that the coarticulation measures will replicate known manner of articulation effects in coarticulation and will follow a hierarchy of coarticulatory resistance: alveolar fricatives > alveopalatal affricates > labiovelar glides. The children will coarticulate the most in those sequences that exhibit the least amount of resistance (labiovelar glides) and the least in those sequences that exhibit the most amount of resistance (alveolar fricatives).

To evaluate the place of vowel articulation hypothesis, two mixed effects linear regression models were fit using the `lme4` package in the R computing environment (Bates et al. 2015). Each model included **Speaker** as a random effect. One model predicted the temporal coarticulatory measure and the other the spectral. In both cases, the effect of **Context** significantly improved baseline model fit. Specifically, for the spectral model, there is a smaller distance between phones in the sequence [su] than [sæ] (β=-1.56, t=-3.31, p=.002), indicating greater coarticulation between [s] and [u] than [s] and [æ] as children anticipate the roundedness of [u] (Figure 3.1). In the temporal model, the transition duration between [s] and [u] is longer than [s] and [æ] (β=1.08, t=1.99, p=.05), again indicating greater coarticulation between the segments in [su]. Thus, both the temporal and spectral measures capture coarticulatory differences by place of articulation in fricative-vowel sequences in the vertical dimension (i.e. backness) and by vowel quality (roundedness), but the spectral model may be a more reliable indicator of anticipatory coarticulation for these segments.

Next, I evaluate the hypothesis that the coarticulatory measures should predict coarticulatory differences by consonant manner in CV sequences, after controlling for vowel identity. Two mixed effects linear regression models were again fit as before with Speaker as a random effect. The fixed effect Consonant Manner improved both model fits. Specifically, in the spectral model, [sɪ] reliably differed from [ʧɪ] (β=-2.67, t=-4.74, p<.001) and [wɪ] (β=-3.19, t=-5.93, p<.001) – [s] and [ɪ] were less acoustically overlapped than the segments in [ʧɪ] or [wɪ], suggesting less coarticulation. However, a post-hoc test with [ʧɪ] as the reference level demonstrated that [ʧɪ] did not differ significantly from [wɪ] (p=.78). Still, the trend by consonant manner follows the anticipated direction: there was a larger acoustic distance between segments in [ʧɪ] (median=7.39, σ=4.19) than [wɪ] (median=6.72, σ=2.00) suggesting

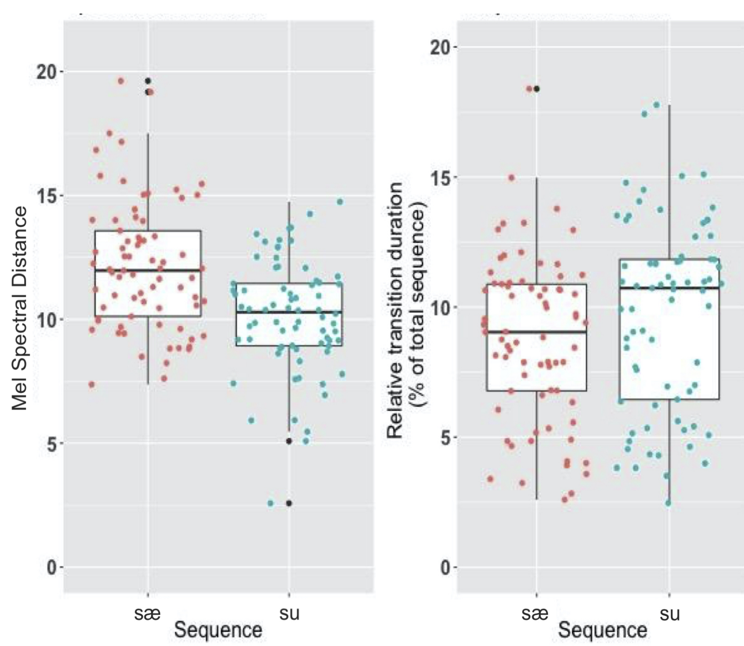


Figure 3.1: Fricative-vowel coarticulation by place of vowel articulation. Computed temporally (R) and spectrally (L).

less coarticulation in [ʧɪ] than [wɪ] (Figure 3.2). For the temporal model, [sɪ] reliably differed from [ʧɪ] (β=1.98, t=3.42, p<.001) and [wɪ] (β=7.71, t=14.04, p<.001). Another post-hoc test also demonstrated that along the temporal dimension, [ʧɪ] differed significantly from [wɪ] (β=5.56, t=5.71, p<.001). The transition between segments in [wɪ] was longer than the transition between segments in [ʧɪ]. These results suggest that both the temporal and spectral coarticulation measures reliably capture known coarticulatory differences by consonant manner, after controlling for vocalic environment.

Validation study discussion

In this validation study, I used two relatively novel acoustic measures of coarticulation, one temporal and one spectral, to diagnose previous acoustic correlates of segmental coarticulation. Through a series of comparisons, I demonstrated that both of the tested acoustic measurements were generally robust enough to capture known patterns of coarticulation.

I first tested the hypothesis that the coarticulation measures would capture differences in fricative-vowel coarticulation by place of vowel articulation and vowel quality. Specifically, speakers are known to anticipate vowel quality, especially roundedness, in fricative-vowel sequences, and should exhibit increased coarticulation in sequences such as [su]. Furthermore, speakers should show anticipate the upcoming vowel in sequences with segments that differ

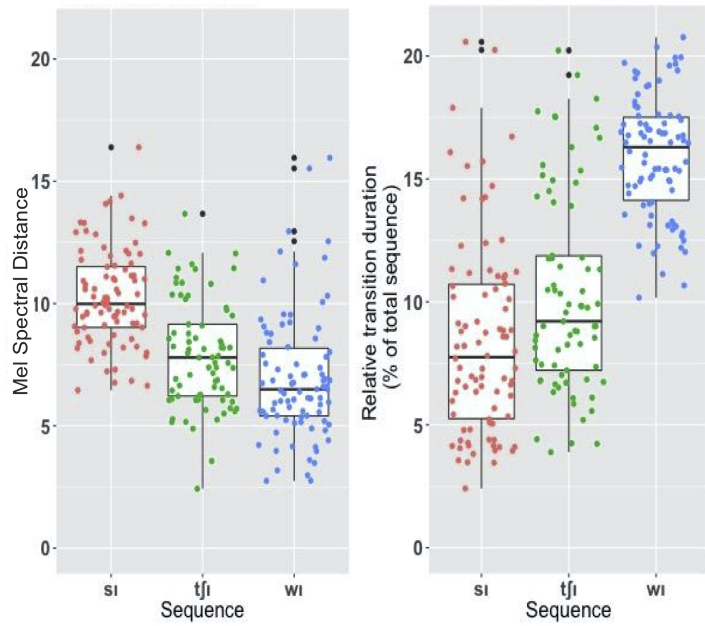


Figure 3.2: CV coarticulation by consonant manner. Computed temporally (R) and spectrally (L).

in place of articulation, such as [su], than with segments that do not, such as [sæ], because the articulation of the former requires a transition from a lingual articulation at the alveolar ridge to an articulation towards the velum.

The current measures captured both of these coarticulatory patterns, though the spectral measure was more reliable. I found that speakers showed more acoustic overlap of phones, and a longer transition duration between phones, in the sequence [su] than the sequence [sæ], replicating known coarticulatory patterns by place of vowel articulation and vowel quality (Hoole et al. 1993; Mann and Repp 1980; Soli 1981; Whalen 1981). However, acoustic measures of coarticulation are imperfect and acoustic similarity/transition duration does not necessarily indicate greater coarticulation. For example, if a speaker were already halfway to hitting a vowel target at the beginning of a VC sequence, then their transition to the following consonant could be faster than a speaker who did not start at the same halfway point. Yet acoustic measures might say that these speakers "coarticulated" in different amounts, without acknowledging the underlying reasons for their dissimilarity.

Next, I attempted to capture differences in coarticulation by manner of consonant articulation. Consonants whose manner requires less lingual contact, particularly when realized with the tongue blade such as alveopalatal and alveolar stops, are able to resist coarticulation with adjacent segments more than consonants whose manner requires more lingual contact with the sluggish dorsum, such as rhotics and labiovelar glides (Recasens and Espinosa 2009a;

Recasens et al. 1997). I replicated these patterns by manner using both coarticulation measures. As predicted, speakers coarticulated less in sequences with more resistant consonants in the following hierarchy: [sɪ] < [ʧɪ] < [wɪ].

These coarticulatory measures are important tools for speech research, particularly developmental. Both measures have broad applicability for a variety of consonant types – unlike center of gravity they are not limited to fricatives or sonorant sounds. Furthermore, the measures are relatively immune to the many challenges that children's voices, generally breathy with high fundamental frequencies, bring to traditional acoustic analyses. Finally, these measurements can be made automatically, over small samples of speech, without equipment more specialized than a portable audio recorder. As a result, these measures may have broad applications for clinical populations or understudied groups. Researchers and clinicians can use the measures as an index of speech maturity in children or as a more fine-grained way to measure speech disfluencies in clinical populations on the basis of small samples collected in the home or clinic (however it is stressed that results are heavily dependent upon text to audio alignment choices which can still be time-consuming). Field linguists and clinicians working in under-served communities can use these measures to document speech patterns in populations who cannot feasibly be reached with articulatory apparatuses such as ultrasounds or electromagnetic articulatography. The speed of the measures also evades some of the challenges inherent to articulatory data collection outside of the lab environment or with young children (children are reticent to wear ultrasound stabilization helmets or paste pellets on the tongue for electromagnetic articulatography). Thus, while both articulatory and acoustic measures are valid methods to study child speech development, acoustic tools can broaden the communities that are represented in this research.

Future work in this realm could continue to test these coarticulation measures on additional segments to ensure that they capture other coarticulatory patterns such as nasality which were not tested here (Beddor et al. 2018). I also limited the validation to four-year-olds, so it may be important to evaluate the performance of these measures in other age groups. It is also important to note that the word repetition employed here could have resulted in phonetic convergence between the children and the adult model speaker, though hopefully the presentation of test items in a random order mitigated any effect. Furthermore, future work explicitly contrasting formant-based measurements with those outlined here is warranted. Finally, though anticipatory coarticulation is generally analyzed in adjacent segments and considered a short-distance phenomenon, long-distance coarticulation may exhibit different developmental patterns (Barbier et al. 2013; Goffman et al. 2008). The current measurements could also efficiently shed light on this outstanding question.

3.4 Experimental study

Hypotheses

The primary objective of the experimental study is to compare adult and child Quechua speakers' coarticulatory patterns within and between morpheme boundaries.

1. Do adult Quechua speakers coarticulate more between biphone sequences within morpheme boundaries (e.g. p**ap**a 'potato') versus across morpheme boundaries (e.g. pap**a-p**i 'potato-LOC')?

I predict that adult speakers will coarticulate more between the same biphone sequence *within* morphemes than *across* morphemes. This speech pattern between morphological environments may reflect the decomposition of complex words into component parts, as previous work on both adult (Cho 2001; Tomaschek et al. *under review*) and child coarticulation (Song et al. 2013b) has demonstrated.

2. Do child speakers of Quechua, aged 5;0-10;0, differentiate their coarticulatory patterns across versus within morpheme boundaries? If so, does this change throughout development in this cross-sectional sample?

On the other hand, I predict that child Quechua speakers will not necessarily differentiate their coarticulatory patterns between the two morphological environments. Instead, children may coarticulate equally within and across morphemes. This would suggest, as previous work on child speech has suggested (e.g. Noiray et al. 2019a; Redford 2018; Zharkova et al. 2011), that increased coarticulation reflects more holistic representations. Specifically, the children in the current study may coarticulate similarly in the two morphological environments. If this occurs, it may indicate that they are not always breaking down morphologically complex words in the same manner as adults, but are instead storing items more holistically (Redford 2018).

However, because I hypothesize that adults distinguish between the word environments, I should also anticipate that children's coarticulatory patterns will change over the course of development (age 5;0-10;0). Thus, older children will likewise *begin* to coarticulate differently between the two morphological environments. The timing of this development is unclear as there is limited work on morphophonetic interplays in child speech (see 3.2), but overall we should see increasing ability to distinguish between the environments by age.

In addition to measuring coarticulation in the two morphological environments, the duration of the biphone sequences will also be measured. There are acknowledged interactions between duration and coarticulation; for example, speakers tend to coarticulate more when they speak faster (Gay 1981; Matthies et al. 2001). Thus, measuring how coarticulation interacts with speaking rate could be an important component to the speech patterns evaluated here. Furthermore, given that adults speak faster than children, it may be especially important to measure, and control for, the duration of the biphone sequences when measuring differences in coarticulation between different age groups.

Methods

Participants

Fifty-one children, aged 5;0-10;11, and ten female adults (adult μ_{age}=23, σ=5.46, three did not report) participated in this study. Children's distribution by age was as follows: 10 five-year-olds (μ=5;7, σ=0;4; 6 girls, 4 boys, one did not report), 10 six-year-olds (μ=6;5, σ=0;2; 5 girls, 8 boys), 13 seven-year-olds (μ=7;7, σ=0;4; 3 girls, 8 boys), 8 eight-year-olds (μ=8;8, σ=0;2; 4 girls, 5 boys, one did not report), 5 nine-year-olds (μ=9;4, σ=0;3; 2 girls, 3 boys, two did not report), and 5 ten-year-olds (μ=10;8, σ=0;4; 1 girl, 4 boys, two did not report). The recording for one of the adult participants contained significant wind interference and was removed from analysis leaving 9 adult participants in the final sample. All participants were bilingual Spanish-Quechua speakers living in or around a mid-size town in southern Bolivia.

The child participants were either recruited at the local school where I volunteered (n=13), as described in Chapter 2, or through personal contacts in the surrounding communities (n=38). The adult participants were recruited through local contacts. These are a subset of the same participants described in chapter 2. Additionally, the four-year-olds did not participate in this study.

Most children had typical speech and hearing development, per parental/teacher self-report. The caregivers of 3 children (2 seven-year-olds, 1 five-year-old) stated that their child was late to begin talking.[5] Note that these communities are medically under-served so some language delays/impairments may go unreported. Additionally, 3 children had lost one or more of their top/bottom front teeth at the time of recording.[6] See Chapter 2 for additional details on the children's speech and language development and hearing status.

Information on the central caregiver's education level (usually the mother, occasionally the grandmother) was collected from those families recruited from the surrounding community, but not those recruited at the school.[7] There were 7 sibling pairs (no twins), and 1 three-sibling pair, in the child sample resulting in 43 unique caregivers in the sample. For the 31 caregivers of children recruited from the surrounding community, the caregivers' education levels were: 17 caregivers (59% of caregivers from the community) completed some primary school (less than six years of education), 5 (17%) completed primary school (6 years

[5]Late talker status was not collected from the participants recruited from the school.

[6]I report this because the presence of front teeth could have notable consequences for speech acoustics (e.g. anterior fricatives). This information is not typically reported in speech development research, but arguably should be.

[7]Maternal education level is the usual proxy for socioeconomic status in language development research (Hoff 2003).

of education), 3 (10%) completed the equivalent of a middle school (10 years of education), n=3 (10%) completed secondary/high school (13 years of education) and 2 (7%) had not received any formal schooling. Two caregivers did not report.

A coarse estimation of the central caregiver's level of Spanish-Quechua bilingualism was also collected from those families recruited from the surrounding community: 6 (21% of caregivers from the community) were monolingual Quechua speakers, 5 (17%) were Quechua-dominant but spoke/understood some Spanish, 17 (59%) were bilingual Quechua-Spanish speakers, and 1 did not report. See chapter 2 for details on mother's education level and knowledge of Spanish as correlates of socioeconomic status in Bolivia.

Tasks

The child participants completed four tasks, all prompted with pictures, in the following order: 1) real word repetition, including a morphological extension component (to be explained in the following sections), 2) Quechua nonword repetition, 3) Spanish nonword repetition, and 4) additional real word repetition with morphological extension. For all tasks, children repeated the real words or nonwords after a pre-recorded model speaker. Nonword repetition tasks are not discussed in this dissertation. The relevant tasks are explained in more detail in the following sections.

The adult participants only completed the two real word repetition tasks because even the eldest children approached ceiling on the nonword repetition tasks. Because much of the children's data was collected during the school day, the entire testing procedure had to be relatively short and executable in one sitting. The entire experimental procedure took approximately 30-40 minutes per child and 20 minutes per adult. Compensation for participation is as described in chapter 2.

Stimuli

The real word repetition tasks consisted of 56 high-frequency Quechua nouns (plus 6 training trials for 62 total lexical items) that are familiar to children learning Spanish and Quechua in southern Bolivia (stimuli listed in Appendix B). Neither Bolivian Spanish nor any Quechuan language has an equivalent to the *Macarthur Bates Communicative Development Inventory* (Fenson et al. 2007), which reports stages of age-normed vocabulary development. Nor do these languages have any large, transcribed child-directed speech corpus from which to infer vocabulary development. For these reasons, children's knowledge of the test items was confirmed via a pre-test that demonstrated that children as young as 3;0 recognized all items (Cychosz 2019). Female caregivers also confirmed that children as young as 3;0 should recognize the items.

In addition to selecting high-frequency lexical items, likely to be recognized by the children and easily represented in a photo, these particular lexical stimuli were also chosen because they contained the sequence [ap] or [am] within a morpheme (e.g. p**ap**a 'potato') or crossing a morpheme boundary (e.g. thap**a-p**i 'prairie-LOC'). And to control for acoustic correlates of stress, the phone [a] in the biphone sequence had to fall in the syllable carrying primary stress.[8,9] Thus, of the original 56 real word test items, this study measured coarticulation on a subset that contained the biphone sequence [ap] (n=23) or [am] (n=23) (see 3.16 for list of stimuli that elicited [ap] and Appendix C for the stimuli that elicited [am]).[10] The experimental hypotheses remain the same for both the [ap] and [am] biphone sequences. However, given the acoustic measure of coarticulation employed here - spectral distance - overall there is less "coarticulation" between the phones in [ap] than the phones in [am]. This is due to the increased acoustic similarity of the phones in [am] (voiced, sonorous).

The VC sequences [ap] and [am] were chosen to examine coarticulatory effects for several important reasons. First, instead of CV sequences, VC sequences were chosen because all Quechua nominal case-marking suffixes are consonant-initial (e.g. '*-q*' GENITIVE, '*-manta*' ABLATIVE). Consequently, it is not possible to elicit a CV sequence that crosses a noun-case marker boundary in Quechua. Then, of the possible VC sequences, [ap] and [am] were chosen because coarticulatory measures are highly dependent upon segmentation decisions. The acoustic delimitation between vowels and voiceless stops/vowels and nasals is relatively obvious and not subjective.[11]

The two suffixes *-pi* and *-man* (pronounced [maŋ]) were also chosen for several specific reasons. First, nominal case markers were chosen because nouns are easier to represent in picture prompts than derived word forms (e.g. puñu-y 'to sleep' puñu-chi-y 'to make (one) sleep') or verb conjugations. Second, nouns are grammatical in Quechua with just one suffix. Some conjugated verbs require multiple suffixes (see 'sleep' example above), which would make elicitation and tight control of the experimental stimuli more difficult. In using nominal suffixes, elicitation could be isolated to a single stem+suffix combination. Finally, the locative and allative markers were used because, absent a large corpus of child-directed Quechua speech, it is reasonable to assume that the locative *-pi* and allative *-man* on high-frequency nouns, such as those elicited, will be relatively frequent in a child's input.

Given all of these considerations - the need for high-frequency, child-friendly words, the correct prosodic environment, frequent suffixes, and segments that were easily segmented - it was challenging to identify lexical items for use in the experiment. Still, with (n=35) unique items in the across morpheme boundary condition and (n=11) unique items in the within boundary condition, this study uses more distinct lexical items than most previous studies of morphological effects on speech production in children or adults (Lee-Kim et al, 2013; Song et al., 2013a).

[8] Quechua is canonically open-syllabic, so all VC syllables transcend syllable boundaries.

[9] The only exception to this was the item *ham'piri* 'healer', and its inflected form *hampi'ri-pi* 'healer-LOC', where the [am] sequence does not coincide with primary stress. This item was included because it was difficult to find sufficient items for the within morpheme condition that adhered to the criterion of frequency, recognition by children, etc.

[10] Some lexical items contained both within and across stimuli (e.g. [am] and [ap] in ll**ama-p**i 'llama-LOC').

[11] Much previous work on child coarticulation has studied fricative-vowel sequences (e.g. Zharkova et al., 2011). This was not possible in the current study as there are no fricative-initial nominal case markers in Quechua.

Table 3.2: Real word repetition stimuli to elicit [ap]

Real word*	Translation	Morpheme environment†
chi'**ta-pi**	'sheep-LOC'	across
cu'**ca-p**i	'coca (leaves)-LOC'	across
hatunma'**ma-p**i	'grandma-LOC'	across
imi'**lla-pi**	'girl-LOC'	across
juk'u'ch**a-p**i	'mouse-LOC'	across
lla'm**a-pi**	'llama-LOC'	across
lla'p**a-pi**	'lightening-LOC'	across
ma'm**a-p**i	'mom-LOC'	across
pam'p**a-p**i	'prairie-LOC'	across
pa'p**a-p**i	'potato-LOC'	across
q'a'p**a-pi**	'palm of hand-LOC'	across
sun'kh**a-p**i	'beard-LOC'	across
t'i'k**a-p**i	'flower-LOC'	across
tha'p**a-pi**	'nest-LOC'	across
uhu't'**a-p**i	'sandal-LOC'	across
wa'k**a-pi**	'cow-LOC'	across
wall'p**a-p**i	'chicken-LOC'	across
wa'w**a-pi**	'baby/child-LOC'	across
'p**ap**a	'potato'	within
'll**ap**a	'lightening'	within
'**ap**i	'corn/citrus drink'	within
'th**ap**a	'nest'	within
'q'**ap**a	'palm of hand'	within

* ' indicates stress, ' indicates ejective

† Each "across" item additionally inflected with -*man* (ALLATIVE) (See Appendix C).

The real word stimuli came from recordings of an adult female bilingual Quechua-Spanish speaker. These recordings were digitized at a sampling frequency of 44.1 kHz using a portable Zoom H1 Handy Recorder. Stimuli were normed for amplitude between words, but not duration, since some words had ejectives, fricatives, etc. that are temporally longer. The real word picture stimuli were color photographs of the objects.

Children in these communities have limited exposure to technology (some mothers have flip phones but most children are unfamiliar with computing devices). Consequently, instead of presenting each picture stimulus on a screen, which could have been culturally inappropriate, pictures were presented on individual pages clipped into an 11 x 12.4" plastic binder. For this reason, the words were not entirely randomized for each participant.

Instead, two different randomized lists were created with approximately half of the children and adults completing the first list and half completing the second. Since there were more across-morpheme stimuli than within-morpheme stimuli (it was much more difficult to find stimuli for the within-morpheme condition), participants repeated the within-morpheme stimuli three times in the experiment and the across-morpheme stimuli two times. Repetitions of the same stimulus were always separated by at least two different stimuli and were presented with a novel photo of the item each time.

Data collection

For the experimental phase, participants were seated on the ground or on a stool, side-by-side with the experimenter. Audio stimuli were were played for the experimenter and participant from an iTunes playlist run on an iPhone 6. Each participant wore AKG K240 binaural studio headphones and the experimenter wore Apple earpods to follow along with the experiment; both headphones were connected to the iPhone with a Belkin headphone splitter. See chapter 2 for further details on the setting of data collection.

For data collection, the participant first heard the audio stimulus (a bare noun) and was simultaneously presented with the accompanying photo in the binder. The participant was instructed to simply repeat the bare noun after the model speaker. The participant was then instructed to produce the word again in inflected form. In this way, the researcher could be confident that the children independently knew how to inflect each of the nouns with the tested suffixes and were not simply copying a prompt. For the inflected form, the two morphemes described in section 3.4 were elicited. The locative marker -*pi* was elicited in the first real word repetition task and the allative marker -*man* was elicited in the second task. Ideally, all of the participants' productions would have been spontaneous instead of repeated. However, in a previous version of a similar word elicitation task, with different children, I found that the youngest children frequently became too nervous and hesitant to follow the task when not prompted with the word (Cychosz 2019). Elicited imitation paired with a visual stimulus is also the method used in the previous study on coarticulatory effects within and between morpheme boundaries (Song et al. 2013b).

For the children, the inflected forms were elicited using a large plastic toy insect. For the locative marker -*pi*, the toy insect was placed on top of the picture stimulus and the child was prompted, "Where is the bug?" In response, the child produced the word with the correct suffixal carrier, e.g. *llama-pi (kasan)* (llama-LOC COP-3PS, "(It is) on the llama."). For the allative marker -*man*, the researcher wiggled and moved the toy insect on the page towards the noun in question and prompted the child, "Where is the bug going to?," to which the child would produce the word with the suffixal carrier, e.g. *llama-man (risan)* (llama-ALL go-3PS, "(It is going) to/towards the llama."). The adult participants were merely told to add the relevant morpheme to each prompted word in a carrier phrase. For the first real word task, the carrier phrase was, "I say in the _____ two times" (*Noqa nini _____-pi iskay kutita*). For the second task, the phrase was, "I say to/towards the _____ two times" (*Noqa* *nini* _____*-man iskay kutita*). The experimenter would then manually advance to the next stimulus item.

Eliciting participant responses in quasi-sentential contexts such as these may disguise the contrasts between inflected and base word forms since "phonetic variation between orthographically-distinct homophones increases when the target homophones are dictated in isolated word list or in contrastive sentences" (Seyfarth et al., 2018:35). The elicitation methods used here thus discourage metalinguistic awareness to the degree possible given the number of lexical items needed for the experiment.

Participant responses were recorded with a portable Zoom H1 Handy Recorder at a 44.1 kHz sampling rate. Children were rewarded with stickers throughout the task and many additionally chose to help the experimenter flip through the pages of the binder.

Data analysis

Each participant's audio file was first aligned to the word level in Praat (Boersma and Weenik 2019). A Quechua forced aligner was trained on all of the participants' data using the Montreal Forced Aligner (McAuliffe et al. 2017) to align to the phone level. The phone-level alignment was hand-checked by one of two trained phoneticians. One of these phoneticians was blind to the hypothesis of the experiment. Alignment was conducted auditorily and by reviewing the associated acoustic waveform and broadband spectrogram in Praat.

Coarticulation measures can be sensitive to alignment decisions, so a number of parameters were set prior to alignment to ensure segmentation reliability. Word-initial plosive, affricate, and ejective onset corresponded to the burst. The start of the vowel [a] corresponded to the onset of periodicity and formant structure in the waveform and spectrogram. Nasals were differentiated from vowels by the presence of anti-formants in the spectrogram and a reduction in intensity in the waveform. Additional parameters were set (e.g. for glides, rhotics) but are irrelevant for the segments under analysis in the current study. See chapter 2 for additional segmentation details and the report on inter-aligner agreement.

As described in the validation study, much of the child coarticulation literature employs acoustic measures of coarticulation such as center of gravity for fricatives or formant-based measurements (e.g. transitions or spectral peaks) for vowels. However, measurements of child formants are notoriously difficult to obtain. The following results section reports coarticulation as the the euclidean distance between Mel-frequency log spectra averaged over the middle third of adjacent phones (the spectral measure of coarticulation described in the validation study (Gerosa et al. 2006)).

Results

The primary research question in this experiment asks how child and adult Quechua speakers coarticulate between and within morphemes. The results begin with descriptive statistics concerning the amount of coarticulation by age group (children aged 5 through adults) and morphological environment (within versus between morphemes). Additional descriptive statistics outline the duration of the VC sequences by age group and morphological environment. Then, a series of models are fit to predict coarticulation and duration by age and morphological environment. These models are complemented by an analysis highlighting how coarticulation interacts with duration differently in adults and children in the two morphological environments.

All analyses were conducted in the RStudio computing environment (version: 1.2.5019; RStudio Team (2020). Data visualizations were created with `ggplot2` (Wickham 2016). Modeling was conducted using the `lme4` (Bates et al. 2015), `lmerTest` (Kuznetsova et al. 2017), and `glmmTMB` (Brooks et al. 2017) packages and summaries were presented with `papaja` (**?**) and `Stargazer` (Hlavac 2018). Tests of residual normality were conducted using the `normtest` package (Gavrilov and Pusev 2014). The significance of potential model parameters was determined using a combination of log-likelihood comparisons between models, AIC estimations, and p-values procured from model summaries. In all models, continuous predictors were mean-centered to facilitate model interpretation.

Descriptive statistics

Coarticulation

The degree of coarticulation between the VC sequences [ap] and [am] was measured using the spectral distance metric described in the validation study. For this coarticulation metric, coarticulation is quantified as the Euclidean distance between the spectral vectors of two adjacent phones, henceforth the Mel spectral distance. In this outcome measure, a larger spectral distance between phones equates to *less* coarticulation between the phones.

Table 3.3: Mean spectral distance between [a] and [p] by age and word position

	Across boundary		Within boundary	
Age	Spectral_Distance	SD	Spectral_Distance	SD
5	17.42	4.64	19.63	4.56
6	17.62	4.70	19.78	4.59
7	17.04	4.68	19.18	5.90
8	18.50	6.03	21.35	6.76
9	16.21	7.02	17.51	7.79
10	14.92	5.09	15.93	5.07
adult	14.99	3.64	14.51	3.03

Table 3.4: Mean spectral distance between [a] and [m] by age and word position

	Across boundary		Within boundary	
Age	Spectral_Distance	SD	Spectral_Distance	SD
5	7.94	3.13	8.71	3.39
6	8.44	3.08	9.54	4.76
7	8.43	3.11	8.86	3.80
8	9.89	3.99	9.73	3.78
9	7.43	2.64	8.06	3.11
10	8.21	3.43	8.30	4.25
adult	7.44	4.19	7.00	3.34

Table 3.3 shows the mean Mel spectral distance between the segments in [ap] (words inflected with *-pi* for the across morpheme boundary condition) and Table 3.4 shows this for [am] (words inflected with *-man* for the across morpheme boundary condition). Unsurprisingly, there is a larger average spectral distance between the vowel and plosive in [ap] than the vowel and nasal in [am] because the segments in [am] have increased acoustic similarity (sonority, voicing). Next, looking by age group for coarticulation between [ap], it is apparent that the amount of coarticulation between segments increases as children age. This is likely due to the increased speaking rate in the older cohorts and adults, as will become apparent when these results are crossed with sequence duration. There is less within age group variability in [ap] productions in the adult speakers as well (reflected in the smaller SD of the mean for the adults). Variability does not appear to decrease linearly as both the nine and ten-year-old cohorts exhibit larger SDs than the five and six-year-olds. However, this could also reflect the smaller sample sizes in the older cohorts (5 ten-year-olds and 5 nine-year-olds but 10 each in the five- and six-year-old cohorts).

For the coarticulation patterns between segments in [am], adults and children appear to coarticulate similarly, irrespective of age group. There is slightly greater variability in the amount of coarticulation (larger SD) in the adult group. Overall, the difference in developmental coarticulatory patterns between the VC sequence [am] and the sequence [ap] could be due to differences in the *-man* suffix use in Quechua (frequency, productivity). Alternatively, differences between these sequences could be due to the acoustic differences between [ap] and [am] as segments in [am] are more acoustically similar than segments in [ap].

Duration

I next turn to descriptive statistics describing the interaction of coarticulation and sequence duration. Duration could interact with coarticulation, and age, as speakers coarticulate more in fast speech (Gay 1981), and adults speak faster than children (Lee et al. 1999).

Table 3.5: Mean duration of [ap] sequence by age and word position

	Across boundary		Within boundary	
Age	Duration (ms)	SD	Duration (ms)	SD
5	228.7	47	339.9	52
6	242.5	45	334.4	59
7	245.1	57	319.8	60
8	226.9	56	329.3	43
9	216.6	36	312.6	56
10	212.8	50	302.3	63
adult	205.7	47	197.0	59

Table 3.6: Mean duration of [am] sequence by age and word position

	Across boundary		Within boundary	
Age	Duration (ms)	SD	Duration (ms)	SD
5	214.4	45	251.2	66
6	217.9	58	247.8	62
7	209.8	42	244.6	60
8	218.2	63	247.1	53
9	199.8	30	231.8	64
10	194.3	33	239.9	57
adult	175.6	36	173.7	39

Table 3.5 maps average sequence duration of [ap] by age and word position and Table 3.6 does similarly for [am]. Overall, duration of [ap] decreases with age, with adult speakers exhibiting the shortest average duration for [ap] for both word positions. Of note is that the average duration of [ap] sequences within boundaries tended to be longer than the duration of [ap] sequences across boundaries for all of the children. This pattern was reversed in the adult speakers, however, whose average [ap] duration within morpheme boundaries was actually *shorter* than across. This pattern will be revisited in the modeling portion of the Results.

Turning to the sequence [am], the average duration of the [am] sequence in adult speakers was shorter than all of the child speakers; and like [ap], [am] duration also appears to decrease with age. This pattern of [am] duration decreasing with age is consistent for both word positions, across and within morphemes.

Concerning differences by word position, all age groups showed, on average, shorter [am] durations in the across morpheme condition, contrary to the finding that [ap] sequence duration was longer in the across morpheme condition for adults. However, it is important to note the average duration of [am] by word position in adults only differed by approximately 3 ms while the average within morpheme condition in children was approximately 50 ms greater than the across morpheme condition in children. Thus, for both [am] and [ap] sequences, there appears to be a difference in sequence duration by word position for adult and child speakers. In the following section I turn to the modeling of coarticulation before illustrating how degree of coarticulation interacts with duration differently in the adults and children.

modeling interaction of coarticulation and duration

The central research question in this study asked if child and adult Quechua speakers would coarticulate similarly between VC sequences across versus within morpheme boundaries. To answer this question, a series of generalized linear mixed effect models (GLMMs) were fit to predict degree of coarticulation (Mel spectral distance between each V and C). GLMMs were chosen instead of the more common linear mixed effect models due to the non-normality of the residual **VC Sequence duration** (henceforth simply **Sequence Duration**) which was included in all models. (Shapiro tests of kurtosis and skewness for **Sequence duration** indicated that the null hypothesis that the residual's distribution did not differ significantly from a normal distribution could be rejected. Kurtosis t=5.53, $p<.001$ and skewness: t=1.07, $p<.001$ (Shapiro et al. 1968).) Specifically, the response variable **Spectral distance** and the residual **Sequence duration** are both limited to non-negative values (as all VC sequences had some distance between the V and C and all had a duration), with a resultant right skew to the data distribution.

The choice to fit gamma GLMMs, as opposed to log-normalizing **Sequence duration** and fitting linear mixed models, reflects recent suggestions in cognitive psychology to avoid data transformation, even for commonly-transformed variables such as time/duration, in an effort to facilitate between-study comparison (Lo and Andrews 2015). Consequently, gamma GLMMs were fit using a log linking function to appropriately model the skewed, non-Gaussian distribution of the residual.

A GLMM was fit to predict the spectral distance between segments in the VC sequences. Baseline models included random effects of **Participant** and **Word** (models with random slopes of **Participant** by **Word** did not converge, possibly due to the number of repetitions per speaker; see Methods). Model building then began in a forward-testing manner with predictors added in the following order: **Sequence duration**, **VC sequence** ([ap] or [am]), **Age** (adult or child), **Environment** ([within morpheme or between morphemes]), and interactions. The best model fit included the four-variable interaction of **Sequence duration**,

VC sequence, **Age**, and **Environment** (see Appendix D). This four-variable interaction indicates that the coarticulation and duration patterns differ between adults and children. Consequently, given the difficulty in interpreting four-variable interactions, separate models were fit for adults and children to facilitate coefficient interpretation. The summary for the model containing adults and children together is included in listed in Appendix D.

For both the adult and child groups, models were fit to predict the spectral distance between segments in the VC sequences. Best model fit for the adult and child models included the three-variable interaction of **Sequence duration**, **VC sequence**, and **Environment**: the improvement of models with this interaction over models with the three independent effects was significant for the adult model with alpha level $<.05$ ($\chi^2 = 12.69$, df=7,11 p=.01) and the child model ($\chi^2 = 15.18$, df=7,11, p=.004). Throughout the results, the children's patterns are additionally broken apart by age to view differences between age groups. However, note that in the child model, the addition of the variable **Age Group** (levels: 5, 6, 7, 8, 9, 10) did not improve upon a model with the interaction of **Sequence duration**, **VC sequence**, and **Environment**. This fact suggests that the pattern of coarticulation and duration by morphological environment did not significantly vary by the child's age group.

The final child model summary is listed in table 3.8 and the adult model summary is listed in table 3.7. Note that in the model summaries for the children and adults, the coefficients and standard error measurements were multiplied by 100 to make the otherwise small coefficients more interpretable. This step does not effect the direction or magnitude of the effect between predictors and outcome variables.

In the adult and child models, a positive coefficient for the predictor **VC sequence**, with the reference level '[ap]', shows that there was greater spectral distance between the segments in [ap] than [am], as one would anticipate given the acoustic signatures of [m] (voiced, sonorant) versus [p] (voiceless, transient) (adult model: β=61.78, z=10.78, p$<$.001; child model: β=72.39, z=16.32, p$<$.001).

Also of note in both the adult and child models is the direction of the **Sequence duration** predictor: a positive coefficient for **Sequence duration** indicates that longer duration VC sequences tended to also be less coarticulated (greater spectral distance between phones) (child model: β=0.06, z=3.15, p=0.002; adult model: β=0.33, z=5.52, p$<$.001). The coefficients suggest that when speakers, both adults and children alike, speak slower, they tend to coarticulate less. There is, however, an interaction between several of these predictors, which will demonstrate that children in particular do not always coarticulate less in longer-duration sequences. The direction of the interaction between **Sequence duration**, **VC sequence**, and **Environment** differs between the adult and child speakers so this will be interpreted separately for the two groups in the following section.

Adults

For the adult model, the interaction between **Sequence duration**, **VC sequence**, and **Environment** suggests a difference in the relationship between the response variable - amount of coarticulation - and **Sequence duration** that differs by **Environment** and

Table 3.7: Model predicting coarticulation in adults

term	estimate	S.E.	z.statistic	p.value	95% CI
Intercept	213.56	5.50	38.84	0.00	2.24,2.03
Sequence duration	0.33	0.06	5.52	0.00	0,0
Environment:across morpheme	-7.56	6.89	-1.10	0.27	0.06,-0.21
VC sequence:ap	61.78	5.73	10.78	0.00	0.73,0.51
Sequence duration*Environment:across morpheme	-0.22	0.08	-2.88	0.00	0,0
Sequence duration*VC sequence:ap	-0.21	0.07	-2.95	0.00	0,0
Environment:across morpheme*VC sequence:ap	6.80	8.05	0.84	0.40	0.23,-0.09
Sequence duration*Environment:across morpheme*VC sequence:ap	0.29	0.09	3.20	0.00	0,0

Table 3.8: Model predicting coarticulation in children

term	estimate	S.E.	z.statistic	p.value	95% CI
Intercept	216.80	3.70	58.66	0.00	2.24,2.1
Sequence duration	0.06	0.02	3.15	0.00	0,0
Environment:across morpheme	-0.94	3.33	-0.28	0.78	0.06,-0.07
VC sequence:ap	72.39	4.44	16.32	0.00	0.81,0.64
Sequence duration*Environment:across morpheme	0.07	0.03	2.38	0.02	0,0
Sequence duration*VC sequence:ap	-0.01	0.03	-0.32	0.75	0,0
Environment:across morpheme*VC sequence:ap	-6.06	5.08	-1.19	0.23	0.04,-0.16
Sequence duration*Environment:across morpheme*VC sequence:ap	-0.09	0.04	-2.40	0.02	0,0

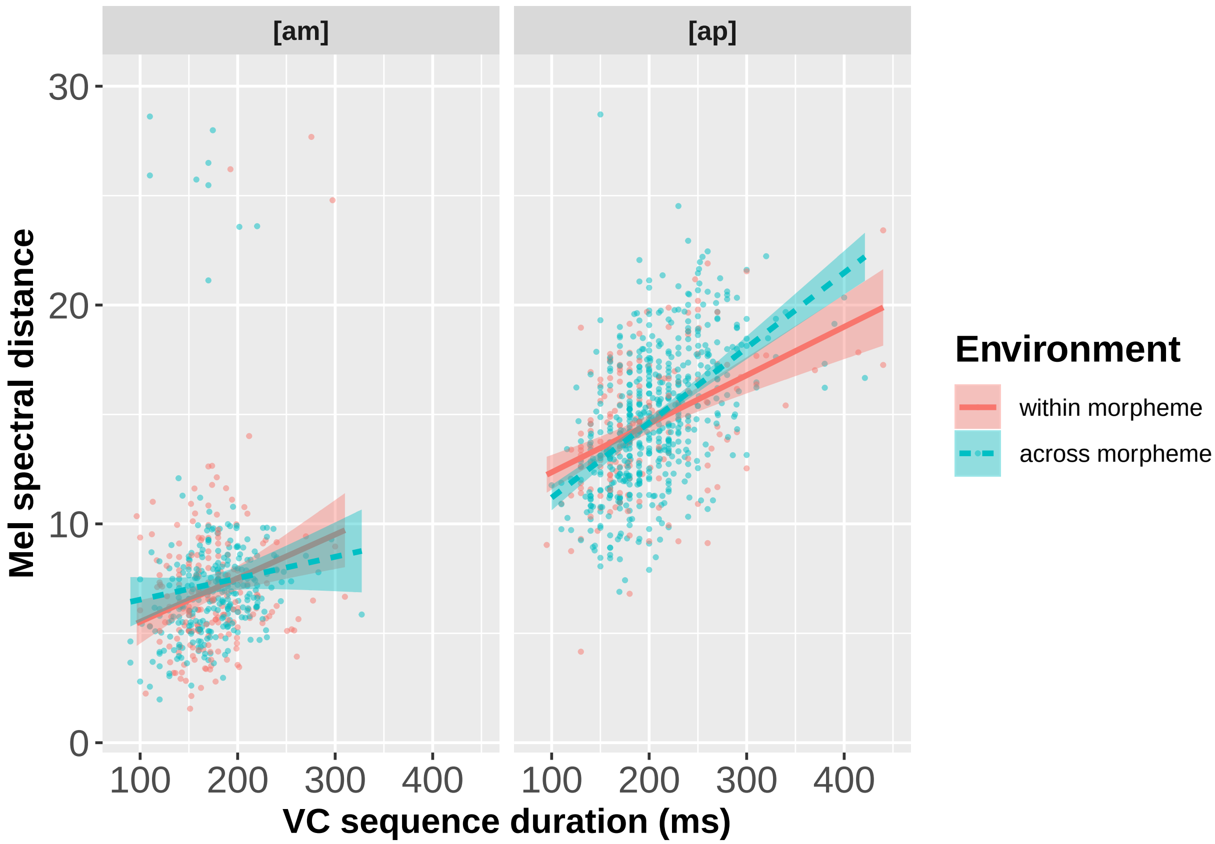


Figure 3.3: Coarticulation within VC sequence by sequence duration and morphological environment in adult speakers

VC sequence. As Figure 3.3 demonstrates, this difference by **Environment** is apparent in the steepness of the slope for the 'across morpheme' and 'within morpheme' conditions for [am] and [ap]. To quantify this difference for the sequence [am], the slopes of the two conditions were calculated. As the [am] panel in Figure 3.3 suggests, the slope for the 'within morpheme' condition was steeper (2.14) than the slope for the 'across morpheme' condition (2.06),[12] suggesting a different relationship between duration and coarticulation between the two word environments in adults.

Overall, the significance of the interaction **Sequence duration**, **VC sequence**, and **Environment** in adult speakers shows two important results: first, adults distinguish by word environment, both for [ap] versus [a#p] sequences and [am] versus [a#m] sequences. Second, complicating this finding, is the fact that adults distinguish between word environments differently depending upon the VC sequence. For [ap], though adults coarticulate roughly

[12]To reflect the data visualizations, these slopes were calculated on the beta coefficients before the coefficients were scaled by 100.

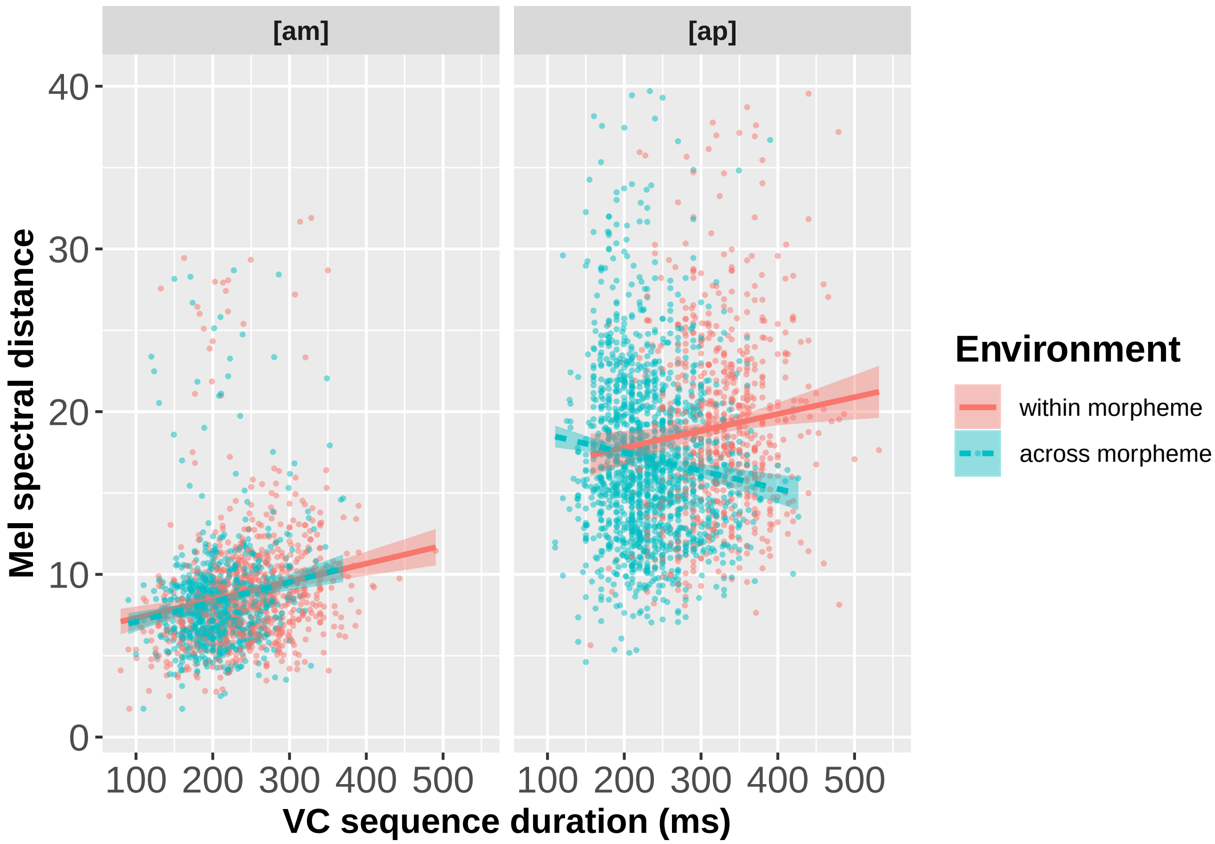


Figure 3.4: Coarticulation within VC sequence by sequence duration and morphological environment in all child speakers

equally across and within morphemes, the relationship between duration and coarticulation (longer duration equates to less coarticulation) is stronger in the 'across morpheme' condition. For [am], adults also distinguish between the two morphological environments by the relationship of VC duration and coarticulatory degree, but the effect of condition is reversed: the relationship between duration and coarticulation is stronger for the 'within morpheme' condition.

Thus, returning to one part of the central research question - does adult coarticulation differ by word environment - the current study finds that adults do coarticulate differently in the two word environments. However, despite these significant differences, there was nevertheless a positive relationship between duration and amount of coarticulation for all combinations of VC sequences and word environments. Adults consistently coarticulate less in longer-duration sequences. This result suggests that adult speakers may have one overarching articulatory plan for all environments and both VC sequences measured. The following section demonstrates how this relationship between duration and coarticulation may not be uniform between adults and children.

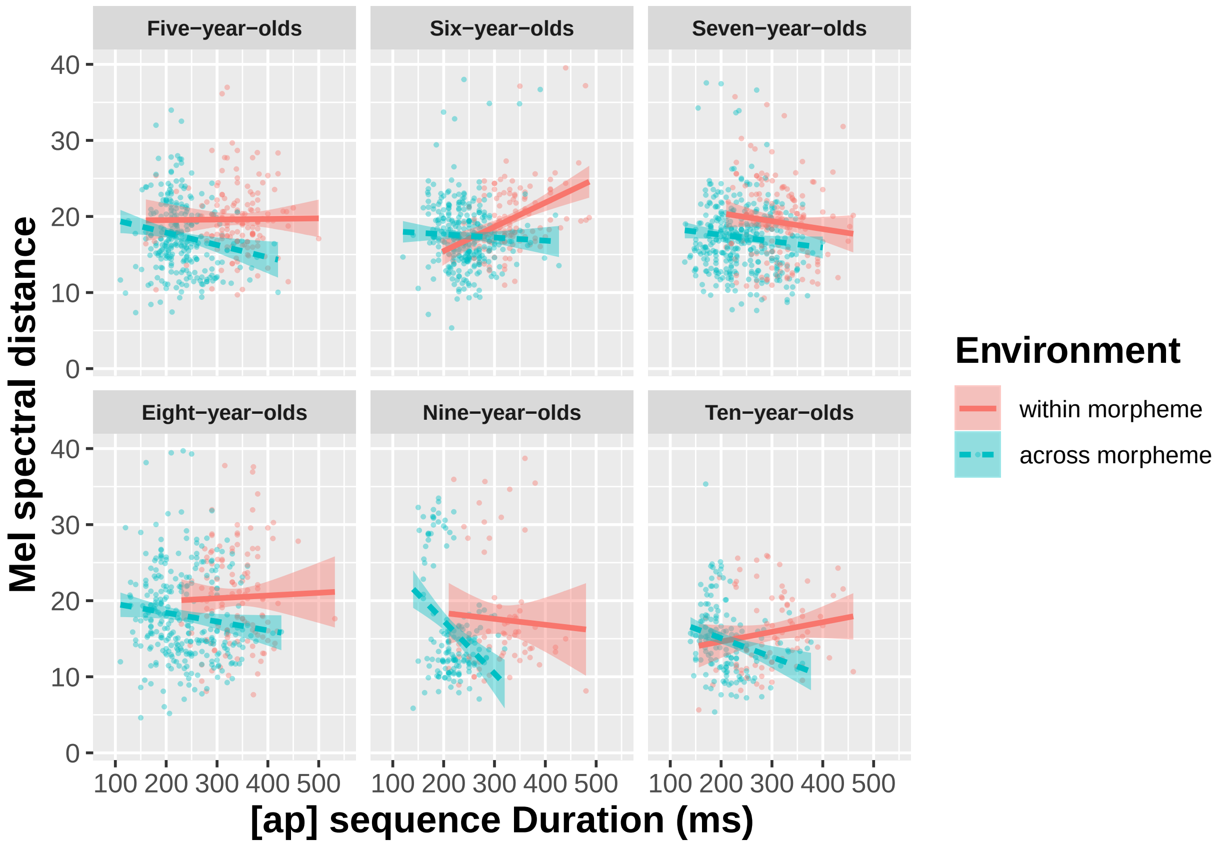


Figure 3.5: Coarticulation within [ap] by sequence duration, morphological environment, and age in child speakers

Children

Turning to the child model, the significant interaction of **Sequence duration**, **VC sequence**, and **Environment** suggests that children do not coarticulate similarly in longer-duration sequences for all combinations of **Environment** and **VC sequence** (Figure 3.4). Specifically, for [ap] sequences that occur across morpheme boundaries, the negative slope indicates that children actually coarticulate *more* in longer duration sequences. The positive slope for the within morpheme boundary condition suggests that children coarticulate less in longer-duration sequences, in line with all of the adult patterns. So, children coarticulate more between segments at morpheme boundaries in words inflected with the locative marker *-pi* than between those same segments that occur within morphemes.

Note that this negative relationship between duration and coarticulation is counter to the positive relationship for every combination of VC sequence and word environment in adult speakers. Adults consistently coarticulate less in longer-duration sequences regardless

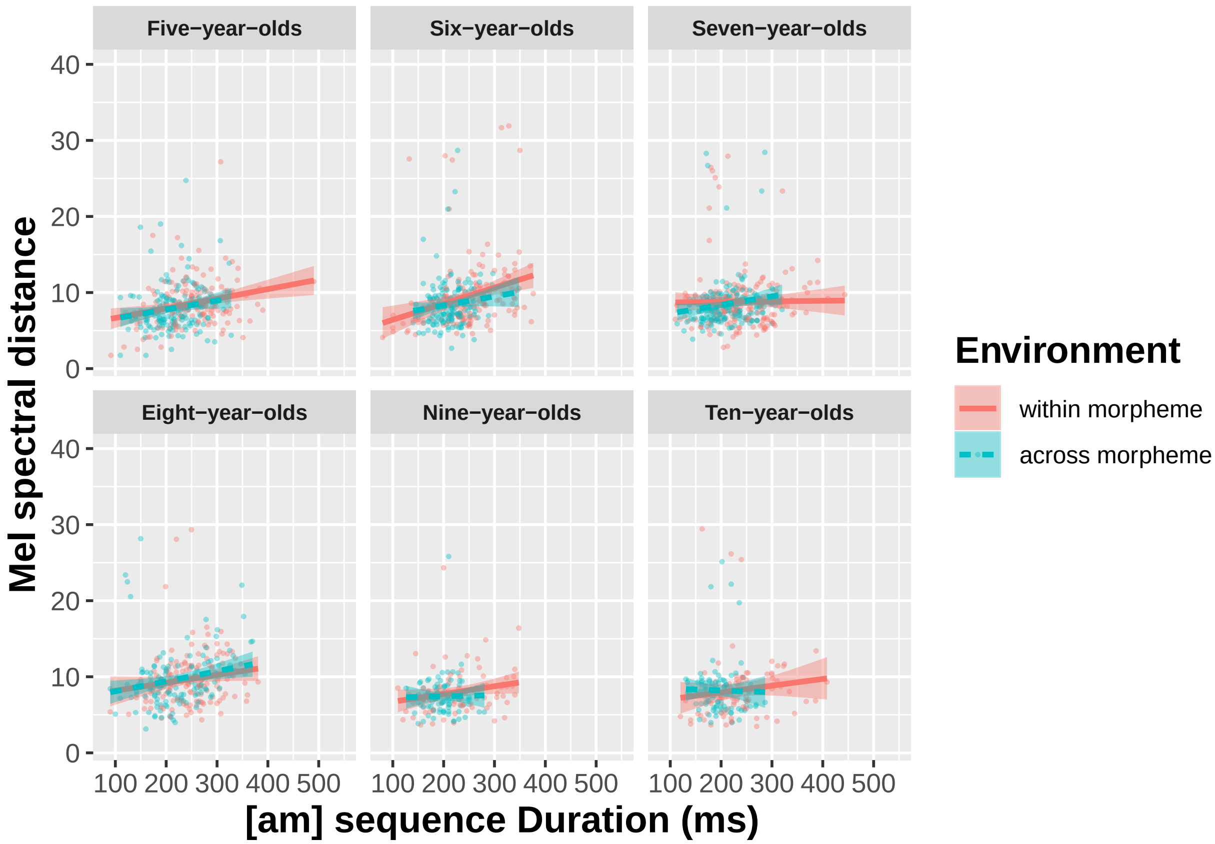


Figure 3.6: Coarticulation within [am] by sequence duration, morphological environment, and age in child speakers

of environment or VC sequence. The facet plot in Figure 3.5 plots this relationship between duration and coarticulation for [ap] for each age group (5-10 years) to ensure a consistent pattern across the groups. All age groups show the same negative relationship: the longer the [ap] sequence, the more the children coarticulate between [a] and [p] in the across morpheme condition.

The results for [am] in children demonstrate broadly similar results to the adult speakers: children coarticulate less between segments in longer-duration [am] sequences. The facet plot in Figure 3.6 once again shows a similar effect for each age group. Given the between-subject variability that typically characterizes child speech, these patterns by environment are further broken apart by individual child for each age group (age 5-10) (Appendix E) to ensure no large outliers with regards to the patterning by word environment. The results by are broadly similar across speakers.

Interim discussion

Comparing between the adult and child models, several preliminary conclusions can be made. First, responding to the original research question - do adults and children coarticulate differently within versus between morpheme boundaries - the results show that both adults and children differentiate by morphological environment. However, they do so in different ways. Adults have a single plan for both environments, and even both VC sequences: adults coarticulate less in longer-duration sequences, and overall, they coarticulate less between [a] and [p] within morphemes than across morpheme boundaries. The stark difference between adults and children emerges in the [ap] sequence patterning. Children differentiate between morphological environments via the relationship between duration and coarticulation as they coarticulate more in longer-duration sequences across morpheme boundaries and coarticulate *less* in longer-duration sequences within morphemes.

For words inflected with *-man*, children show a similar pattern to adults, though children do not differentiate by environment coarticulatorily. Rather, across morpheme sequences are shorter in duration than within morpheme sequences for the children. On the basis of these results, two questions remain. First, why do children differentiate between environments via a combination of duration and coarticulation; specifically, why do children produce shorter duration VC sequences across morpheme boundaries and longer duration sequences within morphemes? Second, why do children coarticulate more in longer-duration [ap] sequences that cross morpheme boundaries (e.g. *llama-pi* 'llama-LOC')? All of the other combinations of morphological environment and VC sequence in the adults and children suggest that the speakers coarticulate less in longer-duration segments.

The finding that children produce shorter-duration VC sequences in the 'across morpheme' condition than the 'within morpheme' condition for [am] and [ap] could be explained by a confound in word size and morphological environment. Coarticulation for the 'across morpheme' condition was, necessarily, measured across morphemes. However, to derive an across morpheme environment in Quechua, nouns are inflected with suffixes (e.g. *llama* 'llama' -> *llama-pi* 'llama-LOC'). As a result, almost all of the stimuli in the 'across morpheme' condition are at least one syllable longer in length than the stimuli for the 'within morpheme' condition. For example, coarticulation between [ap] was frequently measured within two syllable base roots (e.g. *llapa* 'lightening' and *llama* 'llama'). However, for the 'across morpheme' condition, [ap] coarticulation was frequently measured in three-syllable inflections of these nouns (e.g. *llapa-pi* 'lightening-LOC' and *llama-pi* 'llama-LOC'). Even for prosodically longer words where within morpheme coarticulation was measured, such as the three-syllable *hampiri* 'healer' and *hatun mama* 'grandmother', there were equivalent across morpheme stimuli that were one syllable longer (e.g. *hatun mama-mang* 'grandmother-ALL').

Compensatory shortening

To explore the possibility that durational differences between the 'across morpheme' and 'within morpheme' conditions could be due to word length, an exploratory analysis was conducted. I anticipated that sequence duration would shorten in words with more syllables, regardless of morphological context. This well-known tendency for segment durations to shorten in longer-duration/polysyllabic words is known as Compensatory Shortening (Harrington et al. 2015 Lehiste 1972; Munhall et al. 1992). To illustrate how this unfolds in the children's speech production for the current study, Figure 3.7 plots VC sequence duration by the number of syllables for the children and Figure 3.8 plots duration as a function of number of syllables for the adults. As the children's figure demonstrates, children's VC sequences are consistently shorter in words with more syllables, most notably between two- and three-syllable words. The same pattern is not apparent in the adult data: adults have fairly similar sequence lengths regardless of the number of syllables in the word (see Table 3.9 for descriptive statistics of duration by word length in syllables for the children and Table 3.10 for duration by word length results for the adults).

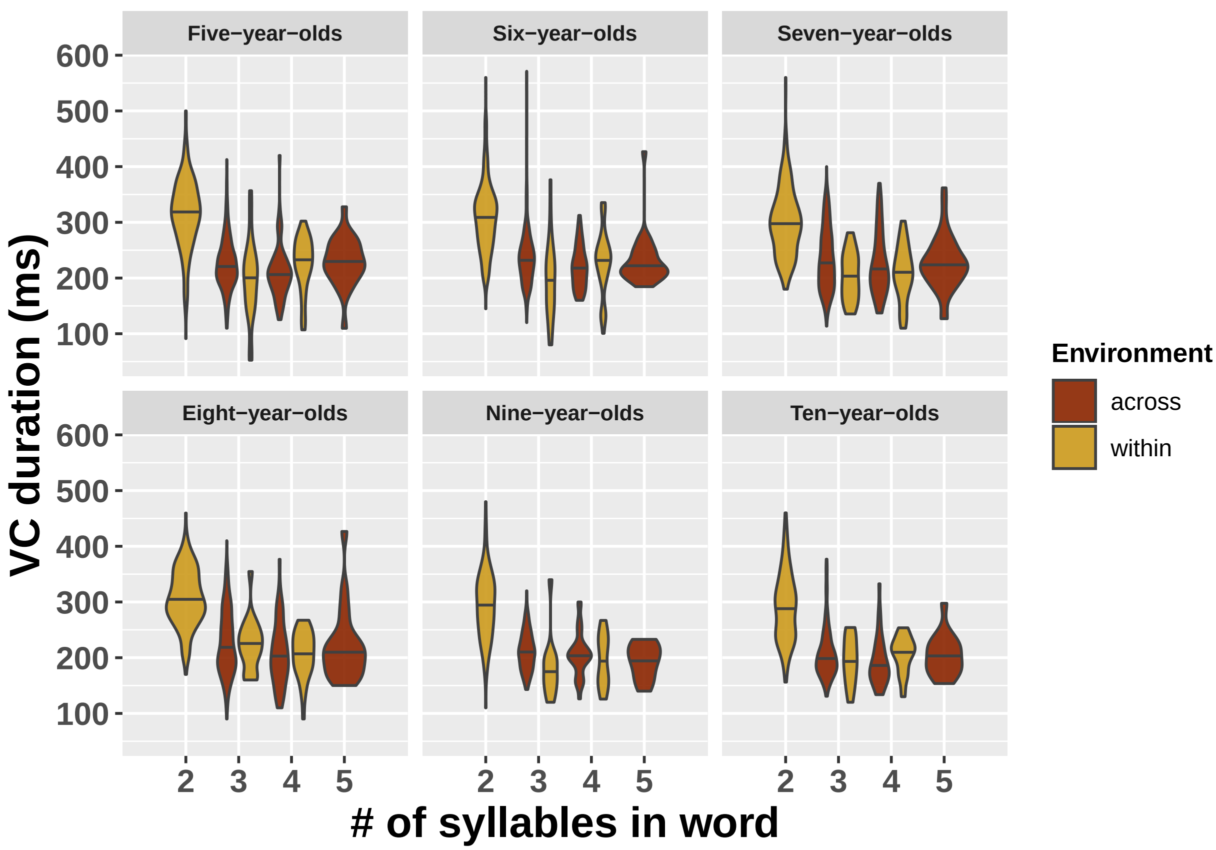


Figure 3.7: Sequence duration by word length and word environment: Children

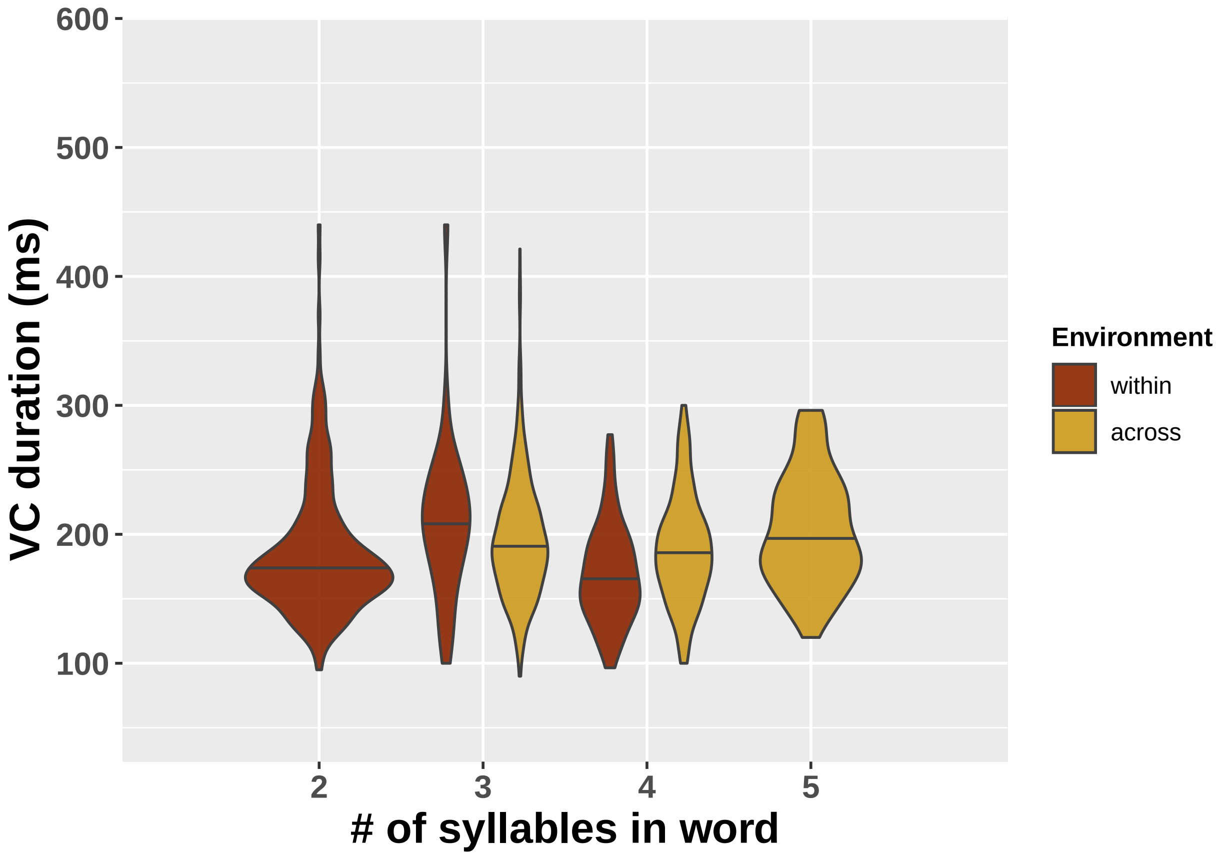


Figure 3.8: Sequence duration by word length and word environment: Adults

Table 3.9: Mean VC sequence duration by number of syllables in word for children

	[am]		[ap]	
Syllables	Duration (ms)	SD	Duration (ms)	SD
2	272.9	52	325.3	57
3	211.6	51	235.2	51
4	207.6	46	220.7	51
5	204.1	35	230.4	51

Table 3.10: Mean VC sequence duration by number of syllables in word for adults

	[am]		[ap]	
Syllables	Duration (ms)	SD	Duration (ms)	SD
2	177.6	39	192.4	57
3	176.5	35	210.0	50
4	168.0	36	195.3	43
5	178.5	51	207.3	40

To further explore the phenomenon of Compensatory Shortening in the children's speech, a linear mixed effects model was fit to predict the VC sequence duration in the children's speech (no skewed/non-negative predictors were included in the modeling so GLMMs were not necessary). Model fitting occurred as before in a forward-testing manner: the base model contained random effects of individual **Child** and **Word** (random slopes of Speaker by Word did not converge). Then, parameters were added in the following order: **Age** (5-10), **Number of Syllables** (2-5), **Environment** (across versus within), the interaction of **Number of Syllables** and **Environment**, and **VC sequence**. Only the predictors **Number of Syllables** and **VC sequence** improved baseline model fit (see Table 3.11 for model summary).

In the model summary, the positive beta coefficient for VC sequence with a reference level [am] indicates that the [ap] sequence was significantly longer than [am] sequences (as previous models demonstrated). Next, the negative beta coefficients for **Syllable Count** with a reference level of '2 syllables' indicated that VC sequence duration was approximately 80 ms shorter in three syllable words than two syllable (β=-79.68, t=-14.61, p<.001). Similarly, VC sequences were approximately 93 ms shorter in four syllable words than two syllable (β=-93.38, t=-13.06, p<.001) and 87 ms shorter in five syllable words than two syllable words (β=-87.01, t=-7.87, p<.001). The insignificance of **Environment** and child **Age** for the modeling suggests that that this relationship between duration and word length is independent of morphological environment and child age.

As these coefficients demonstrate, VC sequence decreases in larger words, with the largest differences between two- and three-syllable words. The diverse stimuli in the two- and three-syllable conditions (many different word types) suggest that this relationship by word length is relatively robust.

The only exception to the tendency to shorten sequences in larger words was that [ap] sequences are slightly longer in duration in 5-syllable words than 4-syllable words. However, in this exploratory analysis, the differences between the four and five syllable words were not tightly controlled: there were only two different five-syllable word stimuli: *hatun mama-man* 'grandmother-ALL" and *hatun mama-pi* 'grandmother-LOC'. I can only speculate that this relationship between sequence duration and number of syllables is strictly linear and would generalize to additional words with more syllables.

Table 3.11: Model predicting VC duration: children

Intercept	289.36***
	(276.14, 302.58)
Three syllables	-79.68^{***}
	(-90.37, -68.99)
Four syllables	-93.38^{***}
	(-107.40, -79.37)
Five syllables	-87.01^{***}
	(-108.68, -65.34)
VC sequence:[ap]	27.66***
	(19.56, 35.77)
Observations	3,877
Log Likelihood	$-20{,}122.98$
Akaike Inf. Crit.	40,261.97
Bayesian Inf. Crit.	40,312.07
Note:	$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

In conclusion, on the basis of this modeling, I propose that children may differentiate by morphological environment in their speech production. However, in Quechua, morphological structure is crossed with prosodic structure: complex words are always structurally longer than base forms. Regardless, children distinguish between morphological/prosodic environments primarily via the acoustic cue of *duration*: in child Quechua, the duration of sequences is shorter in words with more syllables (Compensatory Shortening). This duration pattern by word length was not present in the adult speech, a finding that is explored in the Discussion.

3.5 Discussion

The experiments in this chapter had two goals. The first experiment attempted to validate two relatively novel acoustic measures of coarticulation that could be employed in the higher frequencies of child speech. Spectral and temporal measures of coarticulation were employed over a large corpus of four-year-old children learning English. The results showed that the measures successfully replicated known lingual coarticulatory findings from the literature and thus are reliable measures of coarticulation for child speech.

The second part of this chapter, the experimental study, employed the spectral coarticulation measure from the validation study to measure the coarticulation of adult and child Quechua speakers in two morphological environments. I made two predictions for this experiment: 1) adult speakers would coarticulate less between two phones at a morpheme boundary than between the same phones within a morpheme boundary, and 2) child speakers would coarticulate more than adults between phones at morpheme boundaries. The reasoning behind these hypotheses was that frequency ratios between suffixes, and between roots and suffixes, predict decompostability in adults (Hay 2003; Kemps et al. 2005; Plag and Baayen 2009). Adult speakers have more experience with language than children and have larger vocabularies (Lorge and Chall 1963). Consequently, adults may weigh the ratios of inflected to base forms in their lexicons differently. And children may structure relationships between base and suffixal forms differently as they age and their vocabularies grow. Moreover, children appear to coarticulate less as they age and gain more experience with words and segments (Noiray et al. 2019a). A decrease in coarticulation suggests that phonological representations individuate into segment-sized units. For these reasons, I predicted adults might decompose words differently from children.

The first hypothesis of the experimental study proposed that adult speakers would coarticulate less between two phones at a morpheme boundary than between the same phones within a morpheme boundary. The results of the experimental study broadly confirmed this hypothesis. Adult speakers did, overall, coarticulate less across morpheme boundaries than within. This is to be expected because, as previous studies suggest, speech production - coarticulation and duration - indexes lexical retrieval and composition (Cho 2001; Kemps et al. 2005; Lee-Kim et al. 2013; Plag et al. 2017; Pluymaekers et al. 2010; Song et al. 2013b; 2013a; Sugahara and Turk 2009; Tomaschek et al. *under review*). Specifically, the adult speakers tended to coarticulate less between phones in a VC sequence at a morpheme boundary than within a morpheme because adults are more likely to compose morphologically complex words from their component parts. Though decomposition is probabilistic (e.g. Hay 2003), overall, adults may be less likely to access morphologically complex words holistically.

The second hypothesis in the experimental study proposed that the child participants would coarticulate more than adults between phones at morpheme boundaries. The results and conclusions of this hypothesis proved far more complex, particularly given the interactions between degree of coarticulation and speaking rate/VC sequence duration. In slower speech, adult speakers coarticulated less, both across and within morphemes. This replicates known interactions between speaking rate (here instantiated as VC sequence duration) and coarticulation (Agwuele et al. 2008; Matthies et al. 2001). In children, the same relationship between VC sequence duration and coarticulation appeared in the within-morpheme condition: within morphemes, children coarticulated less when they spoke slower. However, one primary difference between adults and children was that, in children, coarticulation did *not* vary as a consequence of VC sequence duration in the across-morpheme condition. Of further interest was the finding that the children's VC sequence duration was reliably shorter in the across-morpheme condition than the within-morpheme condition. The following section interprets the durational finding from the children's data.

Compensatory shortening

As suggested in the results section, one possible explanation for the shorter sequence duration in the across-morpheme condition for the children is the morphological structure of Quechua words. To construct morphologically complex words in this agglutinating language, additional suffixes must be appended to the root. As a result, almost all of the across-morpheme stimuli were approximately one syllable longer than the within-morpheme stimuli. I suggested that the longer prosodic (and temporal) length of the stimuli in the across-morpheme condition caused children to shorten the duration of the phones in each of the stimuli items (or, variably, lengthen phone duration in the shorter, within-morpheme stimuli). This aligns with the well-known phenomenon of compensatory shortening, or segment duration shortening in longer-duration/polysyllabic lexical items (Harrington et al. 2015; Lehiste 1972; Munhall et al. 1992).[13] This finding reinforces previous work on adult speech, as described in the literature review, that has demonstrated how roots have different variants in their bound and free forms (Kemps et al. 2005). One challenge in morphophonetics has been to identify the explanatory mechanisms behind morphologically-conditioned speech variation; at least in the current data, compensatory shortening could be one of these explanatory mechanisms.

This relationship between-prosodic word size and segment duration in the children's speech is notable. For example, it is unclear if these data mean that the children have some minimal planning unit size causing them to elongate prosodically short words (or shorten prosodically long words) to fit within the unit's temporal domain. However, even more interesting perhaps than the children's patterns is the *lack* of compensatory shortening in the adults. While children demonstrated a trade-off in sequence duration and prosodic word size, adults were insensitive to word size. Yet previous findings on compensatory shortening came from adult populations. Why don't adults compensate for word size in their speech production as the children, even the eldest ten-year-olds, appear to?

There are several potential explanations for the difference in compensatory shortening between adults and children, some related to language experience and others reflecting sociolinguistics in Bolivia. First, adults speak faster than children (Lee et al. 1999), leading to more extreme phonetic reduction in their speech. Adult Quechua speakers may speak so much faster, and reduce so much more, than children that there could be insufficient freedom in their speech duration to differentiate sequence duration by prosodic word size.

Children do not approximate adult-like speaking rates until early puberty (Lee et al. 1999; Smith 1992). This increase in speaking rate is replicated in the present study, even in

[13]The term compensatory shortening is also variably used in the literature to refer to shortening of stressed vowels compared to unstressed vowels in the context of polysyllabicity (Harrington et al. 2015), or the shortening of stressed vowels in the context of unstressed vowels and consonants (Fowler 1981).

a tightly-controlled experimental setting, as the average VC sequence duration for the adult speakers was just 192 ms (σ=47) compared to 252 ms (σ=69) for the five-year-olds, 255 ms (σ=70), for the eight-year-olds, and 231 ms (σ=63) for the ten-year-olds.

Speakers reduce significantly more in fast speech compared to slower, controlled speech. In fast speech, the vowel space is smaller and more centralized (Fourakis 1991; Tsao et al. 2006) - which may compromise contrasts (Koopmans-Van Beinum 1980), strengthen coarticulation between phones (Agwuele et al. 2008; Matthies et al. 2001), and cause the omission of entire segments or syllables (Johnson 2004). In the current study, recall that the average VC sequence duration in the adults was just 192 ms (σ=47) compared to 252 ms (σ=69) for the five-year-olds. The data suggest that all adult speech, regardless of prosodic or morphological structure, is maximally fast and reduced.

Though there is little work on the phonetics of highly morphologically complex languages, I can anecdotally attest to how speaking rate interplays with word structure in Quechua. Just as the orthographic forms of English words rarely correspond to their phonetic realization in fast, spontaneous speech, spoken words in Quechua deviate from their citation form. In Quechua, this is especially true with large words that contain numerous suffixes, where the most extreme reduction can be seen. In Quechua, the further a suffix is found from the root, the *more* likely it is to be reduced (shorter in duration, omission of segments/syllables). The explanation for this is partially aerodynamic (e.g. airflow). But there are also likely perceptual and information-theoretic explanations. When suffixes are highly reduced compared to stems, speakers can more easily demarcate between suffixes and roots, and identify word boundaries (Zingler 2018). Furthermore, word meanings are likely increasingly predictable as additional suffixes are added and the available semantic space of the word narrows.[14] Variability in the phonetic realization of English morphology (e.g. plurals) is dependent upon the predictability of plurality given the sentence frame (Cohen 2014), so it is possible that speakers likewise make probability calculations over multiple suffixes. Much more work is needed to understand probabilistic reduction based on word structure, however, overall one explanation for adults' lack of compensatory lengthening is their speaking rate and phonetic reduction.

A second explanation for the lack of compensatory shortening in adults may be that the adults are more dominant in Quechua than the children. This means that the adults may speak faster and, again, may be unable to differentiate prosodic structure via duration. Recent changes in Bolivia's educational policy, as well as the country's general sociolinguistic situation, may have led to adults' increased fluency. Bilingual education in Bolivia became mandatory in 1994 and has, in theory, been relatively widespread since the early 2000s (Benson 2004). This means more children, especially indigenous students and young girls, are attending and completing more schooling than ever before (Hornberger 2009).

In practice, however, bilingual education often takes the form of Spanish-only classrooms. In Quechua-speaking areas, many trained teachers do not speak Quechua fluently or are not

[14]For English, however, Plag and Baayen (2009) found that those suffixes farthest from the stem are also the most productive and available for use in novel environments.

provided with teaching materials and textbooks written in Quechua. The result - students completing more schooling but instructed in Spanish - has been rapid language shift (Hornberger and King 1996). This has been apparent even in the last decade, as the first generation of women educated in this system are now raising their children using both Quechua and Spanish, instead of monolingual Quechua, in an increasingly Spanish-dominant environment.

These sociolinguistic dynamics could manifest in the present sample as the adult female participants (many of whom are mothers) may be more Quechua-dominant than some of the children in the sample. Though the adult females who participated in the study were only, on average, 13 years older than the eldest children (adult μ_{age}=23 years; σ=5.46 years), and all adults and children identified as bilingual Quechua-Spanish speakers, their language practices may reflect the recent changes in educational policy. It is also important to note that all of the children in the sample attended school, in Spanish, for 3-4 hours per day. However, only one of the adult females was still attending school (taught in Spanish). This difference could also explain usage patterns between the age groups.

If the adult females were more Quechua-dominant, or used Quechua more frequently, they would speak faster, more fluently, and could reduce more, as described above. Thus, although the adult speakers in this study undoubtedly did speak faster than the children for speech maturation reasons - it takes time and practice to master the articulatory speed of an adult (Lee et al. 1999) - the adults may also have spoken Quechua faster, and thus failed to compensate for prosodic structure to the degree that the children did, because they use Quechua more frequently. This proposal may not entirely explain the differences in compensatory shortening between the adult and child Quechua speakers - previous studies on compensatory shortening (Lehiste 1972; Munhall et al. 1992) reported on highly-fluent, monolingual adult speakers who *did* compensate for prosodic structure - but it is one explanation for the differences observed between age groups in the present work.

While this study did not report participants' bilingual language dominance, all speakers identified as Quechua-Spanish bilinguals and reported using both languages.[15] The decision not to conduct a traditional language usage survey for several reasons. Traditional measures of self-reported bilingual dominance, such as the bilingual language profile (Birdson et al. 2012), often rely on participant literacy or familiarity and comfort with written behavioral research surveys. Also, the traditional stigmatization of indigenous languages in Bolivia may render self-reports of language dominance unreliable. Nevertheless, one potential difference between the adults and children could be their Quechua language usage or dominance.

Do children distinguish between morphological environments in their speech production?

The original research question in this study asked if adults and children coarticulated differently across versus within morpheme boundaries. As the discussion of compensatory

[15]Computation of the children's language dominance is underway using naturalistic recordings of the children's daily language usage. This method skirts the issue of self-reported usage, as outlined above.

shortening outlined, children appear to distinguish by morphological environment, suggesting that they decompose words. However, this could be a consequence of prosodic structure, which is correlated with morphological structure in Quechua. The different temporal and coarticulatory patterns seen in the across- and within-morpheme conditions could reflect morphological structure and lexical access, or the patterns could reflect prosodic structure. Children may modulate their speech temporally to demarcate between morphological environments - faster segments across morphemes and slower segments within. Or, as compensatory shortening suggests, children may shorten the duration of VC sequences in prosodically longer words.

One way to evaluate these two explanations is to control for prosodic structure by measuring duration and coarticulation across morphemes versus within morphemes, but always in words of the same length (in number of syllables). This is made more difficult in Quechua because the language has canonical penultimate stress: even if duration/coarticulation were measured between [am] in a three-syllable within-morpheme stimulus (e.g. ll**a**.'**m**a.-man 'llama-ALL'), and a corresponding three-syllable across-morpheme stimulus (e.g. pa.'p**a**.-**m**an 'potato-ALL'), the stress would not be controlled. In other words, any differences in duration/coarticulation patterns between the across-morpheme and within-morpheme stimuli could be attributable to known acoustic correlates of stress, and might not reflect morphological *or* prosodic structure. This issue reflects some recent acknowledgements in morphophonetics that it may be nearly impossible to design studies that perfectly isolate morphological effects from the plethora of other correlated variables (Strycharczuk 2019, cf. Seyfarth et al. 2018).

In the current dataset there are three lexical items that control for prosody (and thus stress) and allow us to tease apart the prosodic versus morphological explanations for the children's patterns. Specifically, there are three four-syllable stimuli - two across morpheme (*imi'lla-man* 'girl-ALL' and *juk'u'cha-man* 'mouse-ALL') and one within morpheme (*hatun'mama*) - where the sequence [am] falls in stressed position. In the following exploratory analysis, the duration of [am] and the coarticulation between [a] and [m] were measured for all of the children and adults. Results are listed in tables 3.12 and 3.13.

Table 3.12: Mean duration of [am] across and within morphemes in prosodically-controlled stimuli

	Across boundary		Within boundary	
Age	Duration	SD	Duration	SD
5	193.3(ms)	32	245.7	30
6	199.0	25	242.9	50
7	188.8	34	229.8	36
8	205.3	49	225.2	30
9	189.4	37	226.4	32
10	170.9	26	216.6	17
adult	165.4	30	230.6	33

Table 3.13: Mean spectral distance between [a] and [m] across and within morphemes in prosodically-controlled stimuli

	Across boundary		Within boundary	
Age	Spectral Distance	SD	Spectral Distance	SD
5	7.34	2.52	8.49	2.45
6	9.78	5.84	9.51	5.39
7	7.85	1.30	8.82	5.05
8	9.61	4.00	8.89	3.37
9	7.62	1.88	6.63	2.15
10	8.34	1.57	8.29	4.66
adult	8.90	4.17	5.48	1.20

As the results demonstrate, both the children and adults have shorter [am] sequences in the across-morpheme condition (stimuli *imilla-man* and *juk'ucha-man*). The duration of [am] also decreases with age, as the adults speak fastest. Thus, when controlling for prosody, it appears that children do pattern like the adults and may distinguish by morphological environment using durational cues. However, for coarticulation, the results differ between adults and children. While the children do not appear to distinguish greatly between morphological environments, coarticulating roughly equally in the two environments regardless of age, the adults distinguish between the two areas. As anticipated from previous work (e.g. Cho 2001), the spectral distance between [a] and [m] is greater in the across-morpheme stimuli (*imill**a-m**ang* and *juk'uch**a-m**an* than the within-morpheme stimulus (*hatunm**am**a*).

Unlike the durational results, the exploratory coarticulation results suggest that adults, but not children, distinguish between the the morphological environments.

These data are too sparse to make definitive conclusions; only three distinct lexical items were tested. Concerning the central research question - do children distinguish between morphological environments - the exploratory analysis gave conflicting results: children distinguish between environments like adults durationally, but not coarticulatorily. However, the exploratory analysis does suggest that children's compensatory lengthening is likely morphological in nature, not prosodic, since the [am] was shorter in duration in the across-morpheme condition in the four-syllable stimuli.

Thus, using a combination of coarticulatory and temporal cues in their production, children appear to distinguish between across-morpheme and within-morpheme stimuli. But, in most of the current stimuli, morphological environment is correlated with prosodic environment (to derive morpheme boundaries, extra syllables must be added to roots). It therefore remains unclear if children's different speech patterns across the two morphological environments reflect prosodic planning, lexical planning, or both.

Future work, ideally on languages with different and complex morphological structures, is needed to further explore the relationships between children's speech production and word structure. Going forward, researchers may also benefit from the use of articulatory measures, in addition to acoustic measures, to explore the relationships between coarticulation and morphological structure in children. Given the logistics of ultrasound imaging, and the limitations of fieldwork, the current study was not able to collect articulatory data, which would otherwise be a valuable addition.

I was unable to control for, or examine the effect of, lexical frequency of the base or inflected forms. This is unfortunate because much variability in morphological parsing, and morphophonetics, is attributable to lexical frequency (Seyfarth et al. 2018) or frequency ratios between stems and suffixes (Plag and Baayen 2009; Hay 2003). At this time, however, it is not possible to reliably calculate lexical frequency statistics for Quechua, and indeed for most of the world's languages. A large, naturalistic corpus of child and adult Bolivian Quechua has been collected (Cychosz 2018), but it will be years before it is sufficiently transcribed to calculate lexical statistics. In the mean time, researchers interested in the morphophonetics of underdocumented languages could possibly use age of acquisition as a proxy for word frequency (Morrison et al. 1992).[16]

3.6 Conclusion

The studies in this chapter had two goals. First, two relatively novel measures of coarticulation, that are suitable for the high frequencies of child voices, were validated on a large corpus of four-year-old English-speaking children. Next, one of these measures was

[16] I found that obtaining age of acquisition norms was not possible for the current study because the Quechua-speaking adults were relatively unfamiliar with behavioral research and the methods that would have been required to solicit age of acquisition information.

used to compute coarticulation in adult and child Quechua speakers in two morphological environments: within morphemes and across morpheme boundaries.

Results showed that, using a combination of coarticulatory and temporal cues, adults distinguished between two morphological environments in their speech production. This replicated known speech production patterns by word environment but in an understudied, morphologically complex language. The children showed increased prosodic sensitivity where the adults did not: children shortened the duration of sequences in prosodically longer words, which also happened to be morphologically complex. It was suggested that the difference between adults and children could be attributable to adults' faster speaking rate and increased practice with Quechua. Future work is needed to determine if the children's speech patterns reflect prosodic planning, morphological planning, or both.

Overall, this study has demonstrated some of the complexities that arise in morphophonetic patterning and speech development, and the importance of extending recent studies in this sub-domain of phonetics to languages with vastly different word structures and speakers with different language experiences.

3.7 Appendices

Appendix A

Table 3.14: South Bolivian Quechua Consonant Inventory

	Bilabial	Dental	Postalveolar	Palatal	Velar	Uvular	Glottal
Plosive	p	t	ʧ		k	q	
Aspirated	p^h	t^h	$ʧ^h$		k^h	q^h	
Ejective	p'	t'	ʧ'		k'	q'	
Nasal	m	n		ɲ			
Fricative		s					h
Tap		ɾ					
Approximant	w		ʎ	j			
Lateral		l					

Appendix B

Table 3.15: Stimuli used in real word repetition tasks

Real word*	Translation
'warmi	'woman' *training trial*
'wasi	'house' *training trial*
'qhari	'man' *training trial*
'chita	'sheep'
'p'esqo	'bird'
ju'k'ucha	'mouse'
'waka	'cow'
'wallpa	'chicken'
'mama	'mom'
'papa	'potato'
't'ika	'flower'
'llama	'llama'
'cuca	'coca (leaves)'
u'hut'a	'sandal'
ham'piri	'healer'
i'milla	'girl'
'llapa	'lightening'
'api	'corn/citrus drink'
'ch'ulu	'hat'
'punku	'door'
'thapa	'nest'
'punchu	'poncho'
'pampa	'prairie'
'sunkha	'beard'
hatun'mama	'grandma'
'wawa	'baby/child'
'runtu	'egg'
'qolqe	'money'
'q'apa	'palm of hand'
'alqo	'dog'
'q'epi	'bundle'

* For the real words, ' indicates stress, ' indicates ejective

Appendix C

Table 3.16: Real word repetition stimuli to elicit [am]

Real word*	Translation	Morpheme environment†
chi'**ta-m**an	'sheep-ALL'	across
cu'**ca-m**an	'coca (leaves)-ALL'	across
hatunma'm**a-m**an	'grandma-ALL'	across
imi'll**a-m**an	'girl-ALL'	across
juk'u'ch**a-m**an	'mouse-ALL'	across
lla'm**a-m**an	'llama-ALL'	across
lla'p**a-m**an	'lightening-ALL'	across
ma'm**a-m**an	'mom-ALL'	across
pam'p**a-m**an	'prairie-ALL'	across
pa'p**a-m**an	'potato-ALL'	across
q'a'p**a-m**an	'palm of hand-ALL'	across
sun'kh**a-m**an	'beard-ALL'	across
t'i'k**a-m**an	'flower-ALL'	across
tha'p**a-m**an	'nest-ALL'	across
wa'k**a-m**an	'cow-ALL'	across
wall'p**a-m**an	'chicken-ALL'	across
wa'w**a-m**an	'baby/child-ALL'	across
'm**am**a	'mom'	within
'll**am**a	'llama'	within
h**am**'piri	'healer'	within
h**am**pi'ri-pi	'healer-LOC'	within
'p**am**pa	'prairie'	within
hatun'm**am**a	'grandma'	within

* ' indicates stress, ' indicates ejective

Appendix D

Table 3.17: Model predicting coarticulation in children and adults

	estimate	S.E.	z.statistic	p.value	95% CI
Intercept	2.14	0.04	59.61	0.00	2.21,2.07
Sequence_duration_scaled	0.00	0.00	3.65	0.00	0,0
VC_sequenceap	0.73	0.05	16.20	0.00	0.82,0.64
Environment:across morpheme	-0.05	0.07	-0.67	0.50	0.09,-0.19
Sequence_duration*VC sequence:ap	0.02	0.03	0.57	0.57	0.09,-0.05
Sequence_duration* VC sequence:ap	0.00	0.00	-0.63	0.53	0,0
Sequence_duration* Age:adult	0.00	0.00	2.66	0.01	0,0
VC sequence:ap*Age:adult	-0.10	0.06	-1.71	0.09	0.01,-0.21
Sequence_duration* Environment:across morpheme	0.00	0.00	2.37	0.02	0,0
VC sequence:ap* Environment:across morpheme	-0.06	0.05	-1.17	0.24	0.04,-0.17
Age:adult* Environment:across morpheme	-0.09	0.05	-1.74	0.08	0.01,-0.2
Sequence_duration* VC sequence:ap*Age:adult	0.00	0.00	-1.25	0.21	0,0
Sequence_duration_scaled* VC_sequence:ap:Environmentacross morpheme	0.00	0.00	-2.25	0.02	0,0
Sequence_duration*Age:adult* Environment:across morpheme	0.00	0.00	-3.36	0.00	0,0
VC sequence:ap*A ge:adult*Environment:across morpheme	0.16	0.07	2.31	0.02	0.29,0.02
Sequence_duration* VC sequence:ap*Age:adult*Environment:across morpheme	0.00	0.00	3.44	0.00	0,0

Appendix E

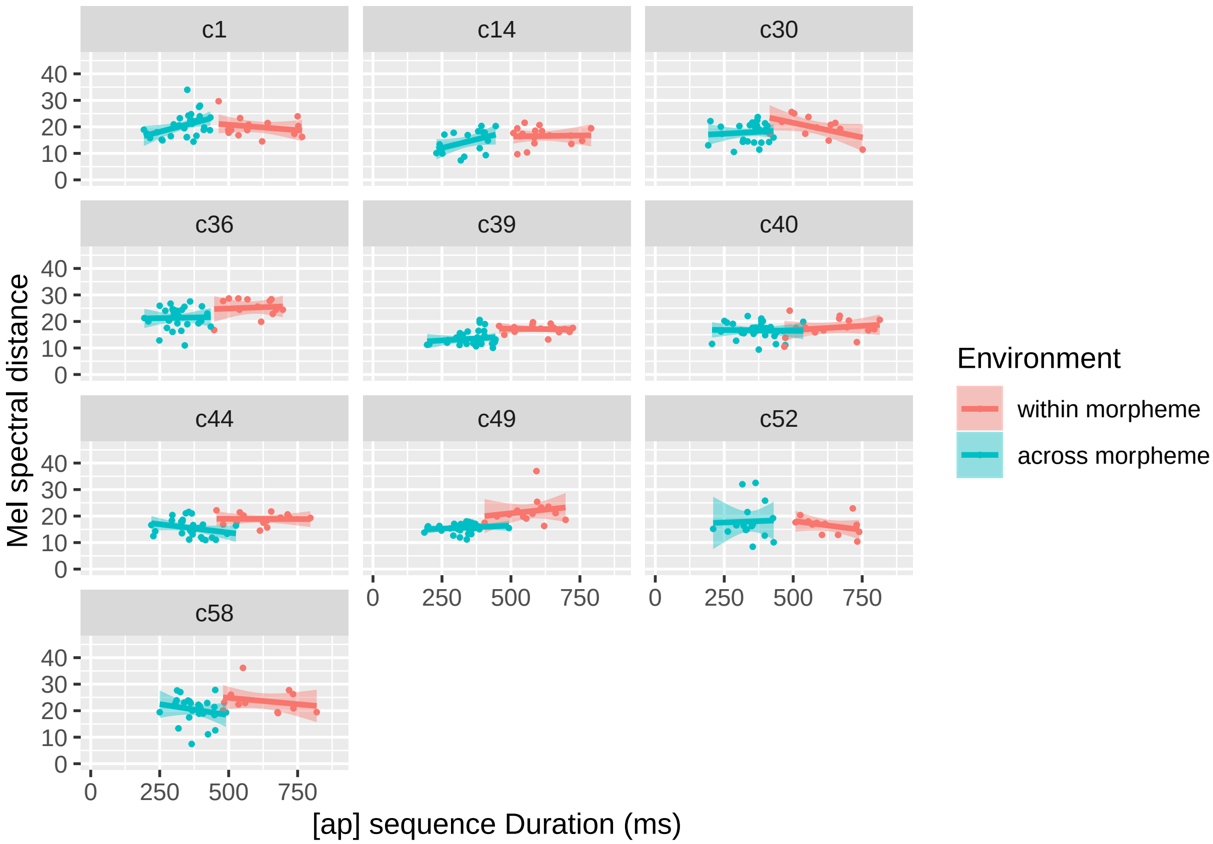


Figure 3.9: Coarticulation by [ap] duration, word, and morphological environment in five-year-old children

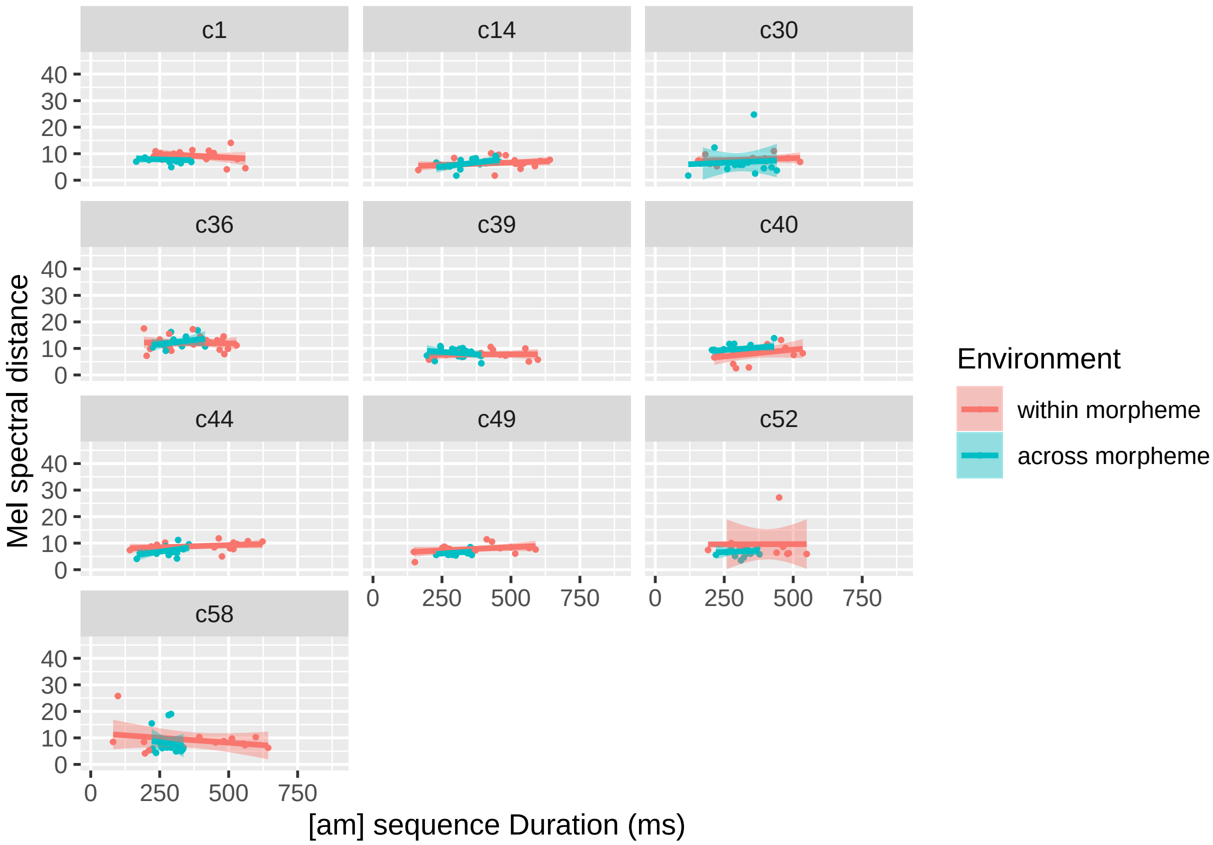


Figure 3.10: Coarticulation by [am] duration, word, and morphological environment in five-year-old children

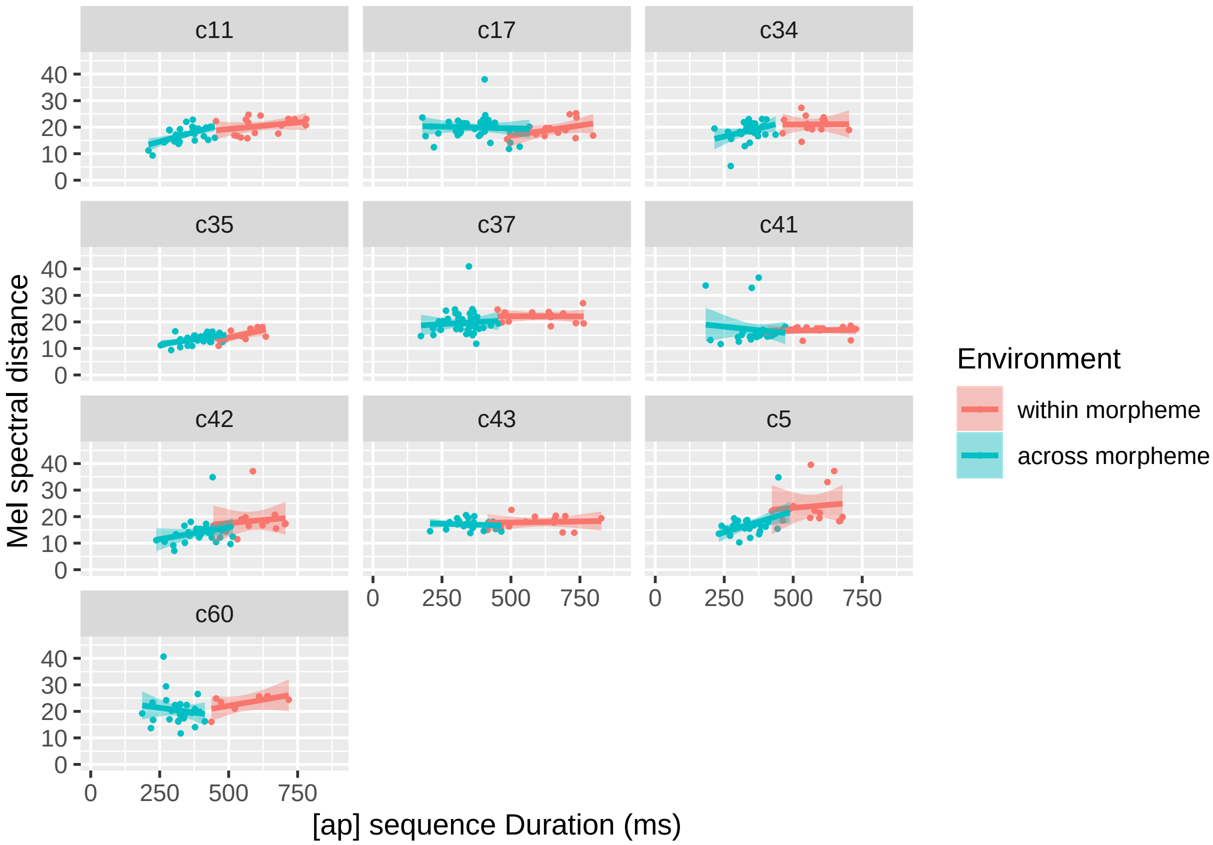


Figure 3.11: Coarticulation by [ap] duration, word, and morphological environment in six-year-old children

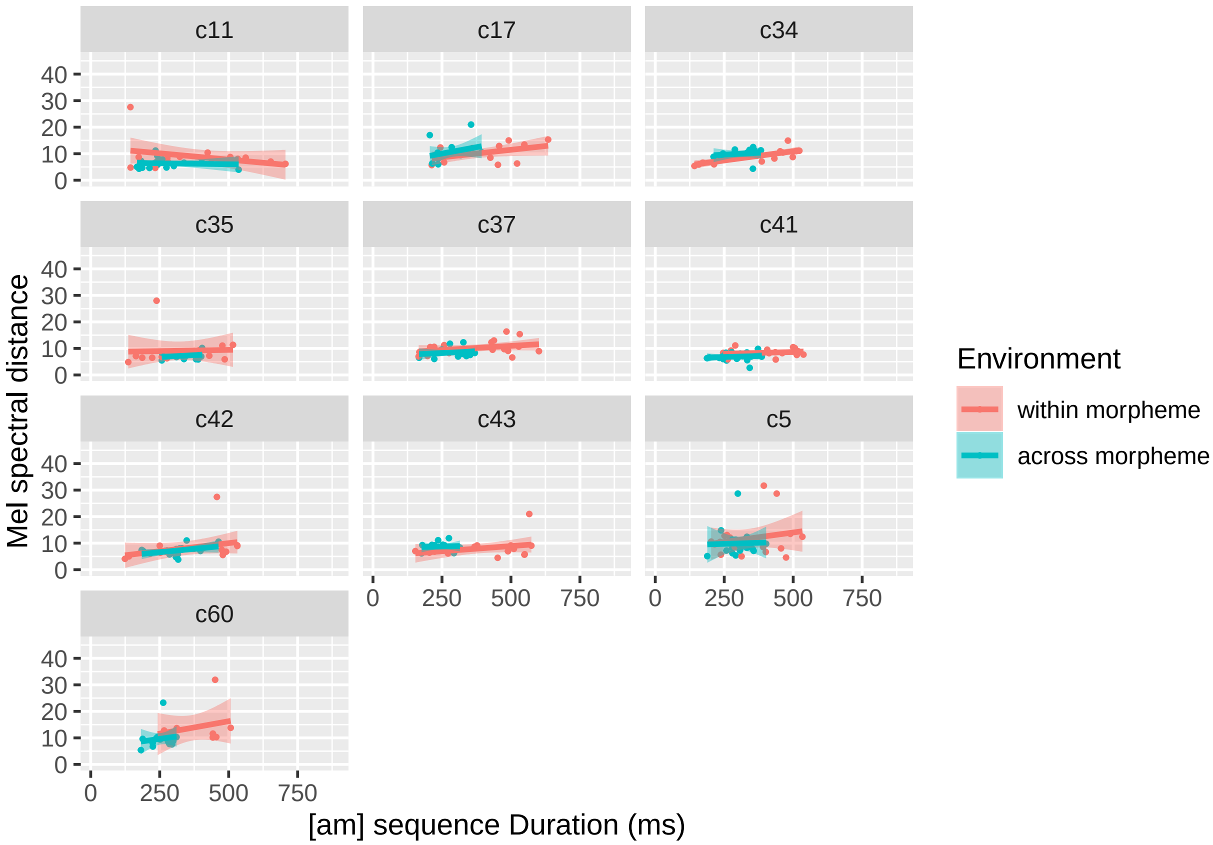


Figure 3.12: Coarticulation by [am] duration, word, and morphological environment in six-year-old children

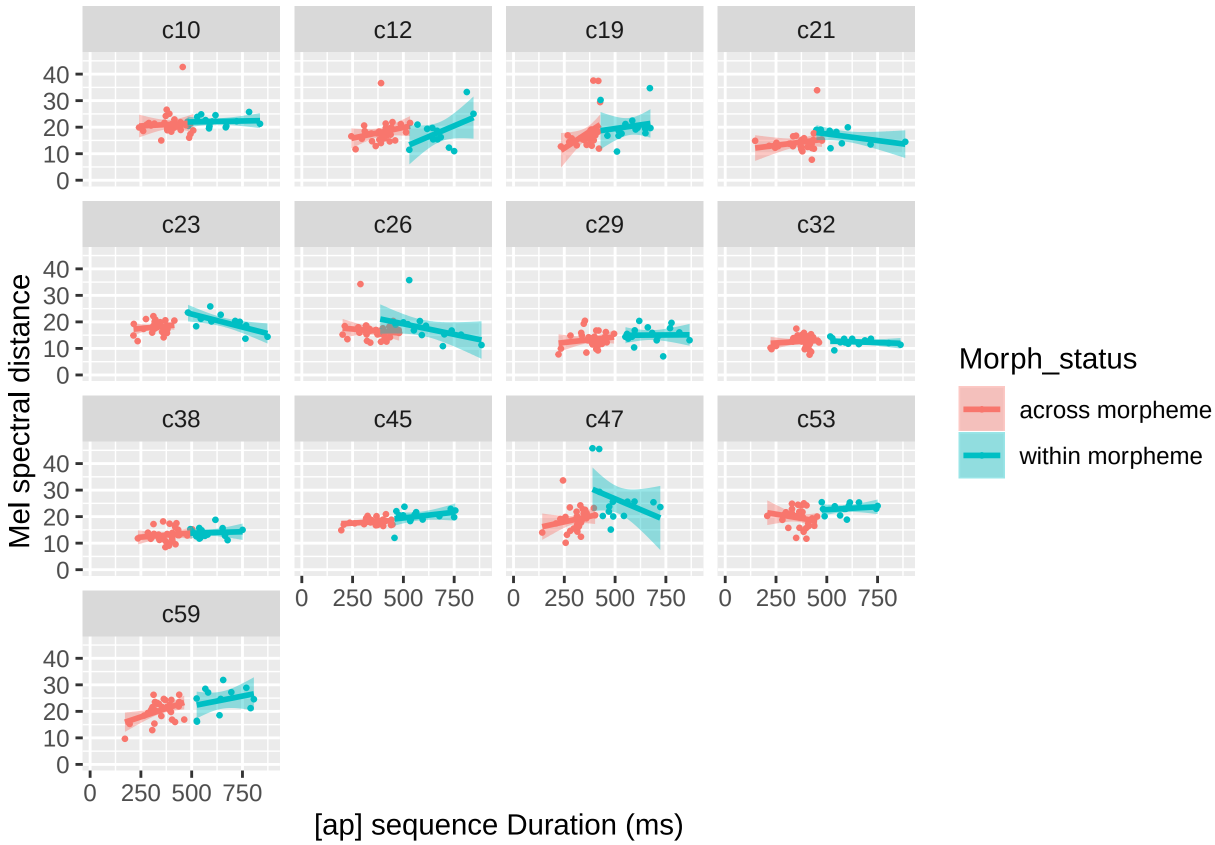


Figure 3.13: Coarticulation by [ap] duration, word, and morphological environment in seven-year-old children

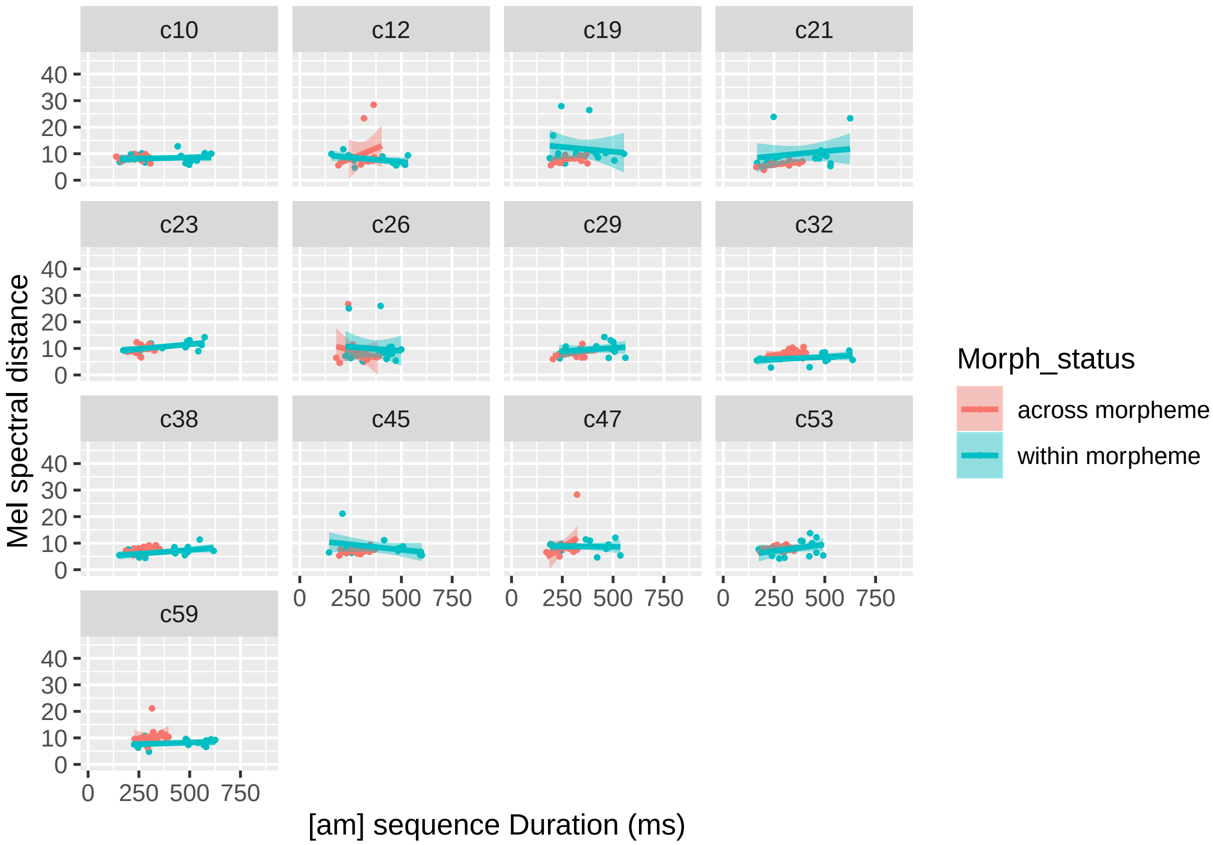


Figure 3.14: Coarticulation by [am] duration, word, and morphological environment in seven-year-old children

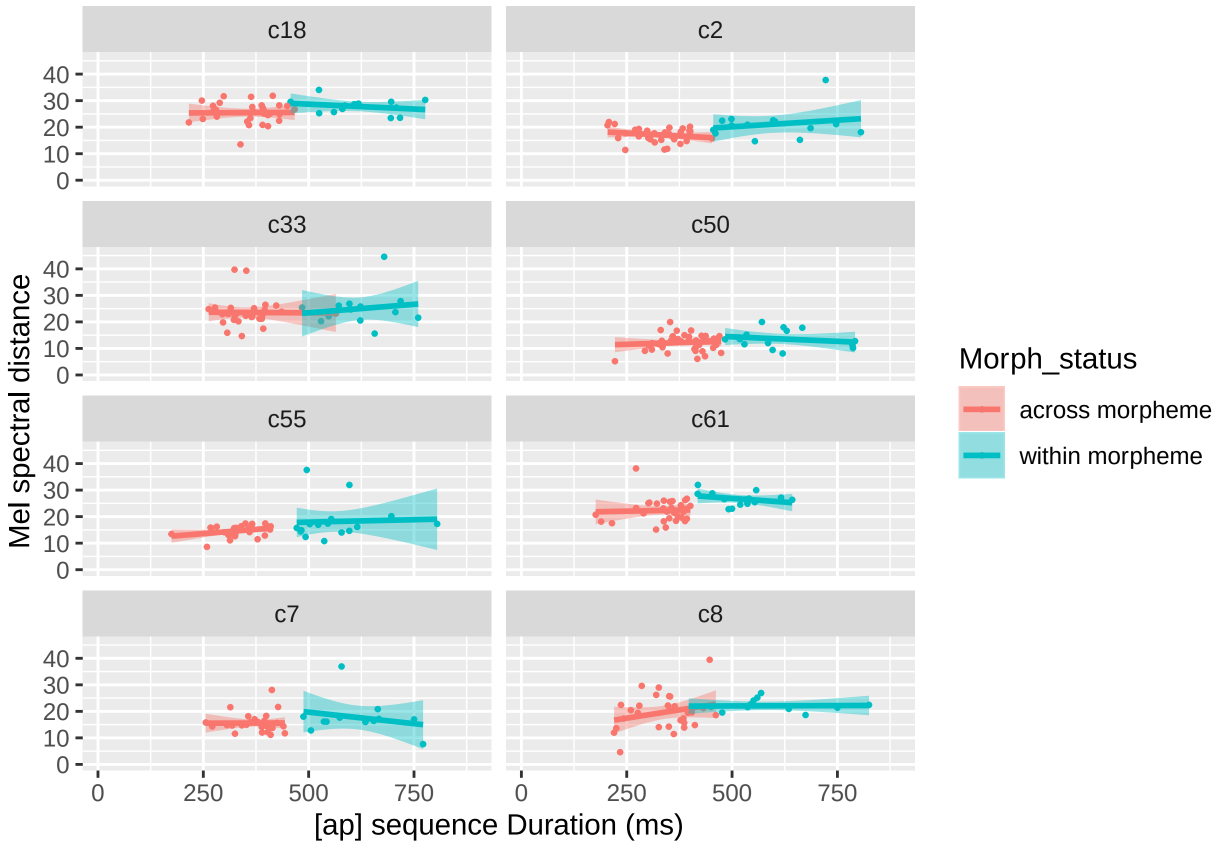


Figure 3.15: Coarticulation by [ap] duration, word, and morphological environment in eight-year-old children

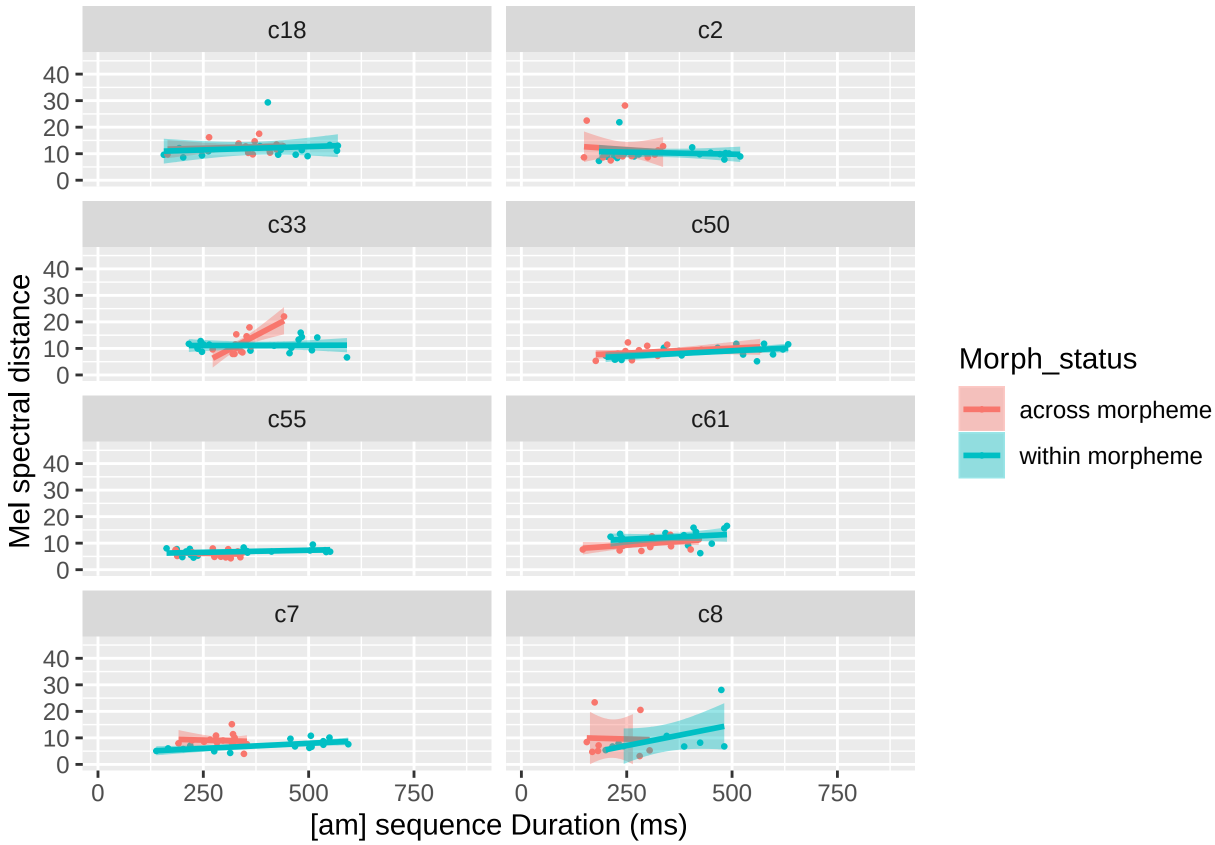


Figure 3.16: Coarticulation by [am] duration, word, and morphological environment in eight-year-old children

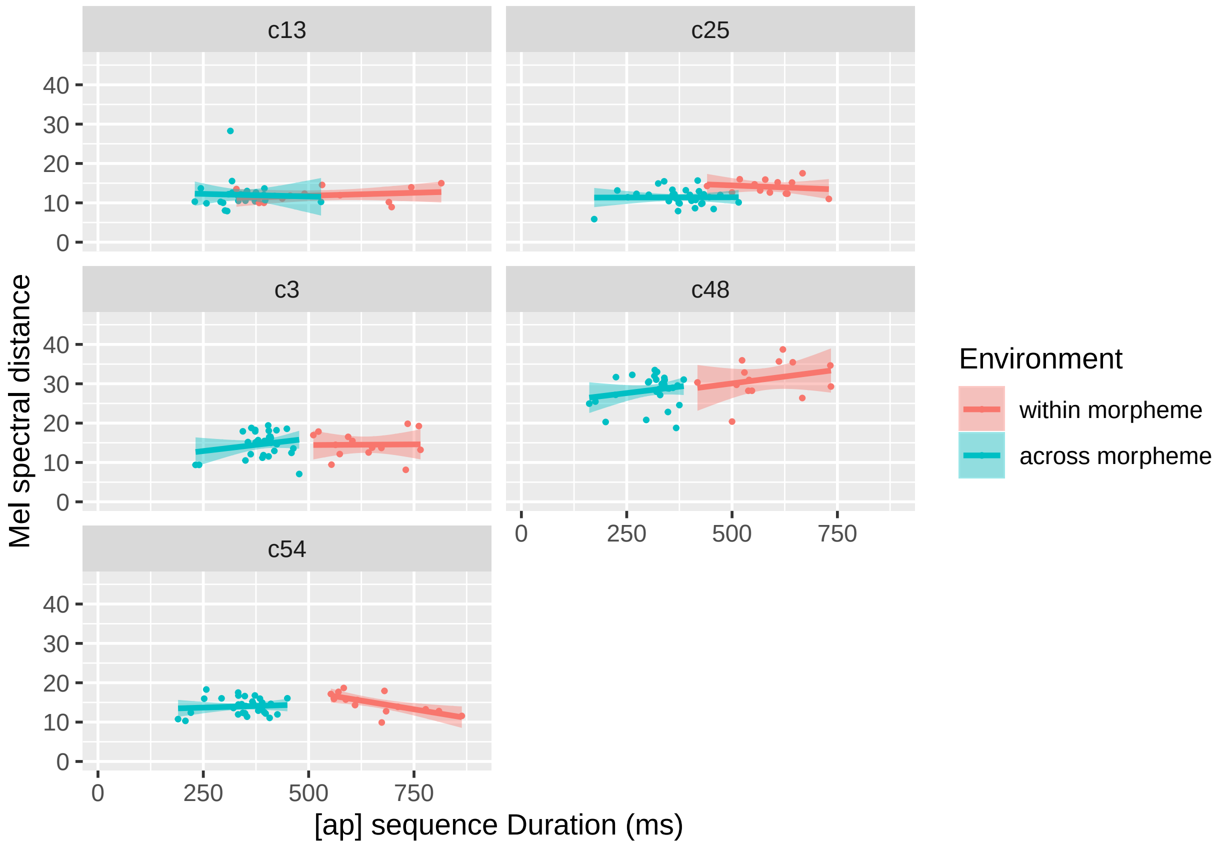


Figure 3.17: Coarticulation by [ap] duration, word, and morphological environment in nine-year-old children

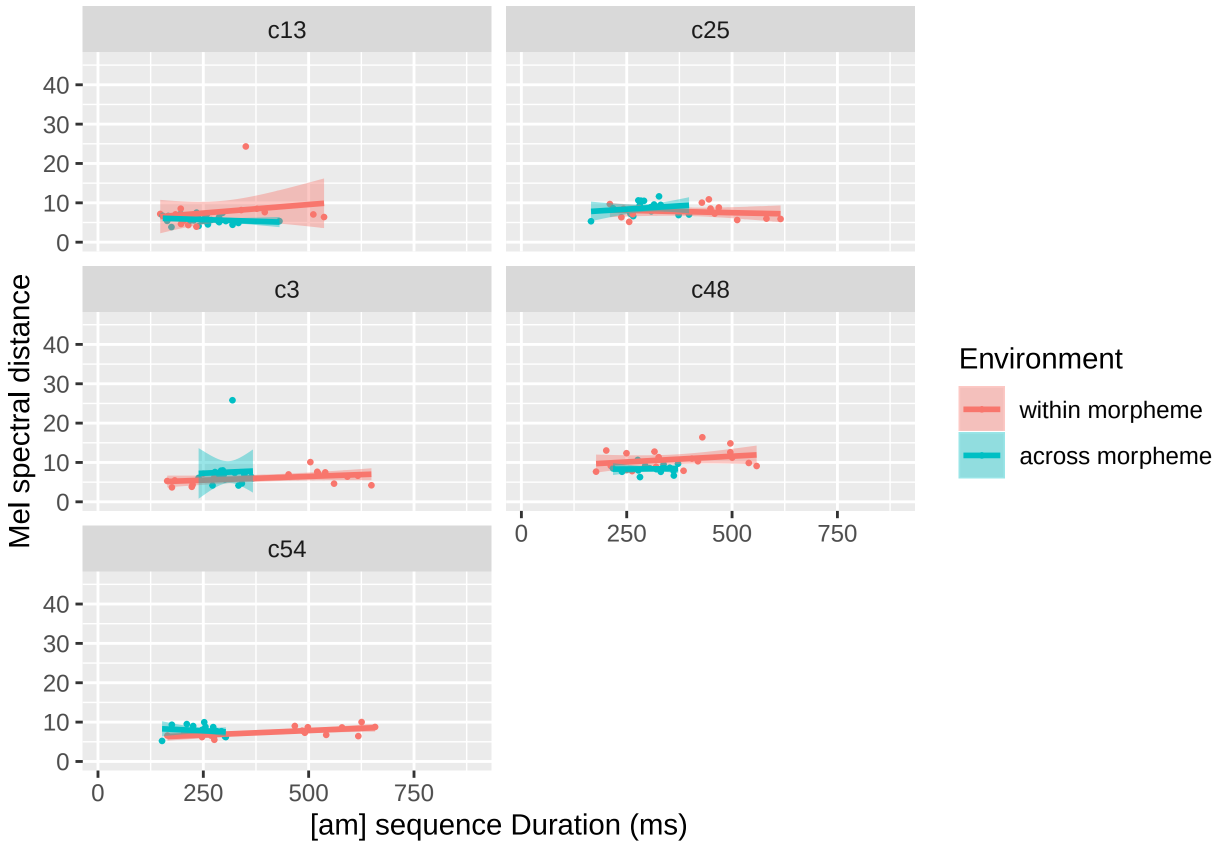


Figure 3.18: Coarticulation by [am] duration, word, and morphological environment in nine-year-old children

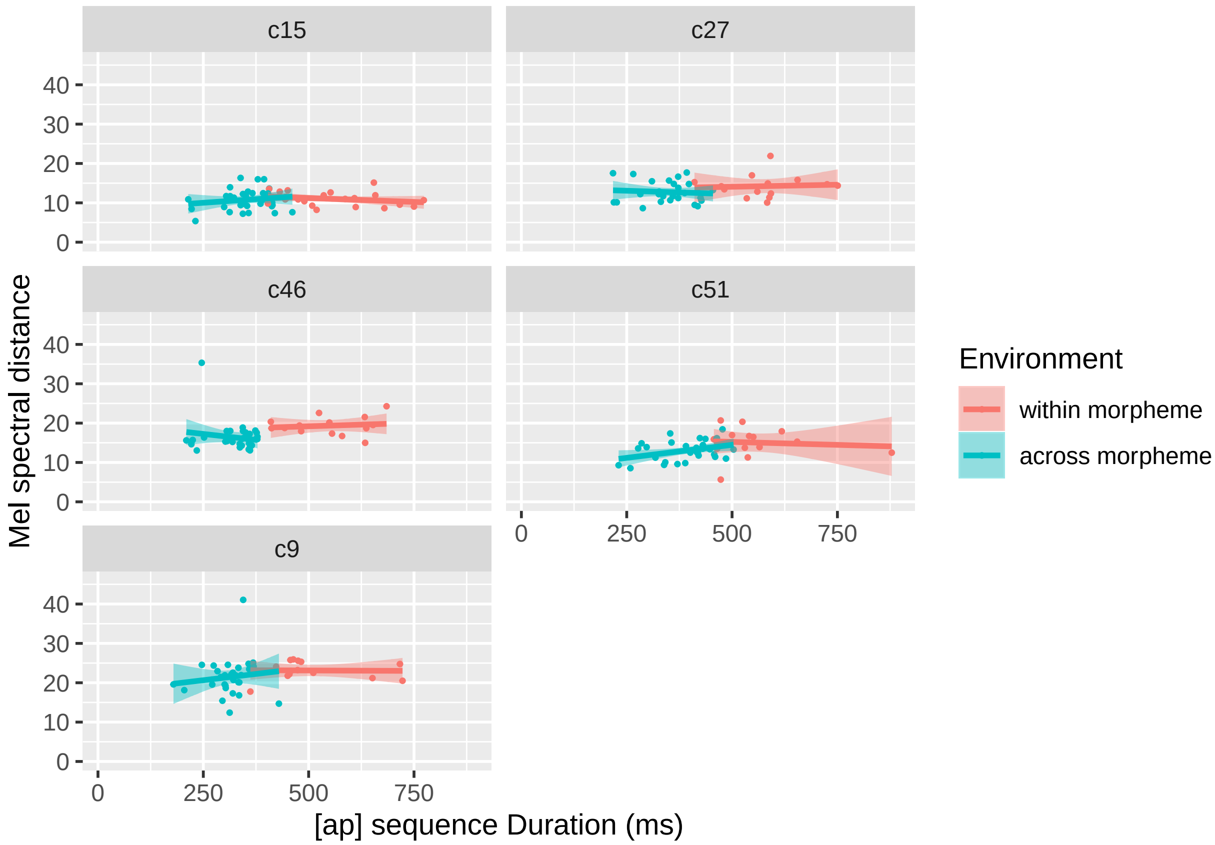


Figure 3.19: Coarticulation by [ap] duration, word, and morphological environment in ten-year-old children

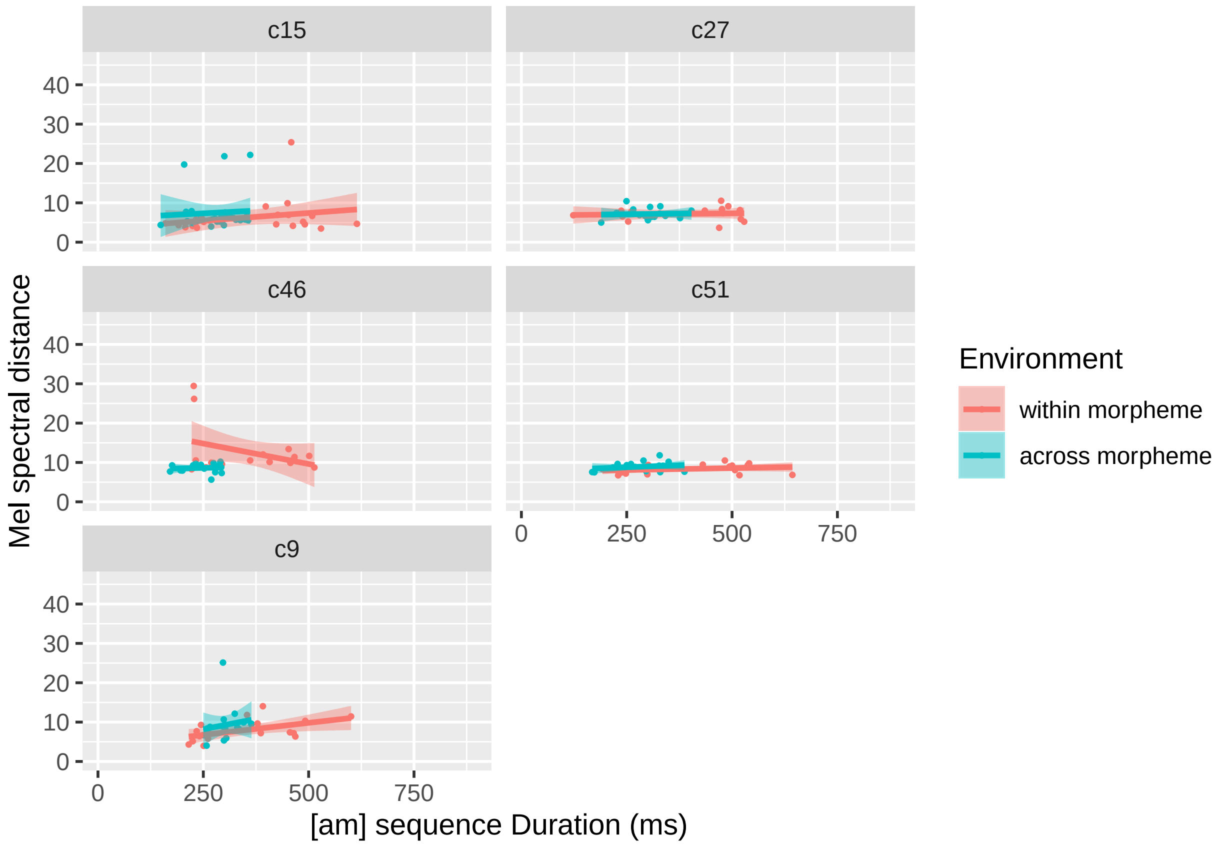


Figure 3.20: Coarticulation by [am] duration, word, and morphological environment in ten-year-old children

Chapter 4

An environmental factor: How bilingual exposure predicts children's speech variation

4.1 Introduction

Young children's speech production is characterized by high intra- and inter-speak variability. Children do not reach adult-like levels of speech variability in domains such as vowel space size, category dispersion, or coarticulation until early adolescence (Gerosa et al. 2006; Lee et al. 1999; Pettinato et al. 2016; Vorperian and Kent 2007; Zharkova et al. 2014). This heightened variability in children has been attributed to a variety of factors such as changes in their perceptual sensitivity to speech categorization (McMurray et al. 2018) and fine motor control (Barbier et al. 2020; Green et al. 2000; Smith and Goffman 1998; Zharkova et al. 2018), as well as anatomical growth (Vorperian and Kent 2007; Vorperian and Wang 2009).

Despite this work, many factors that could potentially account for children's speech variation remain unexplored. In particular, the role of children's language learning environments on their phonetic production has gone almost completely unstudied. The role of the language environment could be highly predictive, however: children's phonetic production patterns could vary based on their receptive language experience, such as the type or quantity of child-directed speech. Children's phonetic patterns could also vary by their expressive language experience, such as how often a child talks or their expressive vocabulary size.

What is the role of the language environment for children's speech variability? And, when evaluating the role of the environment on children's phonetic production, is it relevant to distinguish between receptive and expressive experiences in the environment? Complete answers to these questions require that researchers study children from different language environments - children who have different receptive and expressive experiences. For example, one way to study this could be to measure the phonetic patterns of children with different vocabulary sizes (e.g. Cychosz et al. 2020b; Noiray et al. 2019a) or degrees of phonological awareness (e.g. Caudrelier et al. 2019; Noiray et al. 2018).[1] Another way to respond to these questions could be to study children acquiring two languages. Bilingual children are an ideal population to evaluate the potential roles of receptive and expressive language environments on phonetic patterning because, for bilingual children, the ratio of receptive to expressive language often differs by language. For example, in communities undergoing language shift, children may frequently receive input in the parents' (minority) language - which the children do learn to speak - but the children express themselves at school or with peers in the majority language. As a result, the roles of expressive and receptive language can be evaluated within individual children.

Bearing in mind the potential role of the language environment for children's speech variation, this chapter has two goals. First, it estimates the different bilingual language experiences of Bolivian children simultaneously acquiring Quechua and Spanish. This particular bilingual community is an interesting site to evaluate the roles of expressive and receptive language experience for spoken language variation because it is undergoing rapid language shift. The children's bilingual language exposure is estimated using an ecologically-valid method - daylong audio recordings of children's language environments - that naturalistically captures the children's receptive and expressive language environments. Second, this chapter correlates the differences in the children's bilingual language experiences with their performance on the speech production tasks evaluated in Chapters 2 and 3. As a result, this chapter will demonstrate 1) whether individual differences in bilingual language exposure can predict variability in speech and language production in these children and 2) what types of exposure - receptive or expressive - best predict this variability.

4.2 Background

Environmental effects on language development

The language that children are exposed to in their daily environments has been shown to predict individual differences in speech and language development. The quantity and quality of speech input from adult caregivers predicts children's lexical processing speed (Weisleder and Fernald 2013), syntactic complexity (Huttenlocher et al. 2002), expressive and receptive vocabulary sizes (Hart and Risley 1995; Hoff 2003; Mahr and Edwards 2017), and babbling complexity (Ferjan Ramírez et al. 2019a; Ramírez-Esparza et al. 2014). In recent years, this line of research has increasingly employed a non-obtrusive observational method, daylong audio recordings, which captures the everyday language environments of infants and young children. For this method, the child wears a small, lightweight audio recorder over the course

[1] A related line of research has studied the relationship between children's vocabulary size and repetition accuracy (Cychosz et al. 2020a; Edwards et al. 2004; Munson et al. 2005; Sosa and Stoel-Gammon 2012), but here the discussion is limited to phonetic production outcomes, not accuracy outcomes.

of an entire day, typically 12-16 hours.[2] The recordings can additionally be combined with other modalities such as video (Bergelson et al. 2019a) or photography (Casillas et al. 2019), resulting in an ecologically-valid observation of everything that the child hears and says throughout the day.

Though this work on environmental effects on language development has established a strong correlational relationship between language input and development, two relevant topics in the subfield remain unexplored. First, there is almost no work evaluating a possible relationship between the language environment and children's speech production variability. Yet the relationship between the language environment, especially the quantity of input from adult caregivers, and the many outcome variables outlined above (e.g. lexical processing, babbling) suggests that such a relationship between the environment and phonetic variation is plausible. Second, the language environment is almost always instantiated as input from adult caregivers which leaves open the possibility that the quantity and quality of children's *own* productions could also predict their phonetic variability (DePaolis et al. 2011; Icht and Mama 2015; Zamuner et al. 2018).

To date, only one study that I am aware of has evaluated the predictive role of the environment, and children's own, naturalistic productions in particular, for children's spoken language variation. Cychosz et al. (2020b) measured coarticulation between biphone sequences spoken by four-year-old children. A subset of the children studied also completed a daylong audio recording. The authors found that children with larger receptive vocabularies, and children who vocalized more frequently during their daylong recordings, tended to coarticulate less - a sign of more mature speech development (see Chapter 1). Crucially, the quantity of children's vocalizations was more predictive for their speech patterning than the number of words spoken by adults in the children's environments. This result suggests that the environment does play a role in children's phonetic development. Yet the predictive role of child vocalization quantity suggested that expressive language was more predictive than the receptive measure (adult word count), at least for coarticulatory patterns. This effect is likely to play out differently based on the measure of speech variability (e.g. coarticulation versus vowel category dispersion) and child age.

What is the role of the language environment for children's phonetic production? More specifically, what are the roles of expressive versus receptive language experiences from the environment? Cychosz et al. (2020b) evaluated these questions in a single cohort of four-year-old children who had different receptive (quantity of words spoken by adults) and expressive (quantity of child vocalizations) language experiences. Another way to evaluate the questions of potential environmental effects on production could be to study bilingual speech communities, particularly those undergoing language shift. These communities present an opportunity to evaluate the role of the environment for children's phonetic variation because the quantity and quality of language input and use varies, both by language and between children: most children receive more input in one language than the other and most children

[2]See Mehl and Pennebaker (2003) and Mehl (2017) for details on longform recordings in adult populations.

use one language more than the other. Previous work on bilingual development has indeed found an effect of exposure quantity upon a variety of language development outcomes. Bilingual children who are exposed to more of one or other of their languages tend to show stronger developmental outcomes in that dominant language. The following section outlines the findings from this research.

Exposure effects in bilingual development

The bilingual language learning experience is far from uniform. For bilingual children, learning outcomes, notably vocabulary size, appear to depend upon the relative amount of exposure that the child receives in both of their languages. For example, Pearson et al. (1997) correlated the quantity of Spanish-English input with expressive vocabulary scores in children aged 0;8-2;6.[3] The authors found strong correlations between children's vocabulary sizes in Spanish and English and the amount of input that the children received in each language. Place and Hoff (2011) came to a similar conclusion in another study of Spanish-English bilinguals. Bilingual two-year-old children who were reported to hear more Spanish at home performed better on standardized tests measuring expressive vocabulary size and grammatical complexity in Spanish. And children who were reported to hear more English performed better on similar tests in English.[4]

In a different bilingual population, Carbajal and Peperkamp (2019) found bilingual exposure effects on the receptive vocabulary sizes of French infants aged 0;11. The bilingual infants' language exposure profiles differed widely, with some exposed to more or less French. Overall, these differences were reflected in the infant's vocabulary sizes: more maternal input in the second language resulted in a larger receptive vocabulary in that language. Thordardottir (2011) also found exposure effects on receptive and expressive vocabulary in bilingual children aged 5;0 acquiring French and English. Specifically, that study found a linear relationship between exposure and vocabulary performance. The children who were exposed to more French had larger French vocabularies, especially expressive vocabularies, and vice versa for English.

The quantity of language exposure in bilingual environments also appears to predict young children's speech processing. In a looking-while-listening task, Potter et al. (2019) measured how bilingual English-Spanish children aged 1;6-2;6 processed words in their dominant and non-dominant languages. The children recognized both English and Spanish words embedded in sentences in their non-dominant language. However, when the words were embedded in sentences in their dominant language, they only recognized the words from their dominant language.

[3] Receptive vocabulary information was reported for one child who did not produce any words for the entirety of the family's participation in the study.

[4] For similar conclusions, see also Place and Hoff (2016) who additionally quantified language exposure via number of speakers and language mixing and Byers-Heinlein (2013) who estimated the effect of language mixing on vocabulary development.

Finally, exposure effects have been found for infants' speech perception. Infants aged 0;10 who were bilingual in French and a second language employing lexical stress could discriminate stress contrasts more quickly if they were *less* dominant in French - a language that does not employ lexical stress (Bijeljac-Babic et al. 2012).

The studies summarized above established a correlational relationship between the quantity of exposure in bilingual language environments and a variety of language development outcomes. Yet the outcomes evaluated have been limited to speech perception and processing measures or vocabulary size. It remains unclear if bilingual exposure likewise predicts phonetic variation in children's spoken language. It is possible that bilingual children who are exposed to more of one of their languages, or use more of one of their languages, exhibit less phonetic variability in that language. To that end, the current work estimates bilingual language exposure effects on children's phonetic variability in a bilingual community in southern Bolivia where children's exposure to and use of their two languages varies considerably by household.

Current study

The goal of this chapter is twofold. First, I present a workflow that shows how to process and sample from daylong audio recordings when estimating children's bilingual language environments. This procedure is illustrated using daylong recordings from a corpus of bilingual children acquiring Quechua and Spanish (Cychosz 2018). Second, I use the estimates of children's bilingual environments to predict measures of their phonetic variability, namely within-category vowel dispersion and coarticulation.

I aim to estimate the ratio of Quechua to Spanish to Mixed speech that each child is exposed to in their daily environment. To minimize human annotation, a general sampling-with-replacement technique is employed to annotate portions of each daylong recording. Previous work in the annotation of daylong recordings has consciously sampled from the recording, for example selecting portions from morning, afternoon, and night, that contain high, medium, and low amounts of speech (Casillas et al. 2019; Orena et al. 2019; Orena et al. 2020; Ramírez-Esparza et al. 2017a; Ramírez-Esparza et al. 2017b; Weisleder and Fernald 2013). Random sampling should result in stable estimates of bilingual language exposure, but in a less time-intensive manner.

Using these bilingual exposure estimates, I make one overarching hypothesis for this work: children who are more dominant in Quechua will have more adult-like speech and language patterns. Specifically:

1. Children who are more Quechua-dominant will have tighter, more compact vowel categories in Quechua.

2. Children who are more Quechua-dominant will be more likely to distinguish coarticulatorily between word environments in morphologically complex Quechua words, suggesting that they are analyzing the internal structure of the words.

These hypotheses align with findings on environmental effects on speech processing and vocabulary development: for bilingual children, dominance in one language predicts larger vocabularies and faster speech processing (e.g. Place and Hoff 2011; Potter et al. 2019). It is thus reasonable to propose that bilingual dominance could also predict phonetic production outcome measures.

The children's Quechua experience and dominance will be evaluated in three ways: the percentage of Quechua/Mixed speech from the target child contained in the recording, the percentage of monolingual Spanish speech from the target child contained in the recording, and the mother's language dominance. The percentage of Quechua/Mixed speech and percentage of monolingual Spanish speech are proxies for the children's *expressive* Quechua experience while the mother's language dominance is a proxy for the children's *receptive* Quechua experience. Each of these distinct predictors will be further explained in the Results section. Given that this study is one of the first to evaluate a potential role of the environment for speech variability, no specific hypotheses concerning expressive versus receptive language experience for phonetic variability is made.

4.3 Methods

Participants

Families were recruited through the researcher's personal contacts in communities surrounding a mid-size town in southern Bolivia. Participants ultimately included n=40 children aged 4;0-8;11 (20 girls; 20 boys). Each participant's family reported speaking Quechua at home and lived in the same community during the recording process. See Table 4.1 for further demographic information on participants.

Table 4.1: Demographic information for child participants

Age group	**N**	**Age range**	**Gender**	**N of caregivers w/ $<$ 6 years ed.**
4	5	4;0-4;11	2M; 3F	2
5	7	5;0-5;11	2M; 5F	6
6	8	6;1-6;8	5M; 3F	6
7	14	7;1-7;11	8M; 6F	9
8	6	8;1-8;11	3M; 3F	6

Most children had normal speech and hearing development, per parental self-report. The caregivers of 3 children (2 seven-year-olds, 1 five-year-old) stated that their child was late to begin talking, and caregivers of another 3 children did not report. Note that these communities are medically under-served, so some language delays or impairments may go unreported. Additionally, 3 children had lost one or more of their front teeth at the time of recording. As explained in Chapters 2 and 3, I attempted to complete a hearing test with the children. But it became clear after a few attempts that I was collecting false positives because the children were afraid of making a mistake. Consequently, I cannot say with absolute confidence that all children would have passed a standard hearing screening.

In this sample, 37 children (92.5%) were regularly attending school at the time of data collection. The 3 children (7.5%) who were not attending were four-year-olds, as pre-kindergarten education is available but not compulsory in Bolivia. Most families reported that their children attended school in the morning for an average of 4 hours (range: 3-5 hours). Families of 3 children instead attended school in the afternoon for 6 hours. For ethical reasons, the children did not wear the recorder to school for the daylong recording portion of this experiment. This methodological decision is explained below.

Information on the central caregiver's education level was also collected. There were 30 unique caregivers in the sample due to 8 children being sibling pairs (no twins), and 1 three-sibling group. The distribution of maternal education in the sample of unique caregivers was as follows: 17 caregivers (56.67% of unique caregivers) completed some primary school (less than six years of education), 4 (13.33%) completed primary school (6 years of education), 4 (13.33%) completed the equivalent of middle school (10 years of education), 1 (3.33%) completed secondary/high school (13 years of education), and 3 (10%) had not received any formal schooling. One caregiver did not report. See Chapter 2 for details on mother's education level and knowledge of Spanish as correlates of socioeconomic status in Bolivia.

Caregivers' bilingual language practices

All caregivers in this study spoke Quechua as a first language, and some additionally spoke Spanish, with varying levels of fluency. To get a description of the caregivers' Quechua-Spanish bilingual language practices, the researcher walked each primary caregiver through a brief oral survey (survey included Appendices). For the central caregivers - usually the mother but the grandmother in one family - the level of Quechua-Spanish bilingualism was reported as follows: 7 (23.33% of the 30 unique caregivers) were monolingual Quechua speakers, 4 (13.33%) were Quechua-dominant but spoke/understood some Spanish, 18 (60%) were bilingual Quechua-Spanish speakers, and 1 did not report. For the fathers, the level of Quechua-Spanish bilingualism was as follows: 1 (3.33%) was a monolingual Quechua speaker, 4 (13.33%) were Quechua-dominant but spoke/understood some Spanish, 22 (73.33%) were bilingual Quechua-Spanish speakers, and 3 did not report.

Families were additionally asked about the language practices of the central caregivers' parents, as monolingual grandparents may enhance inter-generational minority language transmission. On this question, 26 caregivers (86.67% of the 30 unique caregivers) reported that both of their parents spoke only Quechua, 2 caregivers (6.66%) reported that their father spoke some Spanish but their mother was a monolingual Quechua speaker, and 2 caregivers (6.66%) reported that both of their parents spoke Spanish and Quechua.

Information on the central caregivers' code-switching habits was also collected. The central caregiver responded to the following four questions on code-switching behaviors:

1. Do you start sentences in Spanish and finish them in Quechua? That is to say, do you use both languages in the same sentence?
2. Do you start sentences in Quechua and finish them in Spanish?
3. Do you use Quechua words when you speak Spanish?
4. Do you use Spanish words when you speak Quechua?

Responses to these questions are listed in Table 4.2. The questions were generally not applicable to the 7 monolingual caregivers.

Table 4.2: Primary caregiver responses to bilingual language practice survey questions

	Yes	No	No response/NA
Do you start sentences in Spanish and finish them in Quechua?	17 (56.67%)	5 (16.67%)	8 (26.67%)
Do you start sentences in Quechua and finish them in Spanish?	17 (56.67%)	5 (16.67%)	8 (26.67%)
Do you use Quechua words when you speak Spanish?	18 (60%)	4 (13.33%)	8 (26.67%)
Do you use Spanish words when you speak Quechua?	17 (56.67%)	6 (20%)	7 (23.33%)

Procedure

There were two phases in the the experimental procedure: the daylong recording collection and the word elicitation tasks. For both phases, families were visited in their homes or in a central area in the community to explain the experimental procedure.

Daylong recordings

The daylong recordings reported here come from a larger corpus of nearly 100 children, aged 0;6-8;11, simultaneously acquiring Quechua and Spanish (Cychosz 2018). The current study reports on recordings from 40 children, aged 4;0-8;11. To collect the daylong recordings, families were given a small, lightweight recorder: either a 3"x5" Language ENvironment Analysis Digital Language Processor (Greenwood et al. 2011), or a 2"x5" Zoom H1n Handy recorder. To explain the recording procedure to participants, the researcher demonstrated how to turn the recorder on and off and how to pause the recording, among other functions. To obtain fully informed consent for the daylong recording, the researcher explained the radius that the recorder could capture and that the families had the option to delete the recording after completing it. Families were encouraged to ask questions, practice using the recorder, and make sample recordings to become familiar with the technology. Per university Institutional Review Board specifications, families were permitted to pause the recording whenever they wanted. Additionally, families were instructed to either remove the recorder or pause it when the child attended school and when the child was sleeping. In practice, some families forgot to comply and so prior to annotation, an additional pre-processing step was taken to identify portions of the recording where the child might be sleeping (described in further detail below).

Children were not required to wear the recorder to school for a couple of reasons. First, the children's schools were led almost entirely in Spanish, and the children spoke Spanish at school (the lead researcher volunteers at a primary school in the community and has observed the language practices of children in the community at school). Second, there was no reliable way to obtain informed consent from everyone who might appear on the recording during the school day. Since all of the school-aged children in this study had similar educational experiences in terms of the number of hours spent in a Spanish-dominant school (see Section 4.3), there was no need to sample school language usage.

After explaining the daylong recording procedure, families were given a small cotton t-shirt. Each t-shirt had a cotton pocket sewed to the front with a Velcro or snap-button flap to close the pocket and hold the recorder inside (Figure 4.1). Families were told to record for at least 12 hours, at which time they could stop the recording. These 12 hours could be non-consecutive, since families were allowed to pause the recording and children could not wear the recorder to school.

Most families completed three daylong recordings.[5] To collect the recordings, families were visited on three different days. At each visit, the researcher checked that the families had completed the recording and exchanged the previous shirt and recorder for a clean shirt and empty recorder. In all, 39 children (98%) successfully completed at least three daylong recordings, while 1 child only completed one recording because the child and his family left for a trip to another community after the first recording day.

Word elicitation tasks

Following the daylong recording procedure, each child participant completed a series of picture-prompted word elicitation tasks in the following order: 1) real word repetition, including a morphological extension component that is explained in Chapter 3, 2) Quechua nonword repetition, 3) Spanish nonword repetition, and 4) additional real word repetition with morphological extension. For all tasks, children repeated the real words or nonwords

[5] Families of children under 1;0 usually completed just one daylong recording but these data are not discussed in this chapter.

(a) Example shirt used to house recorder during data collection.

(b) The LENA recorder was worn inside of the cloth pocket for the duration of the recording.

Figure 4.1: Daylong recording collection materials.

after a pre-recorded model speaker. Nonword repetition tasks are not discussed in this dissertation.

Four children completed a pilot version of the morphological extension task. Consequently only the vowel data from these children will be considered in the current work. These children also did not complete the nonword repetition tasks. Another exception to the testing procedure was for the 5 children in the 4;0 group; in pilot testing, I had previously found that children of this age could not reliably complete the morphological extension task.[6] Consequently, the 4;0 group completed the following tasks in this order: 1) real word repetition, without morphological extension, 2) Quechua nonword repetition, and 3) Spanish nonword repetition. The entire word elicitation procedure took approximately 30-40 minutes per child. For the daylong recordings and word elicitation tasks, each family was compensated with a small monetary sum. The families also kept the t-shirt and the children could pick one item from a toy bag.

The stimuli and procedure for the word elicitation tasks are explained in Chapters 2 and 3. Data cleaning and formant tracking information for the vowel data is detailed in

[6]This is not to say that these children were not morphologically productive. They just had a much harder time understanding the task.

Chapter 2 and the acoustic coarticulation measures are explained and validated in Chapter 3. Inter-rater annotation scores for the speech production task data are likewise reported in Chapters 2 and 3. Vowel data normalization and additional cleaning procedures are described in Section 4.4 of the Results in this chapter.

Data analysis

Recording selection

Because each child completed multiple daylong recordings, the first step in data analysis was to select one recording per child. To get the best estimate of each child's environment, the research team decided to annotate the longest duration recording for each child. In one case, it was not possible to use the longest duration recording because the child went to sleep with the recorder on, resulting in almost all of the recording occurring when the child was not awake. In that case, the second-longest recording was used. In almost all cases, the recordings selected for annotation were made with the LENA digital language processor. Two children's annotated recordings were instead made with the Zoom H1n Handy recorder either because the child only completed one day of recording and it was done with the Zoom recorder (n=1) or because the child's LENA recording contained the child sleeping overnight (n=1). The average duration of the daylong recordings used for bilingual language estimation was 12.12 hours (range 7.63-16 hours), with no notable durational outliers within any age group (Table 4.3).

Table 4.3: Daylong audio recording information by age group

Age group	**Average recording duration (hrs.)**	**Range (hrs.)**	**Average # of potential 30-second clips to annotate**
4	12.17	8.77-16	1400
5	12.29	8.92-16	1454
6	12.12	7.63-16	1369
7	12.13	9.48-16	1469
8	11.00	11.24-13.75	1274

Sampling

As described in the background literature, researchers who collect daylong recordings rarely transcribe the recordings in their entirety. Researchers instead transcribe consciously-sampled portions of full recordings. Some laboratories have, for example, selected audio samples from different parts of the day within the recording (e.g. Weisleder and Fernald

2013), samples containing large numbers of words spoken by adults (e.g. Ferjan Ramírez et al. 2019b), or samples with large amounts of infant vocalizations (e.g. Casillas et al. 2019).

Instead of conscious sampling, this study uses random sampling to select portions of the recording for annotation. It was hypothesized that this method would result in the most efficient, least biased, estimation of bilingual language exposure. The entire recording selection and annotation workflow used is outlined in Figure 4.2.

First, a series of custom Python scripts were written to process the recordings. Since families were permitted to pause the recording, the first step was to stitch together the pieces of the recording to create one longer recording. There were, on average, 2.12 recording "pieces" that made up each recording (range: 1-4). The stitched pieces were interspersed with a 1000ms clip of white noise to mark boundaries between recording pieces.

After stitching the recording pieces together to create a longer recording, the entire recording was then chopped into 30-second clips. (See Table 4.3 for description of number of clips derived from the recordings by age group). The 30-second clip length was chosen because previous work on daylong recordings has annotated 30-second clips from recordings to estimate bilingual language exposure (Orena et al. 2019) and child-directed speech (Ramírez-Esparza et al. 2014).

Once the recordings were chopped, a standard vocal activity detector (Usoltsev 2015) was run over all of the clips and the percentage of the clip containing vocal activity was reported. In practice, the vocal activity metric served two purposes. First, clips that contained 0% vocal activity were not drawn for annotation. Second, prior to annotation, the lead researcher listened to portions of the recordings containing extended stretches of low vocal activity to determine if the child was napping. The researcher determined if the child was napping by listening for relative quiet in the background, lack of speech from the target child, and heavy breathing or snoring. If the researcher found that a clip contained audio of a sleeping child, the clip was marked not to be drawn for annotation.

Annotation

To facilitate annotation of the 30-second clips by the research team, a custom Generalized User Interface (GUI) application was built. (See `https://github.com/megseekosh/Categorize_app_v2` for further details on GUI application structure.) The GUI application led the research personnel through the steps of clip annotation. First, a 30-second clip was randomly selected, with replacement, from a given participants' clips. Again, to speed up annotation, clips were drawn but not annotated if they had 0% vocal activity or contained a child determined to be sleeping. Clips that contained speech from a researcher (for example when the recording was turned on as the researcher was leaving the participant's home) were also identified prior to annotation and were likewise drawn but not annotated.

After drawing a clip, the researcher would listen to the clip and then categorize the speaker(s) and language(s) heard. Research personnel had the option to repeat the clip as many times as they would like. The annotation categories were selected from a drop-down menu containing the following choices:

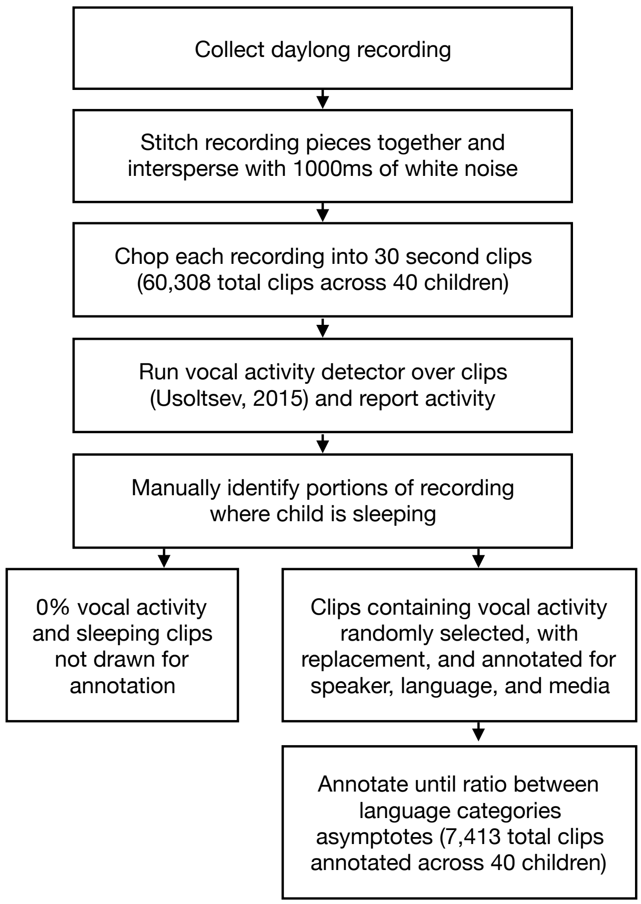


Figure 4.2: Audio clip generation, selection, and annotation workflow.

1. **Language?**: Quechua, Spanish, Mixed, No speech, Personal Identifying Information (PID), Researcher Present, or Unsure

2. **Speaker?**: Target Child, Target Child & Adult, Other Child, Other Child & Adult, Adult, or Unsure

3. **Media Present?**: Yes or No

For the language choice, research personnel were instructed to identify the language being spoken in the clip. If only Spanish was spoken (regardless of the quantity of speech), the researcher marked 'Spanish.' Similarly, the researcher marked 'Quechua' for monolingual Quechua clips. If the research personnel heard both Quechua and Spanish in the clip - either code-switching within a sentence or two separate conversations - they marked the clip as 'Mixed.' For those clips where the speaker or language was not clear, research personnel could select 'Unsure' for either the language category, speaker category, or both. In practice, the 'Unsure' annotation was used most often for clips where a conversation was taking place in the background, far from the child, making it difficult to determine the language and/or speaker categories.[7] Annotators had the option to select 'Unsure' for one category, such as speaker, while still annotating for another category, such as language.

The team defined personal identifying information (PID) as any clip containing first *and* last names (but not just first), street addresses and specific neighborhoods (but not just the name of the town), birth dates, and any discussion of financial information. The 'Researcher Present' annotation indicated that the researcher could be heard on the recording. Finally, if the clip did not contain any speech, the annotator labeled it 'No Speech.' If PID, Researcher Present, or No Speech were selected for the language choice, the annotator did not continue to annotate for speaker. The percentage of the major language annotation categories by individual child is reported in the Results section.

For the speaker annotation, the 'Target Child' was the child wearing the recorder and 'Other Child' was defined as any individual whose voice sounded as though they had not gone through puberty.[8] Personnel were instructed to annotate 'Target Child and Adult,' if a clip contained the target child, another child, and an adult. See `https://github.com/megseekosh/Categorize_app_v2/blob/master/FAQs.MD` for further details on annotation decisions, including a list of frequently asked questions used to standardize annotation between research personnel.

[7]The research team considered the possibility that it may be difficult to ascertain the speaker or language in some clips because those clips are noisier and contain multiple interlocutors. These clips could also be more likely to occur outside of the home. These noisy clips, with multiple interlocutors, might be more likely to contain mixed speech. Thus, disregarding these clips *could* lead the team to inadvertently disregard clips of a certain category (i.e. Mixed speech). In practice, however, the 'Unsure' clips almost always contained background speech without a discernible speaker or language, so the team felt confident in excluding 'Unsure' clips from further analysis.

[8]Usually the team could determine whether a speaker was a child or an adult because they had information on the household members and their ages. However, in the cases where an annotator could not recognize a voice, the team labeled the speaker as a child if their voice sounded pre-pubescent.

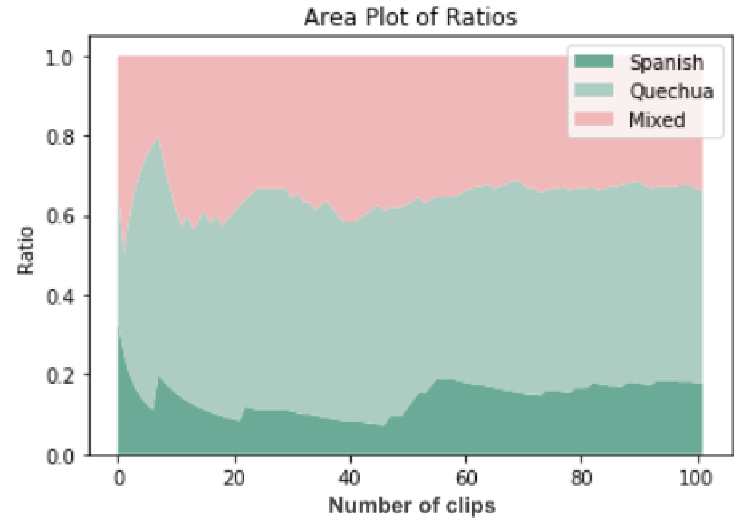


Figure 4.3: Example area plot of ratios between languages by number of clips annotated. Area plots were used to track progress towards language ratio stability during daylong recording annotation.

The choice for media was binary - 'Present' or 'Absent' - because 1) it was often difficult to determine if the media was radio or TV and 2) almost all of the media was in Spanish, making it irrelevant to mark the language of the media. In other words, when media was present, it was in Spanish.

As annotators drew and listened to the 30-second clips, they were simultaneously running a Jupyter notebook (included in the Github repository) that quantified their progress towards annotation. Specifically, after each clip was annotated, a ratio of Quechua to Spanish to Mixed clips was updated for the child. Variance in language assignment was also measured over a moving window of 60 language ratio estimates. Human annotation was cut off when 1) the ratio between language categories asymptoted (exemplified in Figure 4.3) and 2) the variance in language ratios asymptoted (exemplified in Figure 4.4).

As a precautionary measure, since a complete validation of random sampling to annotate daylong recordings is still underway, the team additionally determined that a minimum of 50 language clips must be annotated from each child. Language clips include those annotated as Quechua, Spanish, or Mixed, but not Unsure or No Speech. Given all the pre-determined criteria, the team was more confident that their annotations were accurately reflecting the child's language environment.

A complete validation of the random sampling technique used in this study is outside the scope of this dissertation.[9] Nonetheless, the research team was able to make stable

[9] A validation of the random sampling technique for bilingual language estimates and child-directed

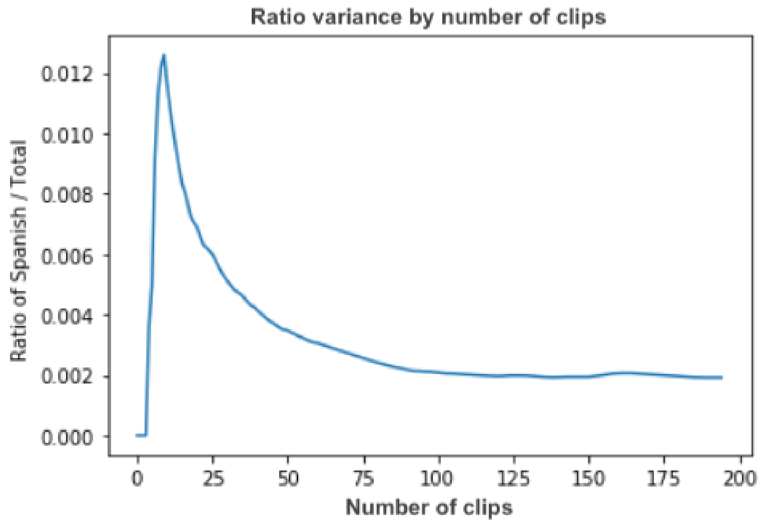


Figure 4.4: Example plot of language ratio variance by number of clips annotated. Variance was computed over a moving window of 60 clips. This plot was used to track progress towards variance stability during daylong recording annotation.

estimations of each child's bilingual language exposure (the ratio of monolingual Quechua to monolingual Spanish to mixed Quechua and Spanish) by listening to and annotating an average of 185.3 30-second clips from a given recording (sd=69.72; range=84-385), or an average of 92.66 minutes total from each recording.[10] Given that the children's recordings varied in length (see Table 4.3), the annotated clips made up an average of 13.13% of each recording (sd=5.47; range=4.38-29.90%). Thus, the number of clips annotated for a given child varied as a function of the unpredictability of language categories in the child's environment. But the criterion for variance between the annotated categories was the same for all children. Overall, this procedure resulted in the annotation of a total of 3,706.5 minutes, or 61.78 hours, across the 40 children.

Personnel

Three undergraduate student research assistants and the lead researcher annotated the daylong recordings. All research assistants were fluent Spanish speakers participating in

speech estimates is underway (Cychosz et al. 2020c).

[10]The figures reported in the text reflect those clips that annotators *actually* listened to. Actually, a grand total of 8,974 clips were drawn, including those that were not listened to because they did not have any vocal activity, the child was sleeping, or the researcher was talking. This amounted to an average of 224.35 30-second clips from each recording (sd=107.06; range=91-618), or an average of 112.15 minutes total from each recording. The number of clips that were listened to are reported in the text because those figures more accurately reflect the time commitment required for annotation.

a linguistics research training program. The annotation personnel underwent a stringent training procedure prior to and during annotation. First, personnel spent approximately three hours practicing annotating 30-second clips from a single recording (the recording was from a child who was not a participant in the current study). After this initial practice, research personnel had to pass an annotation test. The lead researcher selected and annotated 40 30-second clips from the same practice recording. The lead researcher's annotations were considered the gold standard annotations (the lead researcher knows the families and is familiar with both Spanish and Quechua). The clips were selected to represent an array of situations that the research assistants would eventually encounter in the recordings (e.g. no speech, overlapping speech, presence of multiple interlocutors of different ages). The research assistants coded the test clips for Speaker, Language, and Media. The assistants' annotations were compared to the gold standard annotations. Research assistants could not begin annotating recordings for the project until they passed the annotation test with a score of 35/40 correct annotations.

Once the research assistants began annotating, weekly or bi-weekly check-ins were conducted. At these check-ins, the research personnel would discuss clips that they found difficult to annotate and a lengthy list of "Frequently Asked Annotation Questions" was constructed for the team to further standardize annotation between team personnel (see link to FAQ list in Section 4.3). The check-up meetings were also used to listen to clips together and discuss annotation choices to be made.

Inter-rater reliability scores between the lead researcher and all personnel members were calculated to ensure fidelity to the coding scheme. For this inter-rater reliability check, 72 clips were randomly selected from one participant's recording.[11] Each personnel member then annotated the clips according to the established annotation procedure. The inter-rater reliability between each personnel member (lead researcher and three assistants) and the remaining team was as follows: 94.44% agreement (lead researcher), 93.06% agreement, 94.44% agreement, and 98.61% agreement for each of the three assistants (Krippendorff's alpha=0.87 for the entire team). The team found this inter-rater agreement satisfactory and concluded that all team members had been sufficiently calibrated.

Intra-rater reliability was also collected for all personnel involved in annotation: the lead researcher had 99.17% intra-rater agreement (Krippendorff's alpha=0.99), research assistant one had 97.62% agreement (Krippendorff's alpha=0.93), research assistant two had 99.29% agreement (Krippendorff's alpha=0.93), and research assistant three had 100% agreement (Krippendorff's alpha=1.0). In all, these inter- and intra-rater agreement scores were satisfactory to conclude that raters were annotating uniformly.

[11]Due to a bug in the annotation script, 72 clips were annotated by all four team members, though 75 clips were originally selected.

4.4 Results

The first section of the Results presents descriptive analyses of the ratios of Quechua, Spanish, and mixed Quechua-Spanish speech clips (henceforth 'Mixed') from the daylong recordings. This descriptive analysis includes reports of the number of clips labeled Unsure, as well as No Speech (not containing any vocal activity). These analyses quantify the variation in bilingual language exposure between children, as well as how this exposure varies by categories, such as age or the language dominance of the primary caregiver. The second part of the results section examines how individual differences in language exposure predict the language production outcomes outlined in Chapters 2 and 3: within-category vowel dispersion and speech production within and between morpheme boundaries. It is expected that children who use and speak more Quechua will have tighter vowel categories and will be more likely to distinguish between morphological environments in their speech production.

All analyses were conducted in the RStudio computing environment (version: 1.2.5033; RStudio Team 2020). Data visualizations were created with `ggplot2` (Wickham 2016). Modeling was conducted using the `lme4` (Bates et al. 2015) and `lmerTest` (Kuznetsova et al. 2017) packages, and summaries were presented with `Stargazer` (Hlavac 2018). The significance of potential model parameters was determined using a combination of log-likelihood comparisons between models, AIC estimations, and p-values procured from model summaries.

Descriptive analyses of bilingual language exposure

Table 4.4: Clip annotation category counts and percentages by maternal language profile.

	Monolingual Quechua mothers (n=10 children)	Quechua-dominant mothers (n=6 children)	Bilingual Quechua-Spanish mothers (n=23 children)
Mixed	308 (14.6 %)	114 (8.19 %)	568 (14.96 %)
No speech	347 (16.45 %)	349 (25.07 %)	554 (14.59 %)
Quechua	381 (18.07 %)	163 (11.71 %)	353 (9.29 %)
Spanish	782 (37.08 %)	496 (35.63 %)	1459 (38.41 %)
Unsure	291 (13.8 %)	270 (19.4 %)	864 (22.75 %)

The first analyses describe the children's ambient language exposure. Figure 4.5 shows the distribution of the language annotation categories Quechua, Spanish, and Mixed, as well as clips annotated as Unsure or No Speech, by the central caregiver's language profile (see also Table 4.4). (Henceforth the central caregiver will be referred to as the mother, though one of the caregivers was the child's grandmother.) There were three maternal language profiles

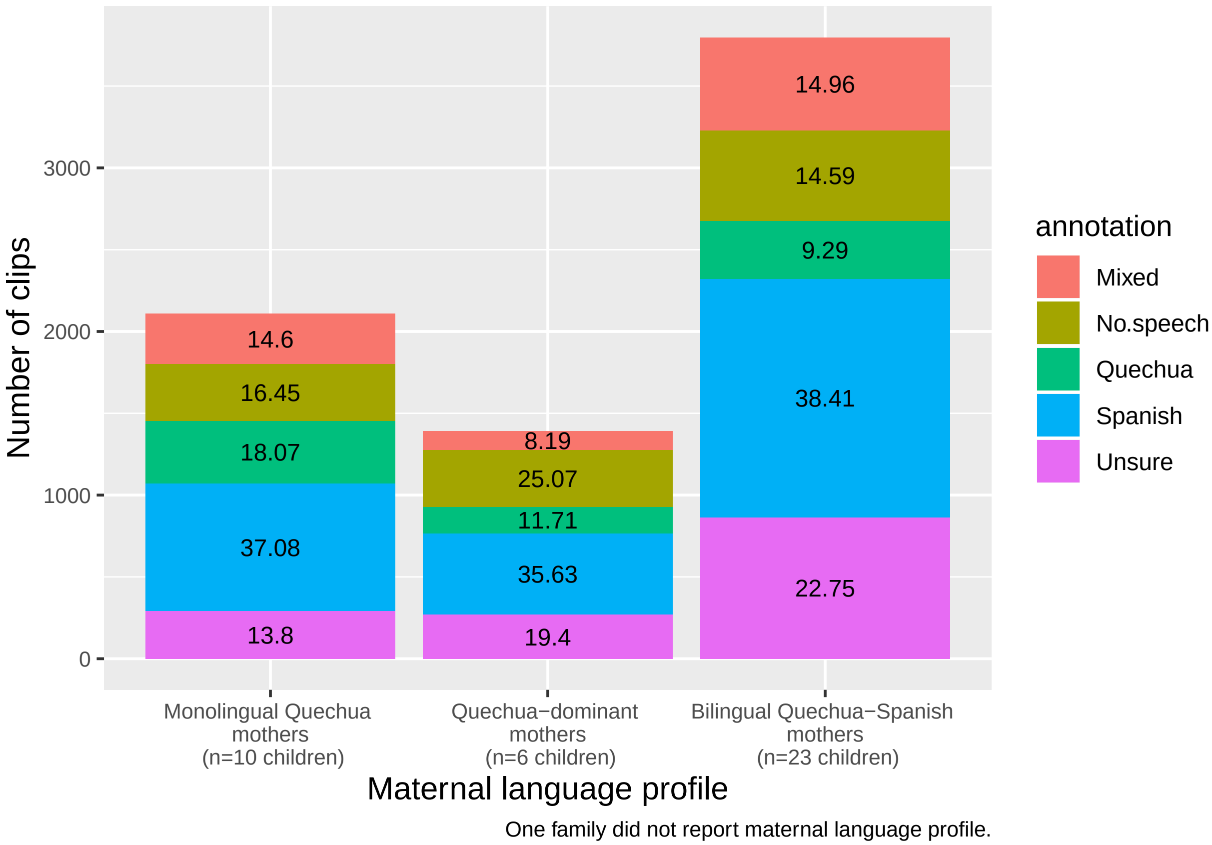


Figure 4.5: Clip annotation category counts by maternal language profile. Percentages listed on barplot.

analyzed: children of mothers who were monolingual Quechua speakers (n=10), children of mothers who were Quechua-dominant (n=6), and children of mothers who were bilingual Quechua-Spanish (n=23). One family did not report the mother's bilingual language profile. As the percentages in the figure and table demonstrate, there was considerable variation between maternal language profiles. Unsurprisingly, there was a larger number of Quechua clips heard in the recordings of the children with monolingual mothers (381 (18.07 %)), than Quechua clips heard from children with Quechua-dominant mothers (163 (11.71 %)), or bilingual Quechua-Spanish mothers (353 (9.29 %)). This suggests that children with monolingual mothers are exposed to more Quechua than children with Quechua-dominant or bilingual mothers. However, the overall percentage of Spanish clips did not vary greatly by language profile. See Figure 4.17 and Table 4.19 in the Appendices for a distribution of clip annotation categories and percentages for each individual child.[12]

[12]Clips containing personal identifying information accounted for between 0 and 0.04% of the total clips from each recording and are not further reported in the analyses.

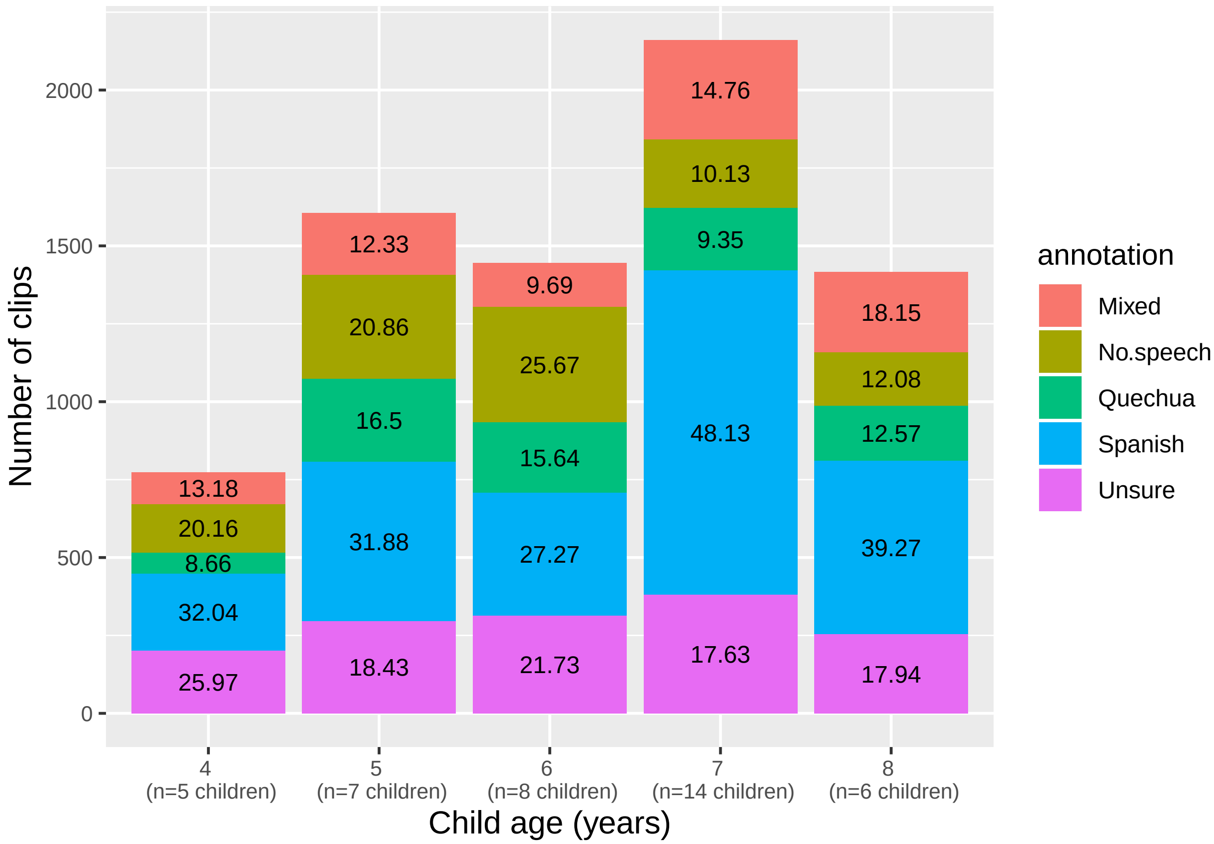


Figure 4.6: Clip annotation category counts by child age (in years). Percentages listed on barplot.

Table 4.5: Clip annotation category counts and percentages by child age (in years).

	Four-year-olds (n=5 children)	Five-year-olds (n=7 children)	Six-year-olds (n=8 children)	Seven-year-olds (n=14 children)	Eight-year-olds (n=6 children)
Mixed	102 (13.18 %)	198 (12.33 %)	140 (9.69 %)	319 (14.76 %)	257 (18.15 %)
No speech	156 (20.16 %)	335 (20.86 %)	371 (25.67 %)	219 (10.13 %)	171 (12.08 %)
Quechua	67 (8.66 %)	265 (16.5 %)	226 (15.64 %)	202 (9.35 %)	178 (12.57 %)
Spanish	248 (32.04 %)	512 (31.88 %)	394 (27.27 %)	1040 (48.13 %)	556 (39.27 %)
Unsure	201 (25.97 %)	296 (18.43 %)	314 (21.73 %)	381 (17.63 %)	254 (17.94 %)

Figure 4.6 presents the distribution of clip annotations by child age group, in years (age 4-8) (See also Table 4.5). As these data demonstrate, there was not a large discrepancy between age groups in clips annotated as Unsure: for all age groups, between approximately 18 and 26% of total clips were marked as Unsure. The data also suggest that language environments

Table 4.6: Maternal language profiles by child age (in years).

Age	Monolingual Quechua	Quechua-dominant	Bilingual Quechua-Spanish	Total
4	0	1	4	5
5	2	2	3	7
6	2	1	5	8
7	2	2	9	13
8	4	0	2	6

become more verbal as children age, since far fewer clips in the seven- and eight-year-old groups did not contain any speech. Larger differences across the age groups were apparent in the language category annotations. In the seven-year-old group, for example, 1040 (48.13 %) clips were annotated as Spanish, compared to just 394 (27.27 %) clips in the six-year-old group and 248 (32.04 %) clips in the four-year-old group. And the largest percentage of Quechua clips was apparent in the five-year-old group.

While there are too few children within each age group to evaluate the effects of maternal language profile by age, a larger proportion of the seven-year-old group had bilingual Quechua-Spanish mothers (Table 4.6). Of the 13 children in the seven-year-old group, 9 had bilingual mothers compared to 3 of the 7 children in the five-year-old group. Nevertheless, the maternal language profile cannot entirely explain the differences between the age groups: 4 out of the 5 children in the four-year-old group had bilingual mothers, but 248 (32.04 %) of their clips were labeled as Spanish.

Overall, these results demonstrate some of the individual differences in ambient language exposure in these bilingual speech communities. The following section will correlate these individual differences with children's speech production outcomes to view how language exposure and practice can affect children's phonetic production and morphological analysis.

Correlating language dominance and speech production

After measuring the children's ambient language profiles, the second objective of the study was to evaluate if and how characteristics of the bilingual environment correlate with speech production patterns. The children's bilingual profiles were correlated with two fine-grained speech production metrics: 1) within-category vowel dispersion and 2) coarticulation between, and duration of, biphone sequences that fell across and within morpheme boundaries in Quechua words. These speech production metrics are derived from the word elicitation tasks described in Chapters 2 and 3. As stated in the Methods, all 40 of the child participants completed a task that elicited vowel measurements. A subset of the children (n=31) completed the morphological extension task. None of the children (n=5) in the 4;0 group completed the morphological extension task. And 3 children (aged 6;8, 7;2, and 8;1) completed an earlier pilot version of the morphology task, so their data from the task were not analyzed.

For the vowel measurements, it is anticipated that children who are more Quechua-dominant will have tighter, more compact vowel categories than children who are exposed to more Spanish. For the coarticulation and duration measurements by morphological environment, it is likewise anticipated that children who are more Quechua-dominant will be more likely to distinguish coarticulatorily and durationally between word environments. In other words, Quechua-dominant children will show a larger difference in their coarticulation and duration patterns in the across-morpheme and within-morpheme condition than more Spanish-dominant children.

Vowel category dispersion

After the word elicitation data were transcribed, a series of steps were taken to prepare the vowel data for analysis. Because this study addresses variation, the number of observations from each speaker was standardized to the extent possible. First, speaker vowel categories with fewer than four F1-F2 observations were removed from analysis (e.g. fewer than four observations of [i] from a given speaker). These categories were removed because category dispersion estimates made over two or three tokens are not likely to be representative of a speaker's variation

The number of tokens per vowel category differed between speakers because of data cleaning procedures and occasional wind interference in the recordings (see Chapter 2). Under this criterion, seventeen peripheral vowel categories were removed (see Table 4.7 for distribution by age group and peripheral vowel) and 29 mid-vowel categories were removed (Table 4.8). This data removal is substantial for some vowels, notably [o], but was preferable to estimating RMS on the basis of one or two vowel tokens.

Table 4.7: Number of vowel sets removed by age and phone to standardize measurements across age groups.

Age	a	i	u
4	NA	1	4
5	NA	NA	1
6	1	1	2
7	1	NA	3
8	1	NA	2

To further standardize the number of vowel measurements between children, a random subset of 10 observations was selected for those speaker vowel categories with more than 10 observations. In this way, no individual child contributed more than 10 or fewer than 4 data points for a given vowel, making the measurements between children more uniform.

Table 4.8: Number of categories removed for statistical modeling by phone.

Phone	Categories Removed
e	12
o	17

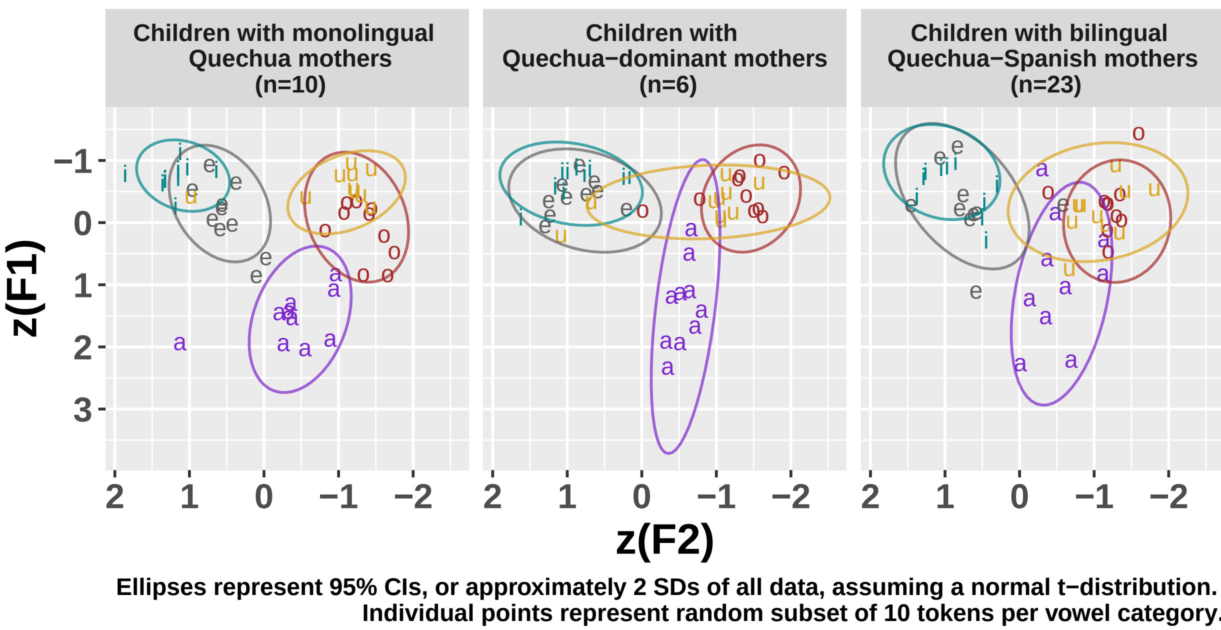


Figure 4.7: Children's vowel spaces by maternal language profile.

The second preparatory step was normalizing the vowel data in order to compensate for anatomical differences between children. This was especially important given the children's different ages, and thus vocal tract lengths. The Lobanov vowel normalization technique, which is essentially z-score normalization, was used to factor out individual anatomical differences. This normalization technique is a vowel-extrinsic, meaning that it takes into account information from all available vowels and multiple formants for the normalization of each formant measurement.[13]

Figure 4.7 plots the children's cleaned and normalized vowel data for Quechua's three phonemic vowels /a, i, u/ and two allophonic vowels [e, o] by maternal language profile. See also Table 4.9 for descriptive statistics of the vowel dispersion by maternal language profile and Table 4.10 for the number of vowel measurements by maternal language profile. As Figure 4.7 and Table 4.9 demonstrate, the children's vowel variability differs by their

[13]Note that as Chapter 2 concludes, even non-uniform formant scaling techniques cannot factor out some developmental differences between children, such as compensation for vocal tract morphology.

Table 4.9: Average and standard deviation of vowel category dispersion (RMS) by phone and maternal language profile.

	Mean (SD) [a]	[e]	[i]	[o]	[u]
Monolingual Quechua	0.69 (0.47)	0.61 (0.36)	0.44 (0.27)	0.63 (0.37)	0.58 (0.41)
Quechua-dominant	1.03 (0.59)	0.7 (0.3)	0.6 (0.31)	0.6 (0.39)	0.85 (0.55)
Bilingual Quechua-Spanish	0.91 (0.57)	0.78 (0.51)	0.63 (0.5)	0.7 (0.4)	0.84 (0.6)

Table 4.10: Number of tokens by vowel category and maternal language profile.

	[a]	[e]	[i]	[o]	[u]
Monolingual Quechua	65	43	80	44	35
Quechua-dominant	41	21	54	22	19
Bilingual Quechua-Spanish	173	95	174	63	69

mother's language profile. Specifically, children with monolingual mothers appear to have tighter, less variable vowel categories in Quechua: the average RMS of the children with monolingual mothers was consistently smaller than the average RMS of the children with Quechua-dominant or bilingual mothers. The only exception to this pattern was the vowel [o], which had an average RMS of 0.6 (0.39) for the children with Quechua-dominant mothers and an average RMS of 0.63 (0.37) for the children with monolingual Quechua mothers. The standard deviation of the RMS also tended to be smaller for the children with monolingual mothers, suggesting that, as a group, they had more uniform vowel patterning.

There were also differences between the children with Quechua-dominant mothers and those with bilingual mothers. For the vowels [e] and [i], children with bilingual mothers had more expansive vowel categories. The dispersion of the [a] category was larger for the Quechua-dominant group. The [u] category was slightly more variable for the children with Quechua-dominant mothers than bilingual mothers. When considering differences between language profile groups, it is important to note that the group with bilingual mothers had almost four times as many children (n=23) as the group with Quechua-dominant mothers (n=6), and more than twice as many children as the group with monolingual Quechua mothers (n=10). Still, the differences in vowel patterning suggest that the mothers' language dominance has an effect on the children's vowel variability.

As the vowel plots in Figure 4.7 show, there was considerable overlap between the allophonic vowels [e] and [o] and their underlying phonemic forms /i/ and /u/, respectively. It is noteworthy that this overlap does not appear to qualitatively differ by maternal language profile. One could expect children with bilingual mothers to have more distinct [e] and [o] categories since these vowels are phonemic in Spanish, which may be the children's dominant language. But this is not the case. There are similar amounts of overlap between the allophonic and phonemic vowels across the three maternal language profiles. This finding suggests that children with bilingual mothers employed a Quechua vowel system during the task. (See Figures 4.19-4.23 in the Appendices for individual vowel plots from each child.)

To further evaluate vowel patterning differences among the three maternal language profiles, a series of linear mixed effects models were fit to predict the children's vowel variability. The objective of this modeling was to determine if the observed differences remained after controlling for phone and individual speaker. The dependent variable in these models was the RMS of each speaker's vowel category or, ideally, five RMS estimations (one for each vowel) per speaker, though some vowel categories were removed due to a low number of observations (see Tables 4.7 and 4.8).

Model fitting proceeded as follows. First, a baseline model with a random effect of **Speaker** was fit to predict the RMS of each speaker's vowel category. The parameter **Phone** was then added to the model, which unsurprisingly improved model fit. Next, the parameter **Maternal Language Profile**, with the levels Monolingual Quechua, Quechua-dominant, and Bilingual Quechua-Spanish, was added to the model. **Maternal Language Profile** improved upon a model fit containing **Phone** and the random effect of **Speaker**, under an alpha level of .10 (model summary presented in Table 4.11). This modeling thus shows a trend that the mother's language profile predicts vowel variability in these bilingual children. More specifically, the positive coefficients for the levels of **Maternal Language Profile**, with a reference level of "Monolingual Quechua", show a trend that the RMS of the vowel category is larger for the children with Quechua-dominant caregivers and bilingual Quechua-Spanish caregivers.

After evaluating the differences in category dispersion by maternal language profile, category dispersion was correlated with the ambient language measures derived from the daylong audio recordings. For this analysis, two ambient language measures, reflecting the child's expressive language experience, were calculated.[14]

1. The first measure was the percentage of monolingual Spanish language clips in the recording where the target child was speaking. This was calculated by dividing the number of clips annotated as 'Spanish' where the child was one of the speakers by the total number of *language* clips where the child was one of the speakers. Language clips

[14] An additional ambient language measure, the percentage of monolingual Spanish clips in the recording where an adult was speaking, was additionally computed. This measure reflected the children's receptive language experience from their environment. For brevity, these results are summarized here: the percentage of monolingual Spanish clips spoken by adults did not predict the children's vowel dispersion. Thus, the discrete parameter of maternal language profile remained the stronger predictor. The percentage of monolingual Spanish clips spoken by adults did, however, predict coarticulation and duration by word environment: children who heard more monolingual Spanish from adults in their environment tended to distinguish less by word environments, both in terms of coarticulation and duration (this in line with the findings from the children's own language use). However, separate models fit to predict the coarticulation and duration data showed that in all cases the child's own language patterning (percentage of monolingual Spanish clips) was the stronger predictor of children's speech patterning by word environment.

Table 4.11: Model predicting vowel category variability

Intercept	0.63^{***}
	(0.47, 0.79)
Phone:[e]	-0.25^{**}
	(−0.40, −0.09)
Phone:[i]	-0.29^{***}
	(−0.43, −0.15)
Phone:[o]	-0.35^{***}
	(−0.52, −0.18)
Phone:[u]	-0.14^{+}
	(−0.29, 0.02)
Lang. Profile: Quechua-dominant	0.20^{+}
	(−0.01, 0.41)
Lang. Profile: Bilingual	0.15^{+}
	(−0.01, 0.30)
Observations	149
Log Likelihood	−55.51
Akaike Inf. Crit.	129.02
Bayesian Inf. Crit.	156.05
Note:	$^{+}$p<0.1; **p<0.01; ***p<0.001

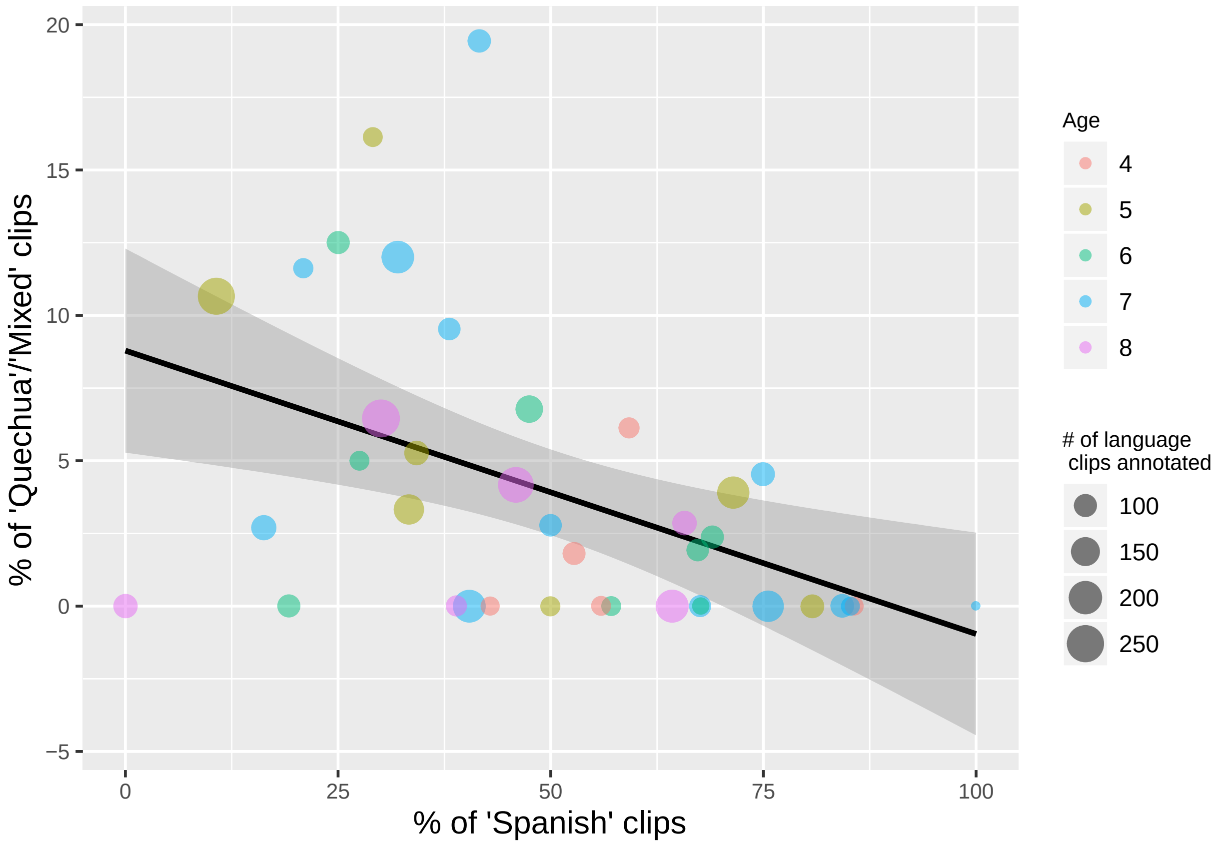


Figure 4.8: Correlation between the number of clips containing the target child annotated 'Spanish' and annotated 'Quechua'/'Mixed'

were defined as those clips annotated as Mixed, Spanish, or Quechua, and not clips annotated as Unsure, No Speech, or containing PID.

2. The second ambient language measure calculated was the percentage of monolingual Quechua and mixed Quechua/Spanish clips in the recording where the target child was speaking. The percentage of Quechua and Mixed clips was calculated by first adding the total number of Quechua clips and Mixed clips where the target child was speaking. This sum was then divided by the total number language clips where the target child was speaking. The decision to use both monolingual Quechua clips and mixed Quechua/Spanish clips, instead of just monolingual Quechua clips, was made because some children had very few or no monolingual Quechua clips where they were speaking. Figures representing the relationship between the speech production outcomes and the percentage of monolingual Quechua clips where the target child was speaking are still included in the appendices.

These two environmental predictors are necessarily correlated. That is, the more monolingual Spanish clips where the child is speaking in a recording, the fewer Quechua/Mixed clips there are likely to be (Figure 4.8). A Pearson correlation coefficient assessing the relationship between the percentage of Spanish clips and the percentage of Quechua/Mixed clips containing the target child in the recordings demonstrates that these two predictors are indeed significantly correlated ($r(37) = -0.46$, $p = 0.004$).

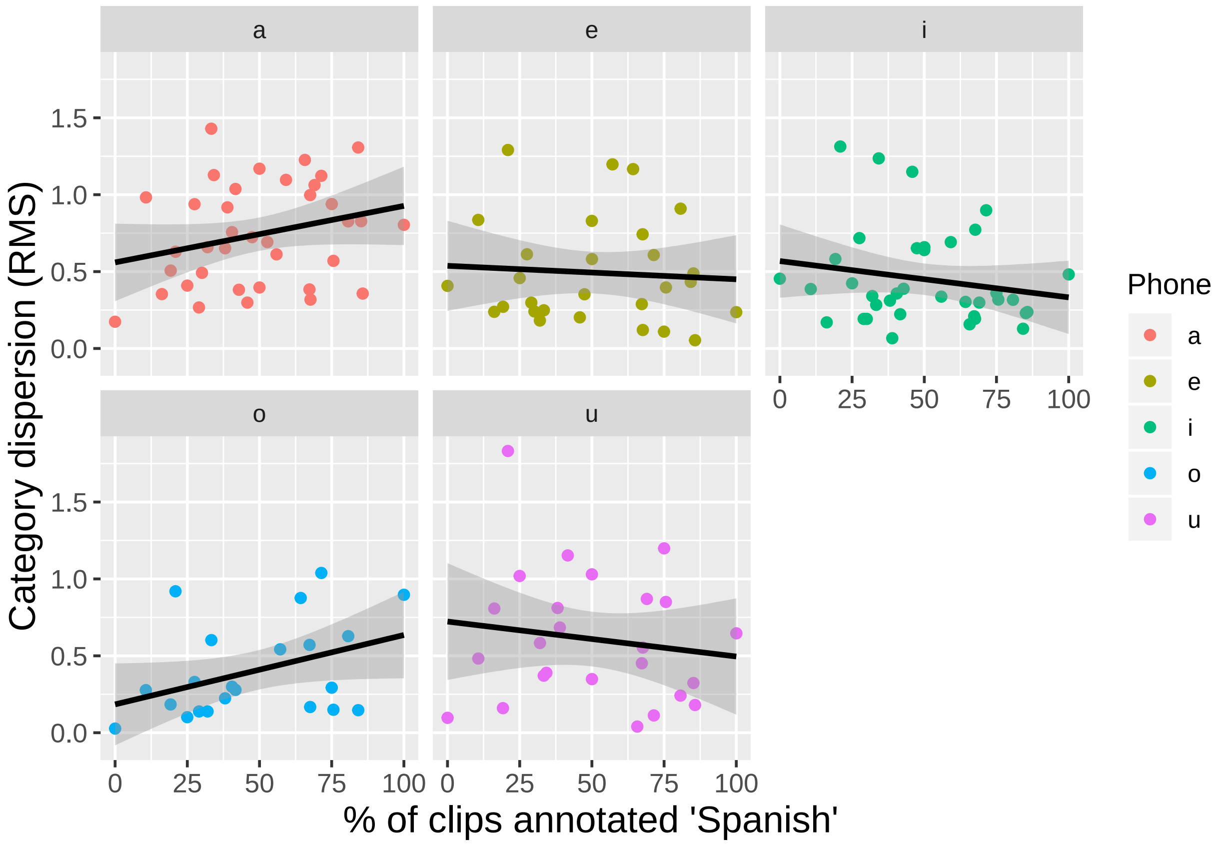


Figure 4.9: Vowel category dispersion (RMS) by percentage of Spanish clips containing target child and phone.

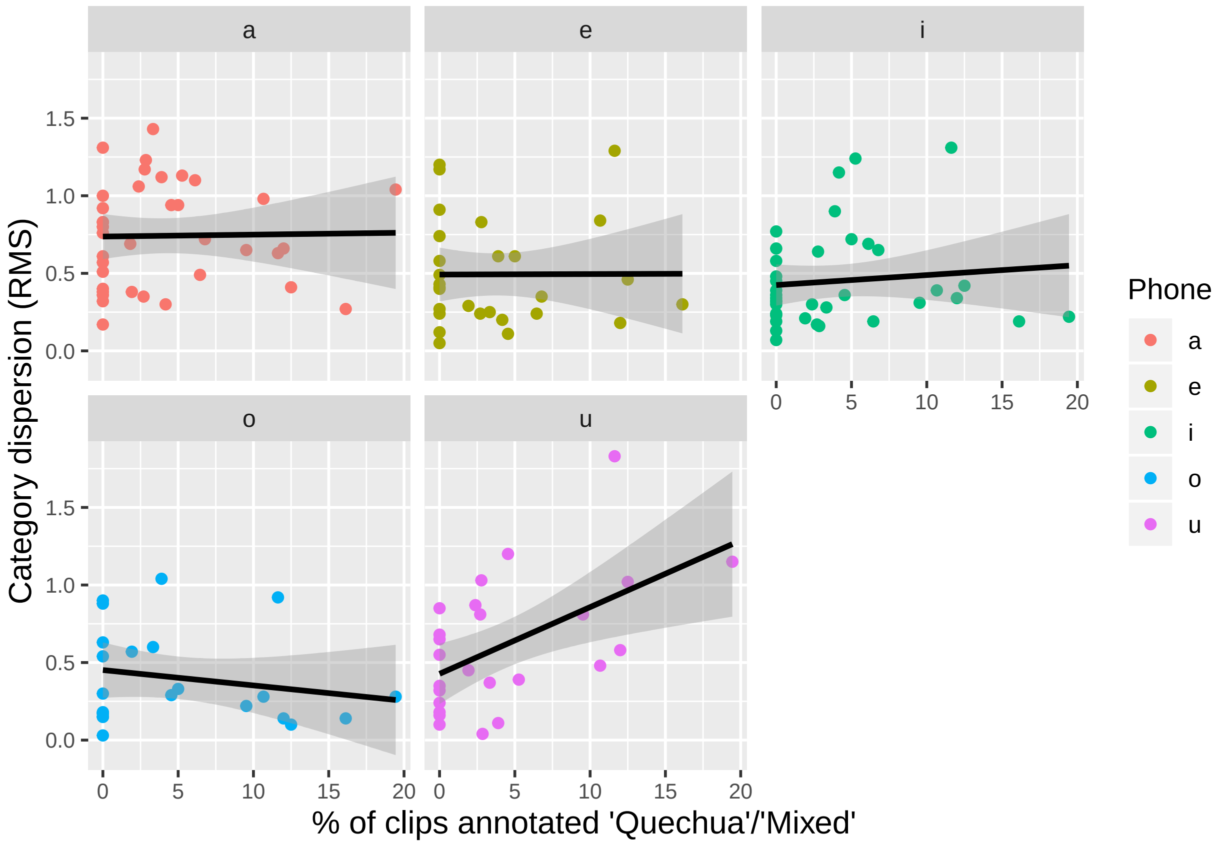


Figure 4.10: Vowel category dispersion (RMS) by percentage of Quechua/Mixed clips.

According to the experimental hypotheses, children who are exposed to and use more Quechua should have tighter, less variable categories. As Figure 4.9 demonstrates, there is some, limited evidence for the idea that children's Quechua language use affects their vowel production, at least for [a]: children who have a smaller percentage of Spanish clips in their recording have less variable [a] categories. However, the figures make it clear that this relationship is not apparent or consistent across all of the vowel categories. Figure 4.10 similarly shows limited evidence for a relationship between ambient language measures and [u] variability, but overall vowel category variability does not seem to vary by the percentage of Quechua clips in the children's recordings. That is, there is no consistent relationship between category variability and environmental effects across all of the vowels. See Figure 4.18 in the Appendices for a display of vowel dispersion by the percentage of monolingual Quechua clips.

A series of linear mixed effects models were fit to evaluate whether the ambient language measures predicted category variability in the children's speech. As with the modeling for mother's bilingual language profile, the outcome variable was the RMS of each speaker's vowel category, so five categories per speaker. Neither of the two parameters **Percentage**

of Spanish Clips or **Percentage of Quechua Clips**, improved upon a model fit that included random effects of **Speaker** and a parameter for **Phone**, suggesting there was no correlation between environmental effects and category variability.

To conclude, the observed relationship between category variability and maternal language profile suggests that the caregivers' bilingual language practices may influence children's spoken vowel patterning. But there was little or no relationship between ambient language measures from the daylong recordings and vowel variability. It is unclear if this accurately reflects a lack of ambient effects or if the methods used to estimate the children's bilingual language environments were inadequate.

Speech production by word environment

The final section of the results evaluates the relationship between the children's language environment and the analysis of morphologically complex words. Here, as in Chapter 3, the analysis and decomposition of morphologically complex words is measured over a biphone sequence, such as [ap], that falls within morpheme boundaries (e.g. *api* 'corn/citrus drink') and across morpheme boundaries (e.g. *llama-pi* 'llama-LOC'). Morphological decomposition is then quantified in two ways: 1) by the *duration* of biphone sequences, such as [ap], that fall in the two word environments, and 2) by the *coarticulation* between phones in biphone sequences, like [ap], that fall in the two environments. Coarticulation here is measured acoustically as the Euclidean distance between the spectrum averaged over the middle third of the adjacent phones (see Chapter 3, Experiment 1).

The overall hypothesis is that children who hear and use more Quechua will be more likely to differentiate their coarticulation and durational patterns by word environment. As a result, these children will show a greater difference between their within and between morpheme coarticulation than the more Spanish-dominant children. Similarly, Quechua-dominant children will show a greater difference between their within and between morpheme biphone sequence duration. If confirmed, these results would suggest that children who hear and use more Quechua are better able to distinguish between word environments, perhaps because they are better able to analyze the internal structure of morphologically complex Quechua words.

The first analysis examines the effect of maternal language profile on the children's speech patterns by word environment. Recall that the Euclidean distance between the phones in the biphone sequence was computed for each word environment - between-morpheme and within-morpheme. To evaluate the difference in coarticulation by word environment, the difference between coarticulation in the across-morpheme environment and the within-morpheme environment was calculated This computed difference is a measure of how *different* the Euclidean distances are in the two word environments. So here a larger difference between the Euclidean distances in the two environments indicates that the child differentiates more between the two word environments. This increased differentiation indicates that the child may be decomposing the morphologically complex words.

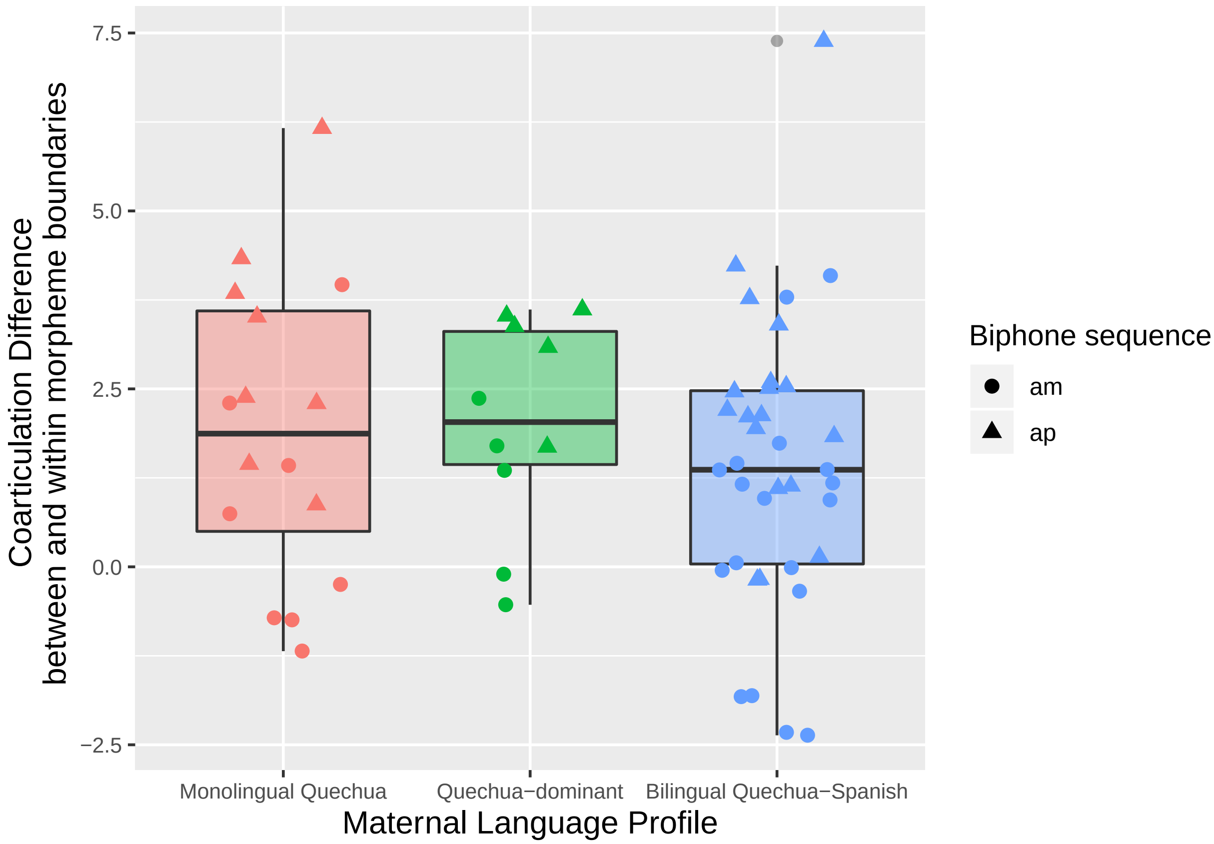


Figure 4.11: Coarticulation difference by maternal language profile and biphone sequence.

Table 4.12: Coarticulation difference by maternal language profile and biphone sequence

	[am]	[ap]
	Mean (SD)	Mean (SD)
Monolingual Quechua	0.69 (1.78)	3.11 (1.71)
Quechua-dominant	0.96 (1.23)	3.06 (0.79)
Bilingual Quechua-Spanish	0.52 (1.84)	2.29 (1.77)

The same difference measure was computed for duration. As with coarticulation by word environment, a larger difference in the duration of a biphone sequence in the within-morpheme position and the across-morpheme position indicates differentiation by word environment. These outcome measures by word environment will be referred to as **Coarticulation Difference** and **Durational Difference** for the remainder of the analysis.

Tables 4.12 and 4.13 and Figures 4.11 and 4.12 present the analysis of the Coarticulation

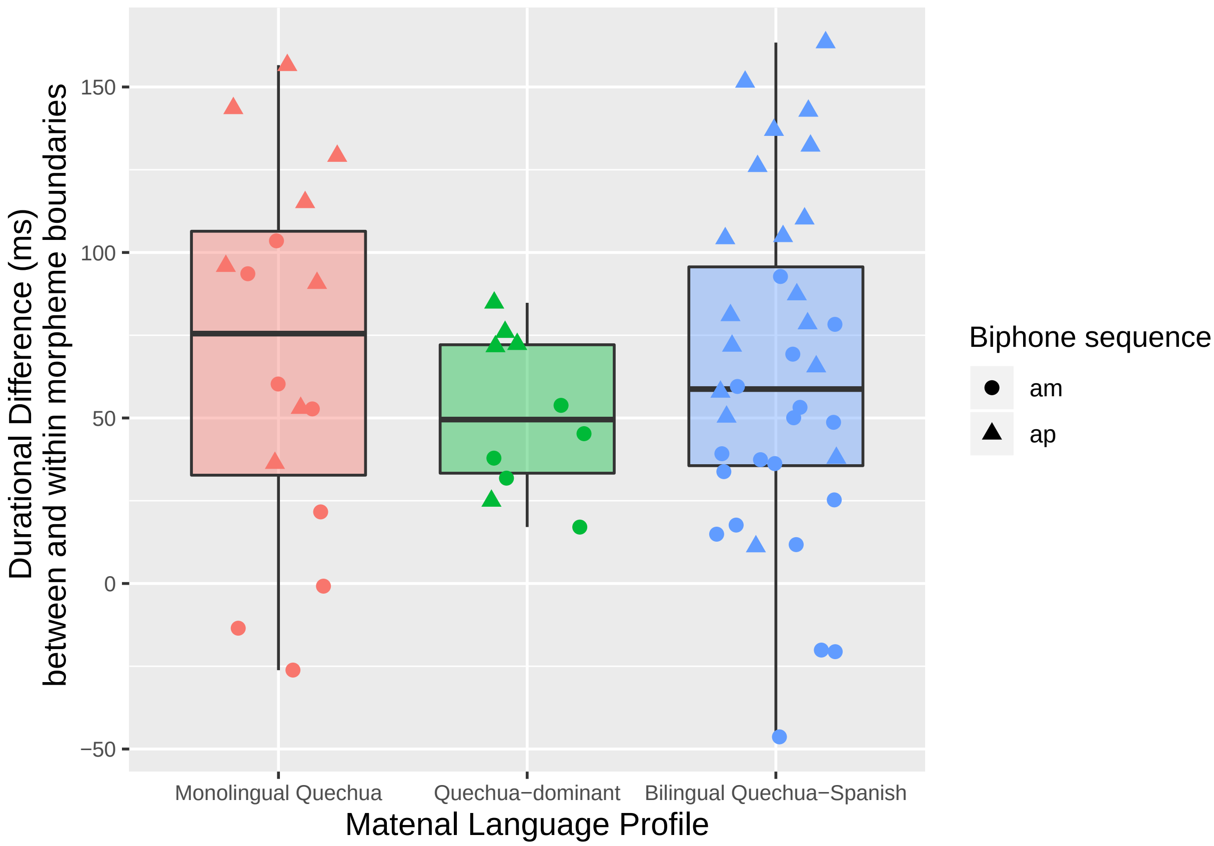


Figure 4.12: Durational difference by maternal language profile and biphone sequence.

Table 4.13: Durational difference (ms) by maternal language profile and biphone sequence

	[am]	[ap]
	Mean (SD)	Mean (SD)
Monolingual Quechua	36.41 (48.72)	102.56 (42.22)
Quechua-dominant	37.16 (13.92)	65.93 (23.49)
Bilingual Quechua-Spanish	32.28 (35.74)	95.22 (42.2)

Difference and Durational Difference by maternal language profile. Overall, the data reported in the tables suggest that there are not large differences in duration or coarticulatory patterns, for either biphone sequence, by maternal language profile. For example, the durational difference between the two word environments for the children with monolingual mothers was 102.56 (42.22) ms, and the difference for the children with bilingual mothers was 95.22 (42.2) ms. This was a difference of approximately 7ms (for the biphone sequence [ap]). These small differences between maternal language profiles suggest that the mothers' language dominance does not predict children's morphological analysis. This hypothesis is evaluated statistically in the following section, where a series of models are fit to predict the children's speech production patterns.

Besides maternal language profile, the bilingual language characteristics from the daylong recordings were also correlated with the children's speech production patterns - duration and coarticulation. Here, as before, the outcome variable is the difference between within- and across-morpheme biphones sequence duration, or the difference between the within- and across-morpheme coarticulation. It is anticipated that children who use more Quechua will exhibit larger differences in their speech patterns between the two word environments, suggesting that Quechua-dominant children analyze the internal structure of morphologically complex words differently from Spanish-dominant children.

Figures 4.13 and 4.14 display the relationship between the the speech production measures and the percentage of annotated clips where the target child was speaking Spanish. The negative relationships apparent in these plots - for both biphone sequences - suggest that children who speak more Spanish exhibit a *smaller* durational and coarticulatory difference between the two word environments. Overall, those children who speak more Spanish in their recordings tend to distinguish less between across-morpheme and within-morpheme word environments.

The other ambient language characteristic was the percentage of Quechua/Mixed clips containing the target child. The relationship between the speech production measures and percentages of Quechua/Mixed clips is presented in Figure 4.15 for coarticulation and Figure 4.16 for duration. The positive relationship apparent in both figures shows that children who have a larger percentage of Quechua/Mixed clips in their recordings show *larger* differences between the two word environments. This positive relationship is apparent for both the Coarticulation Difference and the Durational Difference, as well as for both biphone sequences [am] and [ap]. (Figures 4.24 and 4.25 in the Appendices present these results by the percentage of monolingual Quechua clips, without out any Mixed speech clips and the trend is the same.)

The above results suggest that there are some environmental effects on the children's coarticulatory and durational patterning by word environment. To further evaluate this, two sets of linear mixed effects models were fit, one to predict Coarticulation Difference and one to predict Durational Difference. For both sets of models, the baseline model included a random effect for **Speaker**. Then, the parameter **Biphone Sequence** was added, which improved the model fit for the models predicting Coarticulation Difference and Durational Difference. Once models with **Speaker** and **Biphone Sequence** had been constructed,

Table 4.14: Model predicting coarticulation difference by percentage of Spanish clips.

Intercept	2.22***
	(1.37, 3.08)
Biphone sequence:[ap]	1.99***
	(1.26, 2.71)
Percentage of Spanish clips	−0.05***
	(−0.07, −0.03)
Observations	62
Log Likelihood	−114.05
Akaike Inf. Crit.	238.11
Bayesian Inf. Crit.	248.74
Note:	*p<0.05; **p<0.01; ***p<0.001

Table 4.15: Model predicting coarticulation difference by percentage of Quechua clips.

Intercept	0.12
	(−0.54, 0.78)
Biphone sequence:[ap]	1.99***
	(1.22, 2.76)
Percentage of Quechua clips	0.11**
	(0.03, 0.19)
Observations	62
Log Likelihood	−117.75
Akaike Inf. Crit.	245.50
Bayesian Inf. Crit.	256.13
Note:	*p<0.05; **p<0.01; ***p<0.001

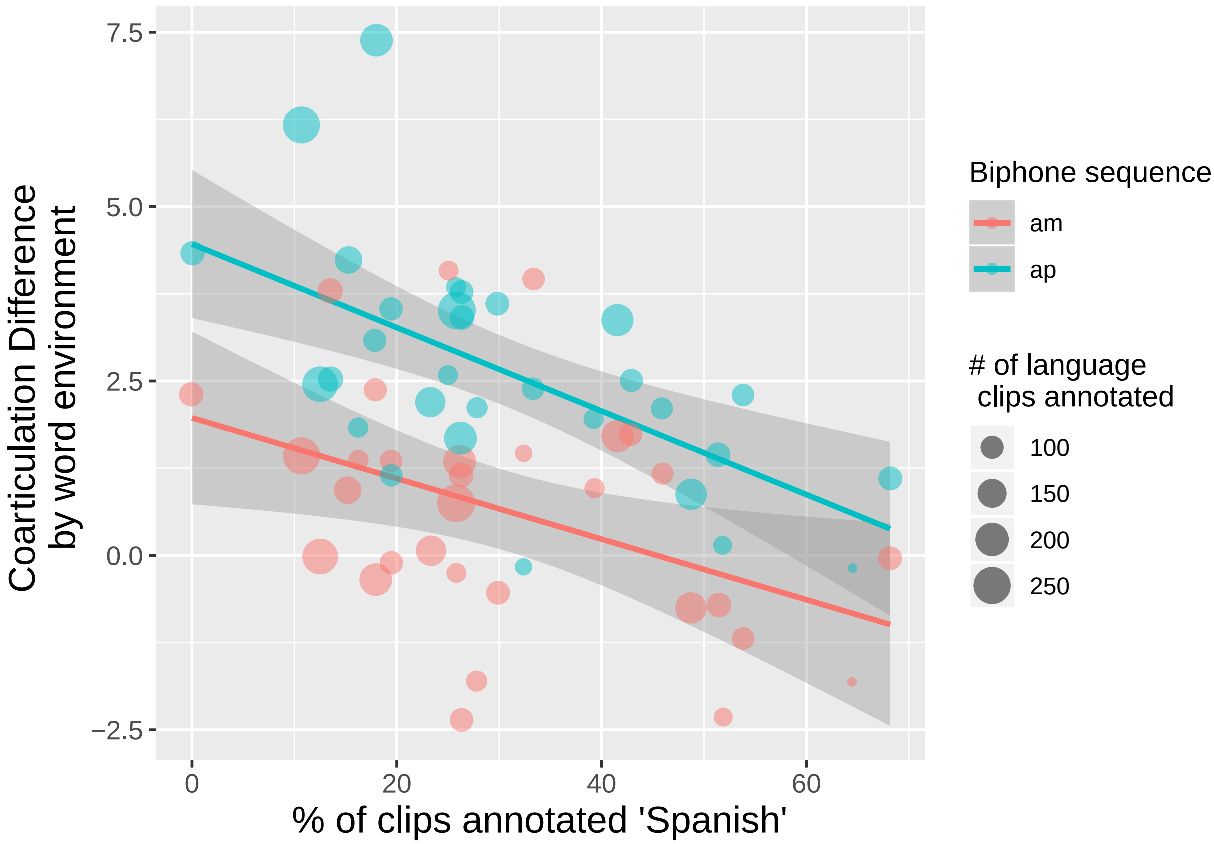


Figure 4.13: Coarticulation difference by percentage of Spanish clips containing the target child.

each of the environmental effects was added in turn. Note that each environmental effect was added *separately* to the models containing the random effect of **Speaker** and fixed effect of **Biphone Sequence**, due to the high correlation between the environmental parameters (see Figure 4.8). For example, the recordings that contained a larger percentage of Quechua/Mixed clips necessarily had a smaller percentage of Spanish clips containing the target child. In addition, for all models summarized below, the interaction between the environmental effect (e.g. **Maternal Language Profile**) and the fixed effect **Biphone sequence** was added to the model. This interaction did not improve upon any model fits with the lone terms.

For the models predicting Coarticulation Difference, the parameter **Percentage of Spanish Clips**, which represents the percentage of clips containing the target child speaking Spanish, improved upon model fit (Model summary listed in Table 4.14). The negative beta coefficient for **Percentage of Spanish Clips** suggests that the more Spanish clips heard in the recording, the smaller the Coarticulation Difference between the two word environments.

Next, the environmental effect **Percentage of Quechua Clips**, representing the per-

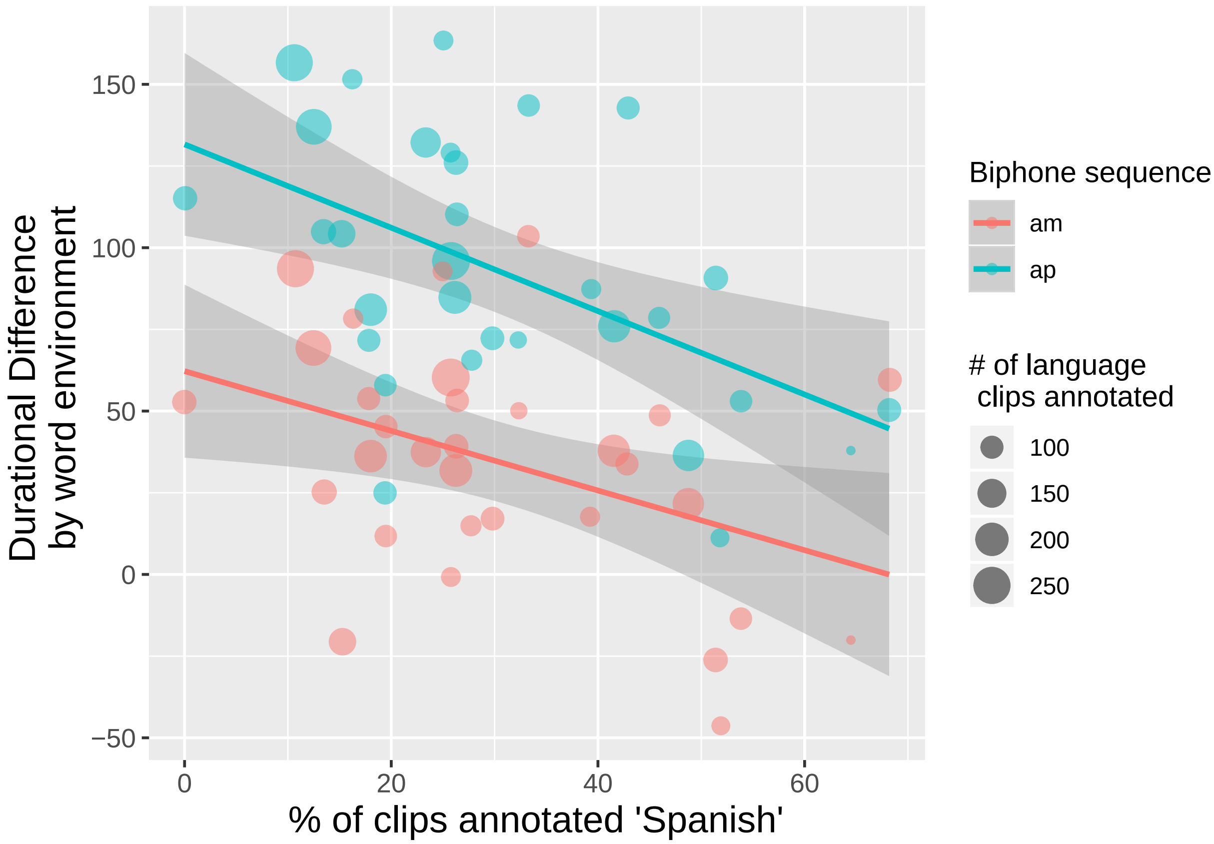


Figure 4.14: Durational difference by percentage of Spanish clips containing the target child.

centage of Quechua/Mixed clips in the recording containing the target child, was added to the model with the random effect of **Speaker** and the parameter **Biphone Sequence**. **Percentage of Quechua Clips** improved the Coarticulation Difference model (Table 4.15). In this model, a positive beta coefficient suggests that the larger the percentage of Quechua/Mixed clips, the larger the Coarticulation Difference between the two word environments.

The final environmental effect, **Maternal Language Profile**, represented the bilingual language profile of the mother (monolingual Quechua, Quechua-dominant, or bilingual Quechua-Spanish). This parameter was added to the model containing the random effect for **Speaker** and **Biphone Sequence**. **Maternal Language Profile** did not improve upon model fit, suggesting that the mother's language status seems to have less of an effect upon the Coarticulation Difference than the other environmental effects **Percentage of Quechua Clips** and **Percentage of Spanish Clips** (Model summary in Table 4.20 in the Appendices).

As a final step in the model fitting procedure, a parameter estimating each child's "talkativeness" was added to the models. This parameter, **Percentage of Child Speaking**, was included to control for the possibility that children who simply talk more - in

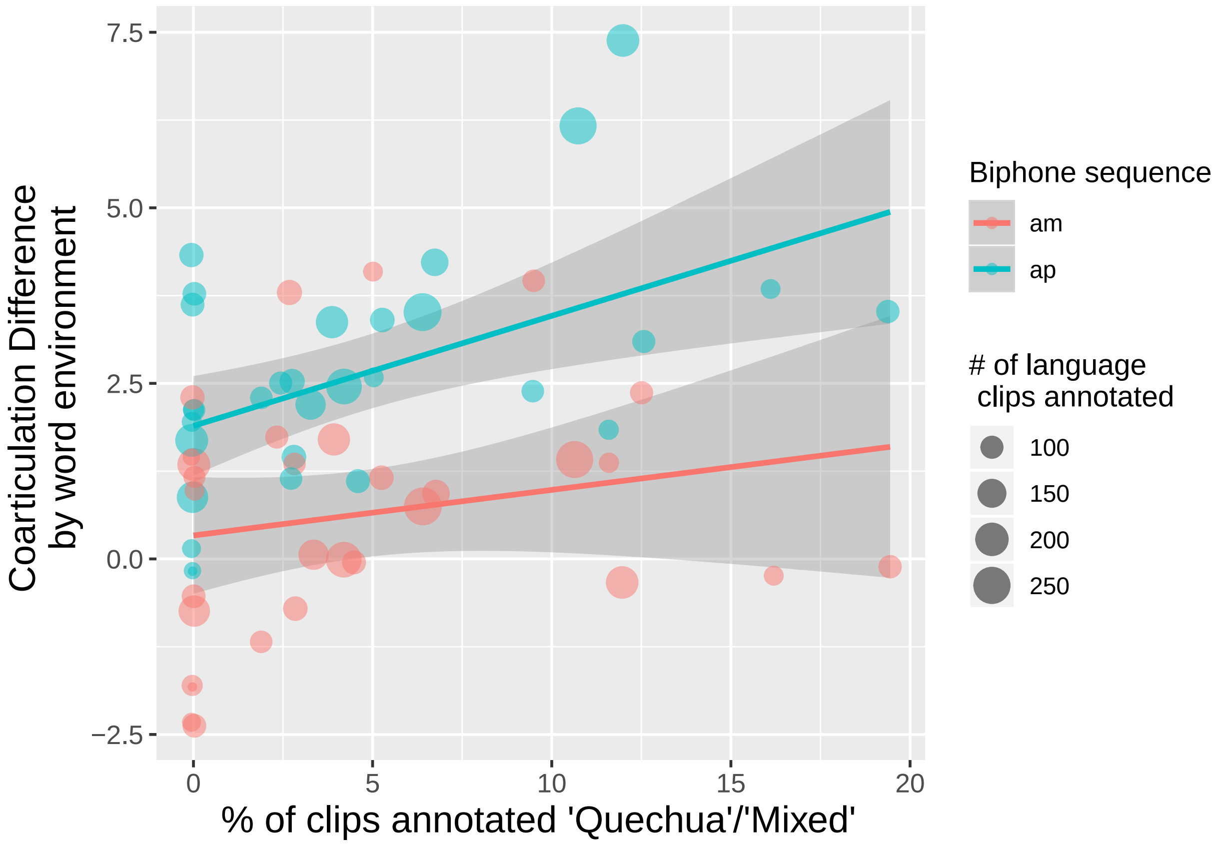


Figure 4.15: Coarticulation difference by the percentage of Quechua/Mixed clips containing the target child.

whatever language - may exhibit different speech production patterns by word environment. **Percentage of Child Speaking** was calculated by dividing the number of clips where the target child was speaking by the total number of language clips (those annotated as Quechua, Spanish, or Mixed, but not unsure or no speech), irrespective of speaker. The parameter **Percentage of Child Speaking** did not improve upon a model containing **Biphone Sequence**, **Percentage of Spanish Clips** or **Percentage of Quechua Clips**, and the random effect of **Speaker**. Furthermore, neither the magnitude nor direction of the effect of **Percentage of Spanish Clips**/**Percentage of Quechua Clips** changed with the addition of **Percentage of Child Speaking**. Though the **Percentage of Child Speaking** is just an estimation of the children's talkativeness, these results suggest that it is the effect of speaking Spanish or Quechua that predicts the children's speech patterns - and not simply how frequently the children talk.

For the models predicting Durational Difference, the parameter **Percentage of Spanish Clips** improved upon model fit (Model summary listed in Table 4.16). The negative beta coefficient for the **Percentage of Spanish Clips** parameter suggests that the more Spanish

Table 4.16: Model predicting durational difference by percentage of Spanish clips.

Intercept	67.79***
	(44.52, 91.07)
Biphone sequence:[ap]	58.26***
	(45.90, 70.61)
Percentage of Spanish clips	−1.09***
	(−1.74, −0.45)
Observations	62
Log Likelihood	−296.66
Akaike Inf. Crit.	603.31
Bayesian Inf. Crit.	613.95
Note:	*p<0.05; **p<0.01; ***p<0.001

Table 4.17: Model predicting durational difference by percentage of Quechua clips.

Intercept	25.33**
	(8.50, 42.17)
Biphone sequence:[ap]	58.26***
	(45.90, 70.61)
Percentage of Quechua clips	1.89
	(−0.35, 4.13)
Observations	62
Log Likelihood	−298.89
Akaike Inf. Crit.	607.77
Bayesian Inf. Crit.	618.41
Note:	*p<0.05; **p<0.01; ***p<0.001

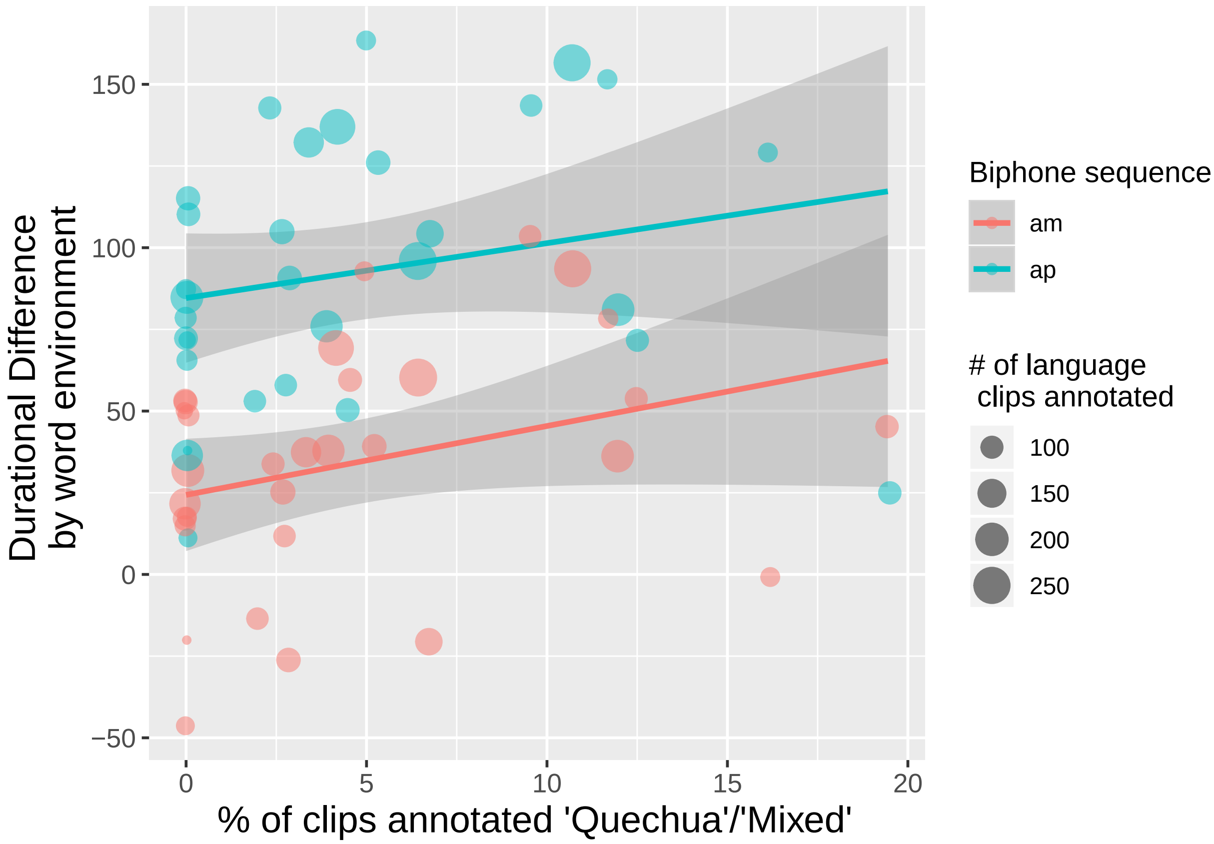


Figure 4.16: Durational difference by the percentage of Quechua/Mixed clips containing the target child.

clips heard in the recording, the smaller the Durational Difference between the two word environments.

Next, the environmental effect **Percentage of Quechua Clips** was added to the model. Unlike the coarticulation models, **Percentage of Quechua Clips** did not improve upon a model containing the random effect of **Speaker** and **Biphone Sequence** under an alpha level of .05 (Table 4.17), though the trend was in the hypothesized direction. This result suggests that the percentage of Quechua/Mixed clips in the recording is not reliably predictive of the Durational Difference between word environments.

The environmental effect **Maternal Language Profile** was also added to the model containing the random effect and **Biphone Sequence**. **Maternal Language Profile** did not improve upon model fit, suggesting that the mother's language status does not have an effect upon the Durational Difference between word environments in the children's speech.

Once again, the parameter **Percentage of Child Speaking**, representing each child's talkativeness, was added to the model containing the parameter **Percentage of Spanish Clips**. **Percentage of Child Speaking** did not improve upon a model containing **Bi-**

phone Sequence, **Percentage of Spanish Clips**, and the random effect of **Speaker**. Nor did the magnitude or direction of the effect of **Percentage of Spanish Clips** change with the addition of **Percentage of Child Speaking**. This again suggests that the relationship between **Percentage of Spanish Clips** and the children's speech patterns exists independent of how frequently the children talk.

Overall, this modeling demonstrates that the **Percentage of Spanish Clips** and **Percentage of Quechua Clips** predict the Coarticulation Difference betwen the word environments, but that **Maternal Language Profile** does not. **Percentage of Spanish Clips** likewise predicts the Durational Difference between the word environments, but neither **Maternal Language Profile** nor **Percentage of Quechua Clips** does. Unlike vowel variability then, coarticulation and duration patterns by word environment are not predicted by the mother's language profile, but instead by the children's own speech practices.

4.5 Discussion

In this chapter, daylong audio recordings of bilingual Quechua-Spanish children's language environments were annotated to estimate the children's exposure to both of their languages. The estimates of language exposure, in addition to the central caregiver's language dominance, were then used to predict the children's performance on a series of speech production tasks.

The primary research question in this study asked if the language environment could predict children's phonetic variation. Specifically, this study examined the roles of expressive versus receptive language experience for phonetic development. Correlations between the children's performance on the speech production tasks and the daylong audio recording measures demonstrated that the environment did play a role, but that the effect of receptive versus expressive experience depended upon the phonetic outcome measure (vowel dispersion or phonetic production by word environment). Variability in children's vowel production was more contingent upon the mother's language dominance than any environmental effect derived from the daylong recordings, suggesting that receptive language is the stronger predictor of vowel variability. The results from the second speech production task, which evaluated the children's ability to distinguish between morphological environments, depended more upon the children's own language production: children who used more Spanish in the daylong recordings tended to distinguish less between the morphological environments in their speech production. Implications for the results of these speech production tasks are discussed in greater detail in the following sections.

One additional contribution of this chapter was the daylong recording annotation workflow that was outlined. This study employed random sampling to efficiently estimate the quantity of each language that the children heard and used throughout the recordings. This annotation technique resulted in efficient, stable estimates of the children's bilingual language exposure. Language exposure and use estimates were achieved after annotating, on average, just 185 30-second clips per child, or 90 minutes from each recording. While validation of this annotation technique for bilingual language exposure and speech register (e.g. child-directed versus overheard speech) is ongoing (Cychosz et al. 2020c), these results tentatively suggest that daylong recordings can be used to make reliable, ecologically-valid estimates of children's dual language exposure. Going forward, researchers may wish to complement parental reports of children's dual language use with more naturalistic estimates drawn from daylong recordings. Or, with populations that are less familiar with behavioral research or written language, researchers could forgo parental reports altogether and instead rely on estimates derived from recordings.

Environmental effects on vowel variability

A central finding from this work was that a receptive language exposure measure, the mother's language dominance or bilingual language profile, correlated with the children's within-category vowel variation. Specifically, children with monolingual Quechua mothers produced smaller, tighter vowel categories than children with Quechua-dominant mothers or bilingual Quechua-Spanish mothers. Likewise, children with Quechua-dominant mothers tended to have smaller, tighter categories than children with bilingual mothers. While acknowledging that these maternal language profile groups were unbalanced - the monolingual mother group had just 10 children compared to 23 children in the bilingual group - this finding still suggests a relationship between a receptive predictor, maternal language patterns, and the children's own speech variation.

Interestingly, the results did not show an effect of the *children's* own language usage upon their vowel variability: there was neither an effect of children speaking more Spanish nor more Quechua/Mixed speech upon their vowel variability. As mentioned in the Results, the absence of a correlation between these language exposure estimates and vowel variation does not mean that the children's language use or expressive language patterns do not affect their vowel variability. It is possible that the daylong recording methods were not sufficiently sensitive to capture those individual differences, or a potential role of expressive language, for vowel variation.

Nevertheless, the finding that maternal language profile predicts some variability in children's vowel patterning warrants some exploration. Here it was found that children with bilingual mothers had more variable vowel categories than children whose mothers reported being more Quechua-dominant or monolingual in Quechua. But the reasons for the direction of the pattern are not immediately clear: why do children with bilingual mothers have more variable vowel categories? After all, those children are likely exposed to more Spanish, a language with five phonemic vowels. This five-vowel contrast could put pressure on the phonological system to *reduce* variability, in an effort to maintain contrast (Bradlow 1995).

There are two potential explanations why children with bilingual mothers tended to have more variable Quechua vowel categories. First, vowel categories could be more variable due to a general effect of bilingualism: all of the children in this study receive input in two languages, potentially rendering their phonological targets in both languages more variable. Second, vowel categories could be more variable due to the specific language combination studied here: Quechua and Spanish. The following section addresses each of these potential explanations in turn. The goal of this discussion is not to distinguish between the general effect of bilingual input and the specific effect of Spanish and Quechua. The current study is not designed to make this distinction. Rather this discussion will highlight two reasons for the direction of the effect of maternal language profile upon vowel variation.

Children with bilingual mothers may have more variable vowel categories because of the type of language that those children are exposed to. Although all of the children tested were bilingual Quechua-Spanish speakers, the primary difference between the maternal language profile groups was, presumably, the quantity of Quechua (and Spanish) that the children heard. The children with bilingual mothers, while probably not exposed to equal amounts of Quechua and Spanish, could have a more mixed linguistic environment than the children with monolingual mothers. Receiving, as an estimation, 50% of their input in one phonological system and 50% in the other could render the children's phonological targets - in both languages - more variable, since those children receive less overall input in both of their languages.

These types of "bilingual exposure effects" have been found in other bilingual populations, for other outcome measures, as summarized in the Previous Literature of this chapter. For example, bilingual infants aged 0;11 who received a larger proportion of their input in a second language had larger receptive vocabularies in that language (Carbajal and Peperkamp 2019). Similarly, bilingual French-English children aged 5;0 had larger vocabularies in their dominant language (Thordardottir 2011) (see also summaries of Bijeljac-Babic et al. (2012) and Potter et al. (2019) in the Previous Literature). To my knowledge, a relationship between bilingual exposure and children's vowel production has not previously been quantified. However, it is reasonable to propose that the relative proportion of two languages in a bilingual environment - whatever those languages are - could predict variability in vowel patterning: the more input the child receives in one of their languages, the less variable their vowel categories might be.

Alternatively, vowel categories may be more variable in the bilingual maternal language profile group because of the specific language combination of Quechua and Spanish studied here. Recall that Quechua has three phonemic vowels /a, i, u/ and two allophonic mid-vowels [e, o] derived only in uvular contexts (Gallagher 2016). Spanish has the same vowel system, /a, e, i, o, u/, but all of the vowels are phonemic.

Because mid-vowels in Quechua are only derived in limited (uvular) contexts, they are less frequent than the peripheral phonemic vowels (Cychosz and Kalt 2018). What effect could this language-specific vowel frequency have on the children's vowel production? An explanation based on language-specific vowel frequencies would predict that the children with monolingual mothers should have smaller, less variable categories. This would hold for all the tested vowels in Quechua, but *especially* for the more frequent, peripheral Quechua /a, i, u/. For children with monolingual mothers, these peripheral vowels are more frequent because 1) the children are exposed to more Quechua in their environment and 2) /a, i, u/ are more frequent than [e] and [o] within Quechua.

The data tend to bear out a prediction based on language-specific vowel frequencies.

Consider Table 4.18, originally presented in the Results section and listed again below for convenience. The table displays the RMS value of each vowel category by maternal language profile. The largest difference in RMS between the monolingual Quechua group and the bilingual Quechua-Spanish group is for /u/ with a difference in category RMS of 0.26, followed by /a/ with a difference of 0.22. The RMS differences between the maternal language profile groups for [e] and [o] are less notable: for [o], there is just a difference of .03 between the monolingual and Quechua-dominant groups, and 0.07 between the monolingual and bilingual groups. In fact, [o] is actually *more* variable in the monolingual Quechua group than the Quechua-dominant group.

Table 4.18: Average and standard deviation of vowel category dispersion (RMS) by phone and maternal language profile.

	Mean (SD) [a]	[e]	[i]	[o]	[u]
Monolingual Quechua	0.69 (0.47)	0.61 (0.36)	0.44 (0.27)	0.63 (0.37)	0.58 (0.41)
Quechua-dominant	1.03 (0.59)	0.7 (0.3)	0.6 (0.31)	0.6 (0.39)	0.85 (0.55)
Bilingual Quechua-Spanish	0.91 (0.57)	0.78 (0.51)	0.63 (0.5)	0.7 (0.4)	0.84 (0.6)

These results suggest that language-specific vowel frequencies may explain some of the variability between maternal language profile groups. Children with monolingual mothers have tighter, more compact vowel categories in Quechua because they receive more input in Quechua, allowing them to more quickly attune to the categories in their speech production. This frequency-based explanation is further supported by the differences in variability between the Quechua peripheral and mid-vowels. Even the children with monolingual Quechua mothers receive relatively less input for the less frequent [e] and [o]; consequently, the differences in variability of [e] and [o] between maternal language profile groups tend to be less notable.[15]

The discussion above leaves us with several conclusions. First, these bilingual children's vowel variability is best predicted by a receptive language measure, maternal bilingual language profile, as exhibited by differences between monolingual Quechua, Quechua-dominant, and bilingual Quechua-Spanish caregiver groups. Differences in the children's own expressive language (percentage of clips annotated as Spanish or Quechua/Mixed that contained the target child) did not predict any individual differences in vowel patterning. A second

[15] One could note that the RMS is roughly equivalent between the mid-vowels and the peripheral vowels. However, there are numerous reasons - independent of frequency or exposure effects - that some vowels may have greater variability than others. For example, /i/ is known to be a highly stable vowel, consistently exhibiting less within-category variability than low vowels such as /a/. This is often attributed to the inflexible lingual posturing required to approximate /i/ without, for example, accidentally articulating a fricative when the tongue is that close to the palate. As a result of these articulatory configuration differences between vowels, it doesn't *necessarily* seem reasonable to attribute within-category differences between vowels to external influences such as frequency of exposure.

important conclusion addresses *why* the children with monolingual mothers have tighter, less variable vowel categories. It was suggested that the relative frequency of vowel exposure explains this pattern. The children with bilingual Quechua-Spanish mothers receive less overall input in Quechua, rendering their categories in that language more variable.

Environmental effects on morphological parsing

This chapter also found an environmental effect of expressive language upon children's ability to differentiate between morphological environments in Quechua. Specifically, children who spoke a higher proportion of Spanish throughout the day showed a *smaller* coarticulatory difference in the within-morpheme and across-morpheme environments of Quechua words. Children who spoke more Spanish also showed a reliably smaller durational difference between the morphological environments in Quechua words. Similar effects were found for relationships between children's Quechua/Mixed speech use and their ability to differentiate between morphological environments, although this effect only trended in the expected direction for the Durational Difference measure and the percentage of Quechua/Mixed clips. Chapter 3 discussed how duration and coarticulation by morphological environment indicate morphological analysis and the parsing of complex words. Consequently, this section assumes a relationship between these speech production metrics and morphological analysis. In doing so, the discussion here will outline what environmental effects upon morphological parsing mean for children's word learning and children's role in language change.

Implications for word learning

The relationship between language exposure and/or use and morphological analysis has clear implications for children's word learning, particularly the learning of morphologically complex words. The current study, for example, showed that children who used more Quechua/Mixed speech were better able to analyze the internal structure of complex words. The bilingual children studied here thus demonstrated how different amounts of language exposure can result in different word learning outcomes.

The result from this chapter also predicts that bilingual children who use less of one of their languages could demonstrate a protracted learning period in analyzing the internal structure of word forms *when compared to their age-matched peers*.[16] Maintaining more of these unanalyzed lexical forms in the lexicon could affect children's abstraction of sub-lexical categories like morphemes and phonemes. For example, a child who predominately uses Quechua may have sufficient experience with the language to reliably parse most Quechua suffixes from roots (recall that even monolingual adult speakers do not always parse morphologically complex words - consider words like *illuminate*, *increase*, or *uncouth* in English). For this child with more Quechua experience, the suffixes then become increasingly abstract,

[16] Vocabulary-matched peers would likely show exceptions to these predictions. For example, a child who doesn't use Quechua very frequently, but has a very large Quechua vocabulary, may still be able to analyze complex word forms more than their language use alone would predict due to the size of their vocabulary.

coinciding with, and perhaps even heightening, the child's meta-linguistic and morphological awareness (Diamanti et al. 2017). The child with relatively less Quechua experience, however, may continue to leave a great many complex word forms in their lexicon unanalyzed.

Children who use Quechua less almost certainly have distinct representations of suffixes and word forms. They do not unilaterally represent words in a holistic, unanalyzed form. Rather, for those children with less experience, the suffixes are simply not as abstracted away from the original inflected form as they are for children who use Quechua more. This analysis suggests that the structure of the lexicon, and connections between items within it, varies as a function of language exposure, specifically language *use*. As such, this conclusion supports previous work that established how the structure of the lexicon differs as a function of language experience (e.g. vocabulary size) in monolingual children (Edwards et al. 2004; Storkel 2002).

Note that these results do not suggest, or demonstrate, protracted morphological productivity in any of the participants. All of the children, regardless of language exposure or use, were capable of completing the morphological extension task. Thus, the task results showed, as it was assumed, that all of the children were morphologically productive Quechua speakers. The results do however suggest something more fine-grained: children who use less Quechua may be morphologically productive and capable of generalizing affixes to novel lexical environments while *simultaneously* having more holistic representations of entire, morphologically complex word forms. Redundant, conflicting representations of this form, at multiple levels in the grammar, are anticipated in models that assume that language categories are constructed gradually, based on experience (Arnon 2010; Davis and Redford 2019; Pierrehumbert 2016). Thus, even Spanish-dominant bilingual children could have abstract morphemes, though the morpheme categories would not necessarily be as abstract as those of more Quechua-dominant children.

Implications for language change

Historical language change is also implicated in the finding that children's morphological analysis is correlated with their language use and exposure. It is well-known that morphosyntactic change can occur as affixes undergo phonetic reduction and fuse to roots during running speech. Change can then occur as speakers subsequently fail to analyze the internal structure of complex words. Some have proposed that this change can occur during first language acquisition (Roberts and Roussou 1999; van Gelderen 2004; see Cournane (2017) and Cournane (2019) for recent arguments). But this assumption of the role of children in language change continues to be the subject of some debate, with some authors arguing that the similarities between language diachrony and ontogeny merely reflect shared cognitive biases between adults and children (Bybee and Slobin 1982; Diessel 2011).

The current study is consistent with the idea that morphological reanalysis could occur during first language acquisition. However, contrary to the viewpoints on children as innovators of language change (e.g. Cournane 2017), the current analysis does not maintain that change occurs because child language learners receive insufficient input. Nor does this analysis suggest that change occurs because child learners overgeneralize from the data they are given. Instead, particular language learning environments - such as bilingual environments - could facilitate different *levels* of morphological analysis. This graded reanalysis, dependent upon the learner's exposure and experience, could lead to change over time. The idea of gradient morphological analysis is fairly consistent with recent research on morphological productivity in adults which has demonstrated that affix productivity depends upon speaker experience - like the ratio of an affix to a stem (Hay 2003; Plag and Baayen 2009).

A graded reanalysis of morphologically complex words might be particularly common during sociolinguistic situations of language shift. For example, in southern Bolivia, rapid language shift to Spanish is occurring in Quechua-speaking communities. But as the current study has demonstrated, this shift does not always occur in a single generation where the adults speak Quechua and the children of the following generation switch entirely to Spanish. Instead, exposure to and experience with Quechua varies amongst the child speakers. The result is that children who use proportionately more Spanish analyze Quechua word forms differently, potentially igniting morphological change.[17] Crucially, these more Spanish-dominant children do not distinguish between the morphological environments in their *spoken* language patterns. It is possible that the more Spanish-dominant children will, with time and experience, differentiate more between morphological environments in their speech. However, if they do not, it is worth considering the structure and acoustic signal of the language input that Spanish-dominant participants might one day provide to their own children.

Future directions

Overheard versus child-directed speech

An important finding in this work is that the receptive measure of maternal language profile predicted the children's vowel variability better than the children's own language use. Maternal language profile even predicted vowel variability better than the proportion of adult Quechua use in the child's environment. This suggests that it is not general language exposure or use but receptive language experience, and in particular the mother's patterning, that predicts children's vowel variability. Given this relationship between variation and maternal language usage, an important next step is to determine the roles of directed and overheard maternal language input for the construction of children's spoken vowel categories. Future work on the role of the bilingual environment in children's phonological development should thus continue to estimate the quantity of each language that is directed to the child versus overheard by the child (see, for example Orena et al. (2020)).

Examining the differences between directed and overheard speech patterns in bilingual environments would have implications for recent work evaluating the importance of child-

[17] Adult second language learners could also exhibit some of these same patterns, though they clearly come to the learning problem with distinct learning biases and experiences (Bergmann et al. 2016; Lupyan and Dale 2010).

directed speech cross-linguistically (Casillas et al. 2019; Cristia 2020; Cristia et al. 2017). It is well known that many cultures do not direct speech directly to their children until toddlerhood or later (Lieven 1994). Yet it is equally acknowledged that the children socialized into these cultures acquire their native language(s) and reach core language development milestones (Casillas et al. 2019; Cychosz et al., under review). If children who receive more direct Quechua speech input, rather than overheard, have tighter, less variable (i.e. more adult-like) vowel categories, this would suggest some benefit for child-directed speech. If, however, the distinction between directed versus overheard speech had no effect on the children's vowel variability, this would suggest that the children were forming phonological categories from all ambient (maternal) language.

Evaluating this question of overheard versus directed speech in a bilingual context could be of further interest because the proportion of directed versus overheard speech likely varies by language. For example, most of a child's Spanish exposure could be child-directed but their Quechua exposure could primarily be overheard (as would be common in many situations of generational language shift). It is plausible that the child's vowel patterning in each language would reflect this difference in input, with Spanish vowels (learned through directed speech) being less variable than Quechua vowels (learned through overheard speech).[18] It is also possible that the distinction between overheard and directed speech is either meaningless for vowel variability, or, alternatively, highly dependent upon the cultural context. The point here is that a bilingual context uniquely permits evaluation of the roles of directed versus overheard speech in phonological development, while controlling for the individual child. This inquiry is a crucial next step for this line of research.

Estimating bilingual environments over multiple days

One important consideration in the use of daylong recordings are the sampling biases built into the methodology. Though the data samples from daylong recordings are robust, and arguably more representative than shorter recordings or recordings made in the lab, even a 16-hour recording in the home does not capture the complexities of a child's language learning environment. The current work computed environmental factors like the target child's Spanish use from a single daylong recording. However, it is unclear if or how these bilingual environment measures might differ from one day of recording to the next.

Addressing this question of recording sampling, Anderson and Fausey (2019) compared environmental language measures taken from multiple days of daylong recordings in infants aged 0;6-1;0. The authors found that the number of words spoken by adults in the target child's vicinity varied by the day of recording, suggesting that some environmental measures

[18] It is important to acknowledge that Quechua and Spanish vowels systems, while very similar, differ in the phonemic status of /e/ and /o/. This could affect the variability of those vowels, independent of learning context. It is likewise important to acknowledge that comprehensive examinations of vowel contrasts in infant-directed speech actually find a slight tendency for mothers to hypoarticulate in the register (Cristia and Seidl 2014; Martin et al. 2015). This means that perhaps the Spanish vowel categories are more reduced, even in a child-directed register.

do indeed vary by day. (The bilingual environment was not measured in Anderson and Fausey (2019); the children were from monolingual English households.)

For bilingual language environments, Orena et al. (2019) and Orena et al. (2020) measured French-English infants' (aged 0;10) language exposure from daylong recordings sampled over three days (typically two weekdays and one weekend day). However, the authors did not evaluate if the bilingual environment measures (language of exposure, etc.) actually differed across the three days, as the studies' objectives were to validate the Language ENvironment Analysis system (Orena et al. 2019) or compare parental reports of bilingual exposure to those derived from daylong recordings (Orena et al. 2020). As a result, it remains an open question how the bilingual language environment varies by day sampled and if these differences have a noticeable effect upon the overall rates of language exposure. If researchers are to employ daylong recordings to make ecologically-valid estimates of bilingual language environments - and the sampling methods employed in this study demonstrate that this is a realistic possibility for labs going forward - then it is crucial to understand how widely these language estimates vary by recording day and if sampling across multiple days is necessary.

It is almost certainly true that the bilingual language environment varies by day of the week or day of the month sampled. For example, when the child participants from this study travel to more rural areas, where more monolingual Quechua speakers live, the children are more likely to speak Quechua. Similarly, the children use almost exclusively Spanish at school.[19] Acknowledging this, researchers collecting daylong audio recording corpora encourage parents to collect the recording on a "typical" day. Still, this discussion has highlighted that a more thorough evaluation of bilingual measures derived from daylong recordings is necessary. This evaluation should ascertain if and how the bilingual environment varies across multiple recording days, including days that the child does and does not attend school as well as weekdays versus weekends. This work could also ascertain how the bilingual environment varies within a single recording day (e.g. morning versus night). The Cychosz corpus contains recordings from children across multiple distinct days, so future studies using this corpus and other available daylong recording corpora should be able to explore day-of-the-week effects in more detail.

4.6 Conclusion

Together, the results from this chapter have led to three important conclusions. First, the language environments of bilingual Quechua-Spanish children vary greatly with respect to the presence and use of each of the children's languages. Second, individual differences in bilingual language environments explain variability in children's phonetic production. And third, the role of the environment varies by the type of language experience (expressive versus receptive) and phonetic outcome measure (vowel category variability versus coarticulation/duration by word environment). Specifically, children's within-category vowel

[19]Neither of these contexts were captured in the recordings annotated for this project. These examples just demonstrate how certain linguistic contexts may be unrepresented in one or even a few daylong recordings.

variability was predicted by a receptive language measure, their mother's bilingual language profile: children with monolingual mothers had tighter, more compact vowel categories in Quechua than children with bilingual Quechua-Spanish mothers. Elsewhere, maternal language profile did not predict individual differences in children's spoken language patterns by morphological environment. However, an expressive language measure, the children's own use of Spanish and Quechua/Mixed speech, did. Children who used more Quechua/Mixed speech tended to differentiate more between morphological environments in their spoken language patterns, suggesting that they were better able to analyze the internal structure of morphologically complex words.

Taken together, the relationship between these spoken language patterns and the bilingual language environment suggests that the abstraction of linguistic categories such as phonemes and morphemes may occur gradually, with experience. However, as evidenced by the successful performance on the morphological extension task by *all* children, these linguistic categories must nevertheless be somewhat abstract, even for children with relatively less Quechua experience.

4.7 Appendices

Descriptive analyses

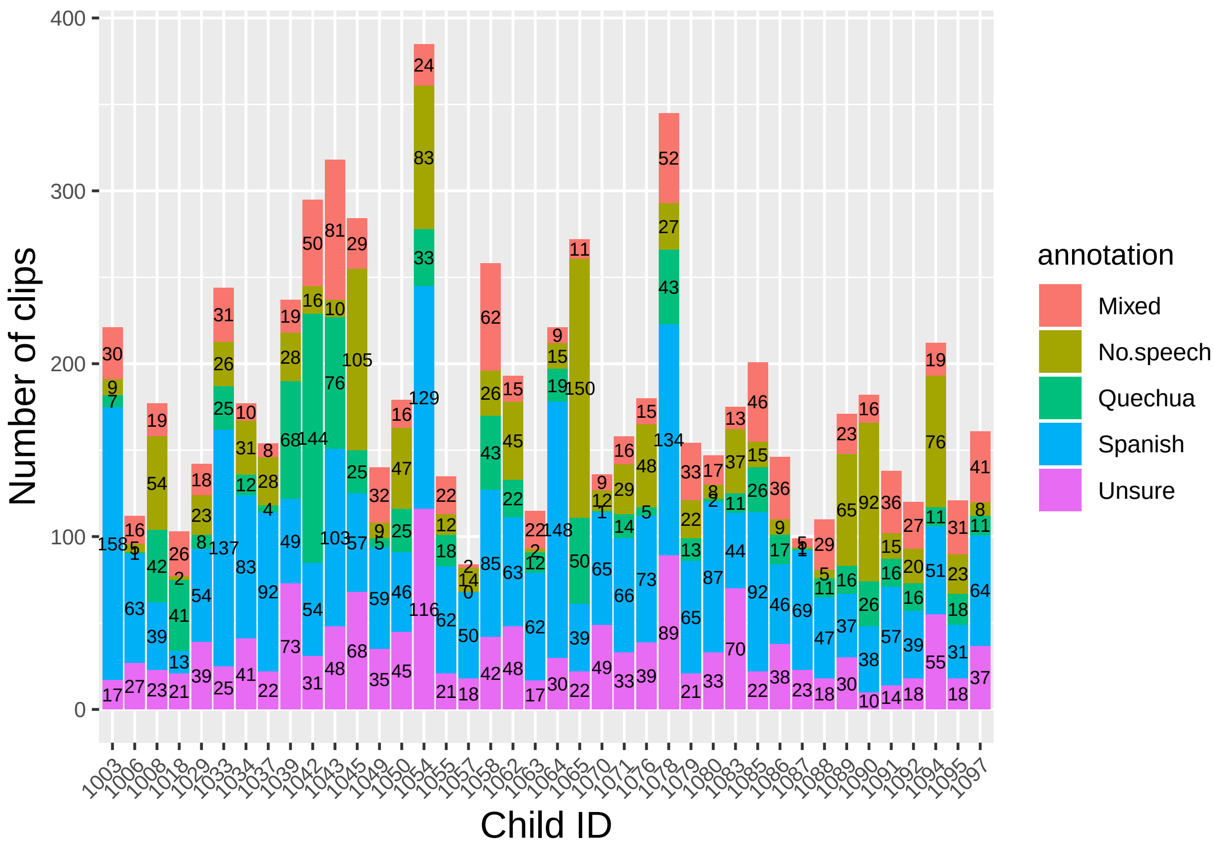


Figure 4.17: Clip annotation category counts for each child. Figures on barplot reflect the number of clips from each category.

Table 4.19: Clip annotation category counts and percentages for each child participant.

Speaker ID	Mixed	Unsure	No speech	Quechua	Spanish	Total # of clips
1003	30 (13.57 %)	17 (7.69 %)	9 (4.07 %)	7 (3.17 %)	158 (71.49 %)	221
1006	16 (14.29 %)	27 (24.11 %)	5 (4.46 %)	1 (0.89 %)	63 (56.25 %)	112
1008	19 (10.73 %)	23 (12.99 %)	54 (30.51 %)	42 (23.73 %)	39 (22.03 %)	177
1018	26 (25.24 %)	21 (20.39 %)	2 (1.94 %)	41 (39.81 %)	13 (12.62 %)	103
1029	18 (12.68 %)	39 (27.46 %)	23 (16.2 %)	8 (5.63 %)	54 (38.03 %)	142
1033	31 (12.7 %)	25 (10.25 %)	26 (10.66 %)	25 (10.25 %)	137 (56.15 %)	244
1034	10 (5.65 %)	41 (23.16 %)	31 (17.51 %)	12 (6.78 %)	83 (46.89 %)	177
1037	8 (5.19 %)	22 (14.29 %)	28 (18.18 %)	4 (2.6 %)	92 (59.74 %)	154
1039	19 (8.02 %)	73 (30.8 %)	28 (11.81 %)	68 (28.69 %)	49 (20.68 %)	237
1042	50 (16.95 %)	31 (10.51 %)	16 (5.42 %)	144 (48.81 %)	54 (18.31 %)	295
1043	81 (25.47 %)	48 (15.09 %)	10 (3.14 %)	76 (23.9 %)	103 (32.39 %)	318
1045	29 (10.21 %)	68 (23.94 %)	105 (36.97 %)	25 (8.8 %)	57 (20.07 %)	284
1049	32 (22.86 %)	35 (25 %)	9 (6.43 %)	5 (3.57 %)	59 (42.14 %)	140
1050	16 (8.94 %)	45 (25.14 %)	47 (26.26 %)	25 (13.97 %)	46 (25.7 %)	179
1054	24 (6.23 %)	116 (30.13 %)	83 (21.56 %)	33 (8.57 %)	129 (33.51 %)	385
1055	22 (16.3 %)	21 (15.56 %)	12 (8.89 %)	18 (13.33 %)	62 (45.93 %)	135
1057	2 (2.38 %)	18 (21.43 %)	14 (16.67 %)	0 (0 %)	50 (59.52 %)	84
1058	62 (24.03 %)	42 (16.28 %)	26 (10.08 %)	43 (16.67 %)	85 (32.95 %)	258
1062	15 (7.77 %)	48 (24.87 %)	45 (23.32 %)	22 (11.4 %)	63 (32.64 %)	193
1063	22 (19.13 %)	17 (14.78 %)	2 (1.74 %)	12 (10.43 %)	62 (53.91 %)	115
1064	9 (4.07 %)	30 (13.57 %)	15 (6.79 %)	19 (8.6 %)	148 (66.97 %)	221
1065	11 (4.04 %)	22 (8.09 %)	150 (55.15 %)	50 (18.38 %)	39 (14.34 %)	272
1070	9 (6.62 %)	49 (36.03 %)	12 (8.82 %)	1 (0.74 %)	65 (47.79 %)	136
1071	16 (10.13 %)	33 (20.89 %)	29 (18.35 %)	14 (8.86 %)	66 (41.77 %)	158
1076	15 (8.33 %)	39 (21.67 %)	48 (26.67 %)	5 (2.78 %)	73 (40.56 %)	180
1078	52 (15.07 %)	89 (25.8 %)	27 (7.83 %)	43 (12.46 %)	134 (38.84 %)	345
1079	33 (21.43 %)	21 (13.64 %)	22 (14.29 %)	13 (8.44 %)	65 (42.21 %)	154
1080	17 (11.56 %)	33 (22.45 %)	8 (5.44 %)	2 (1.36 %)	87 (59.18 %)	147
1083	13 (7.43 %)	70 (40 %)	37 (21.14 %)	11 (6.29 %)	44 (25.14 %)	175
1085	46 (22.89 %)	22 (10.95 %)	15 (7.46 %)	26 (12.94 %)	92 (45.77 %)	201
1086	36 (24.66 %)	38 (26.03 %)	9 (6.16 %)	17 (11.64 %)	46 (31.51 %)	146
1087	5 (5.05 %)	23 (23.23 %)	1 (1.01 %)	1 (1.01 %)	69 (69.7 %)	99
1088	29 (26.36 %)	18 (16.36 %)	5 (4.55 %)	11 (10 %)	47 (42.73 %)	110
1089	23 (13.45 %)	30 (17.54 %)	65 (38.01 %)	16 (9.36 %)	37 (21.64 %)	171
1090	16 (8.79 %)	10 (5.49 %)	92 (50.55 %)	26 (14.29 %)	38 (20.88 %)	182
1091	36 (26.09 %)	14 (10.14 %)	15 (10.87 %)	16 (11.59 %)	57 (41.3 %)	138
1092	27 (22.5 %)	18 (15 %)	20 (16.67 %)	16 (13.33 %)	39 (32.5 %)	120
1094	19 (8.96 %)	55 (25.94 %)	76 (35.85 %)	11 (5.19 %)	51 (24.06 %)	212
1095	31 (25.62 %)	18 (14.88 %)	23 (19.01 %)	18 (14.88 %)	31 (25.62 %)	121
1097	41 (25.47 %)	37 (22.98 %)	8 (4.97 %)	11 (6.83 %)	64 (39.75 %)	161

Vowels

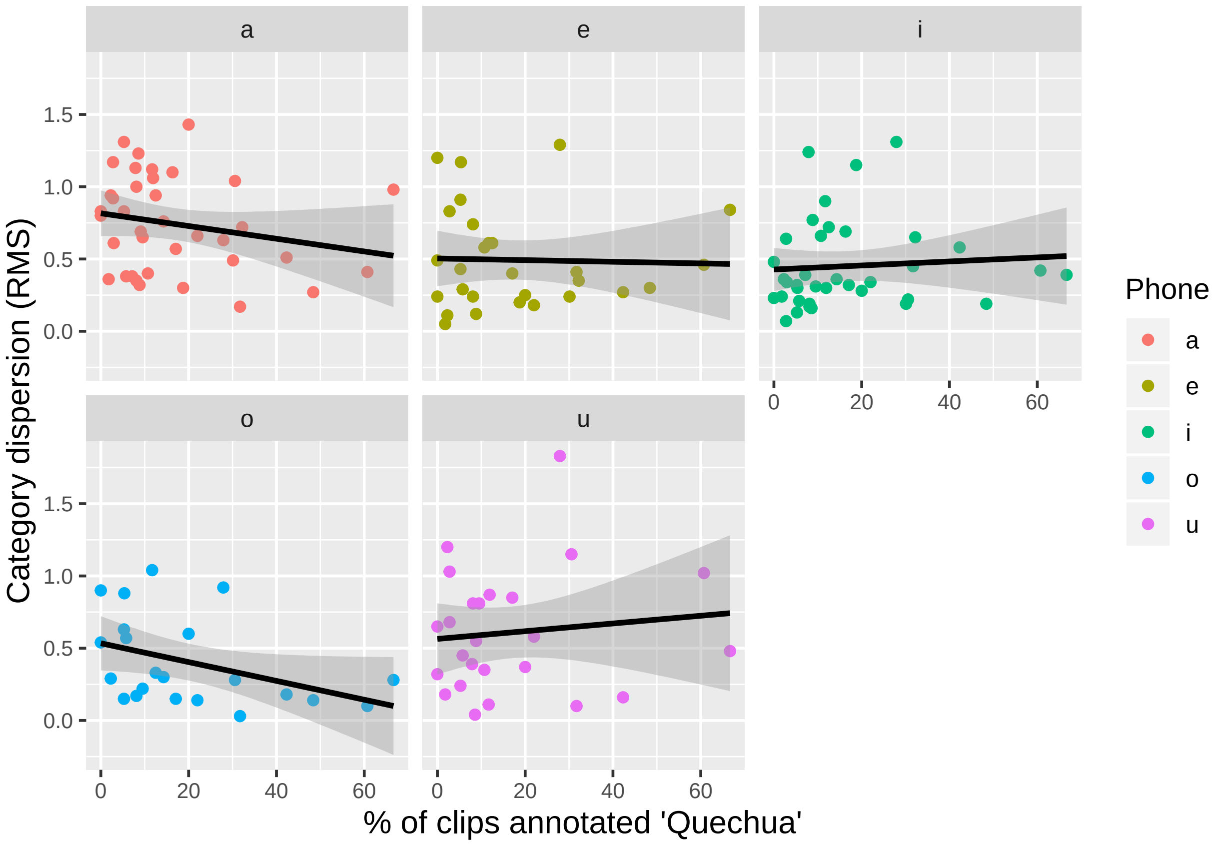


Figure 4.18: Vowel category dispersion (RMS) by percentage of monolingual Quechua clips containing target child and phone.

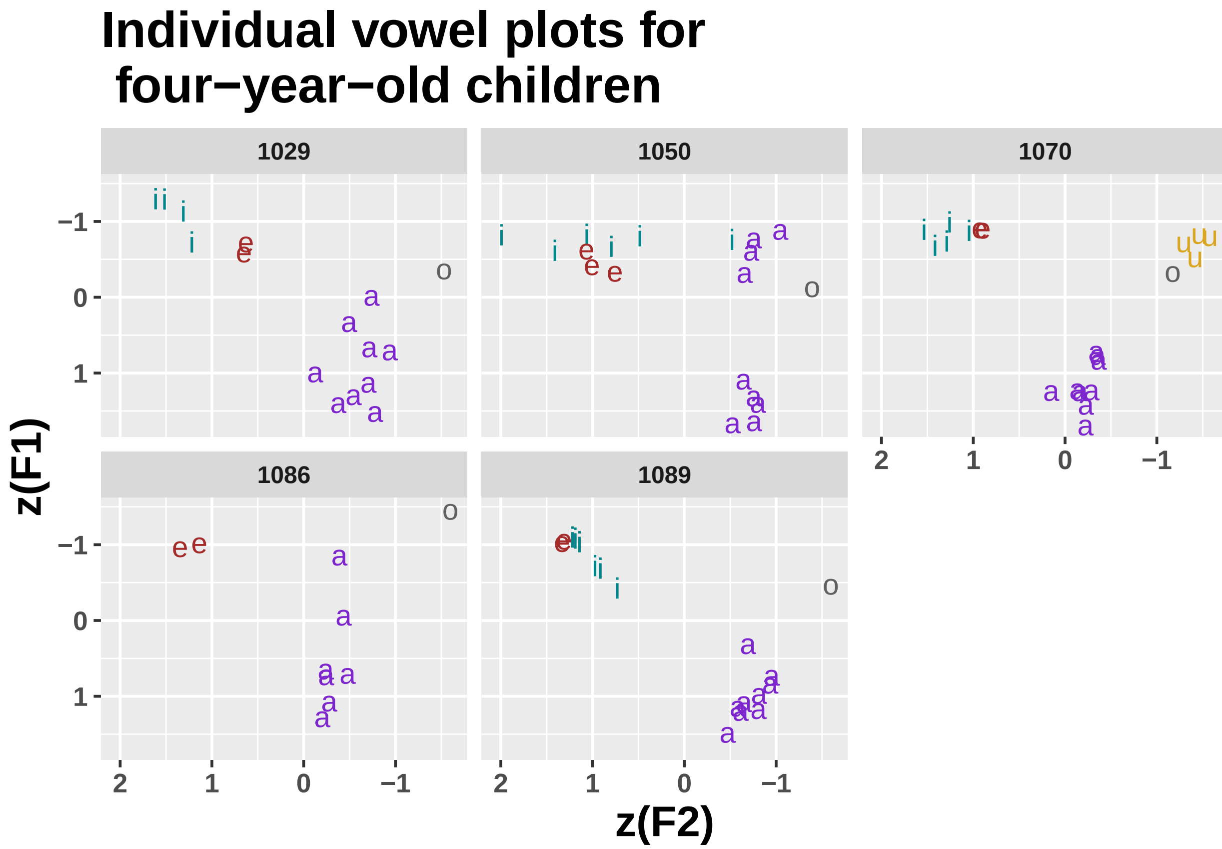


Figure 4.19: Individual vowel plots for four-year-olds.

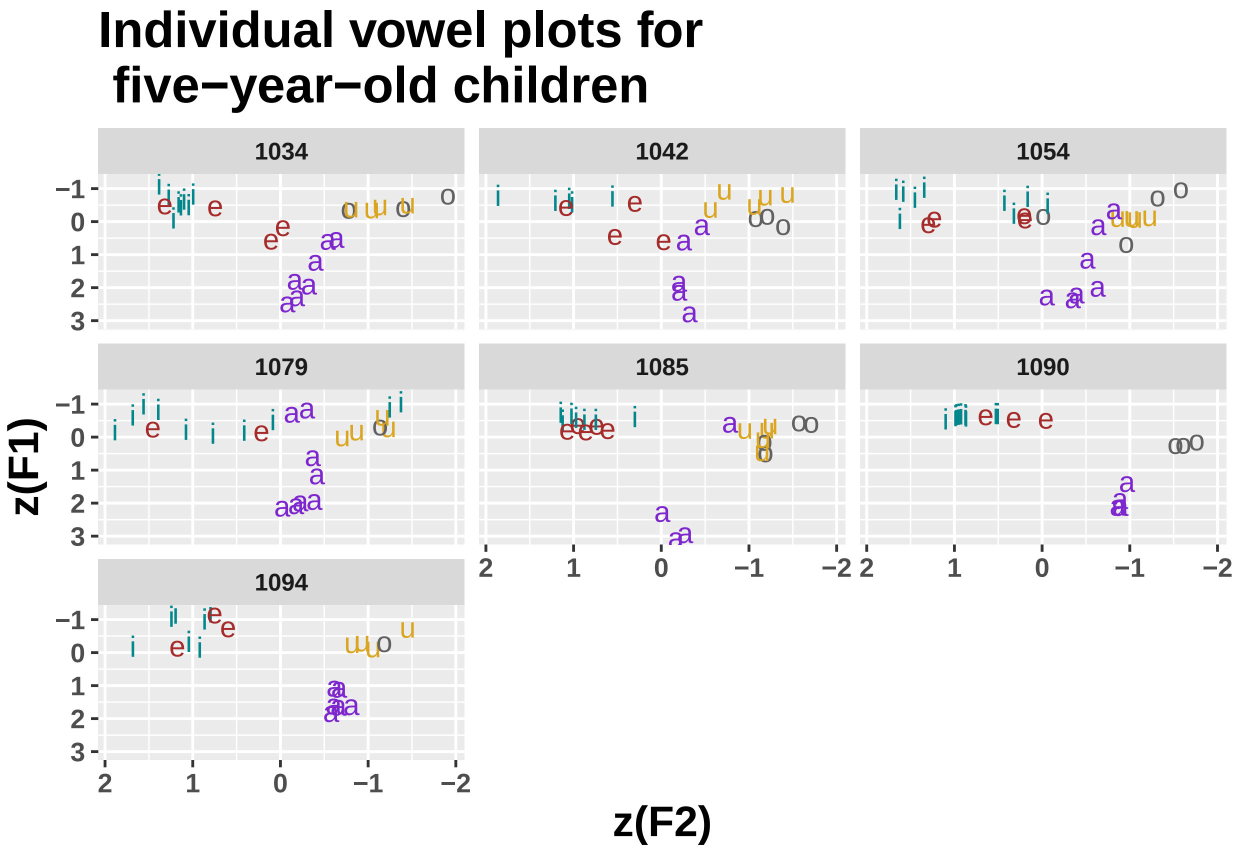


Figure 4.20: Individual vowel plots for five-year-olds.

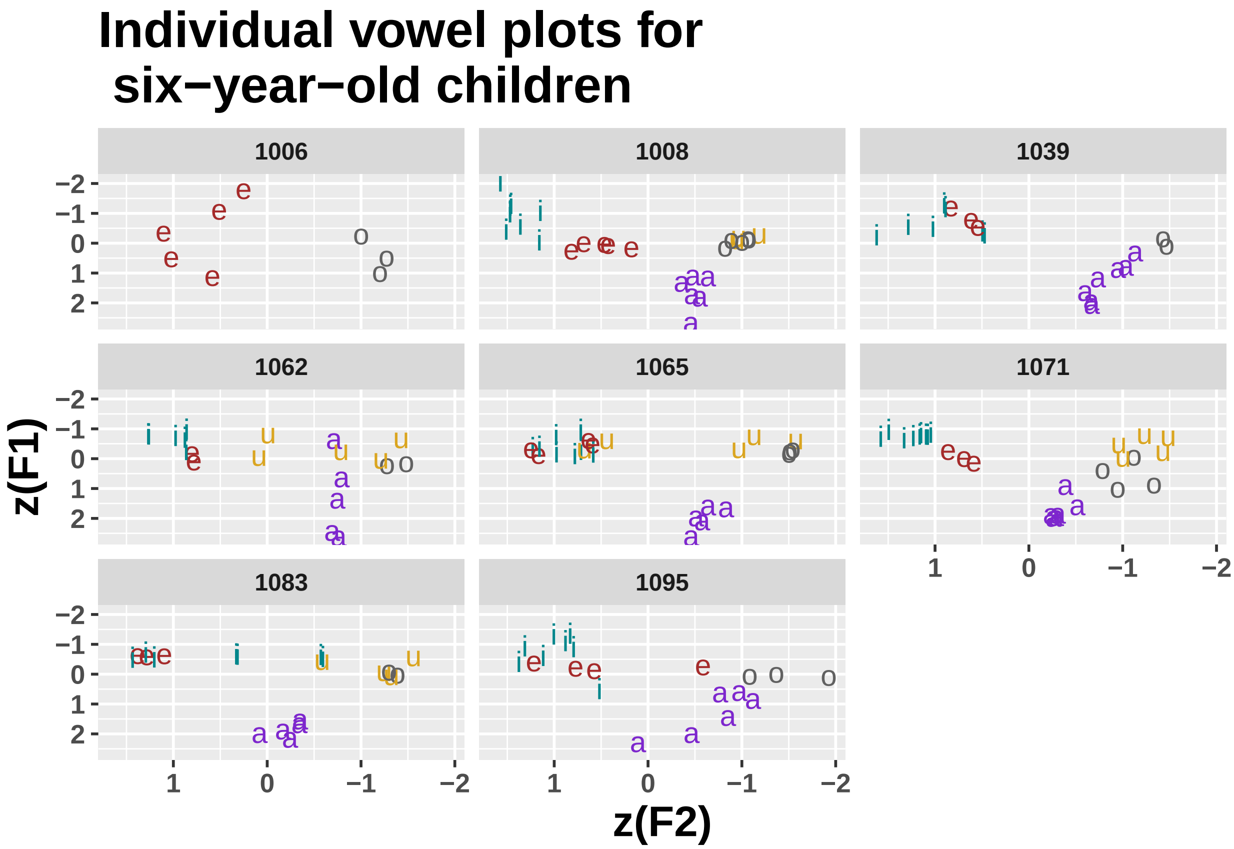


Figure 4.21: Individual vowel plots for six-year-olds.

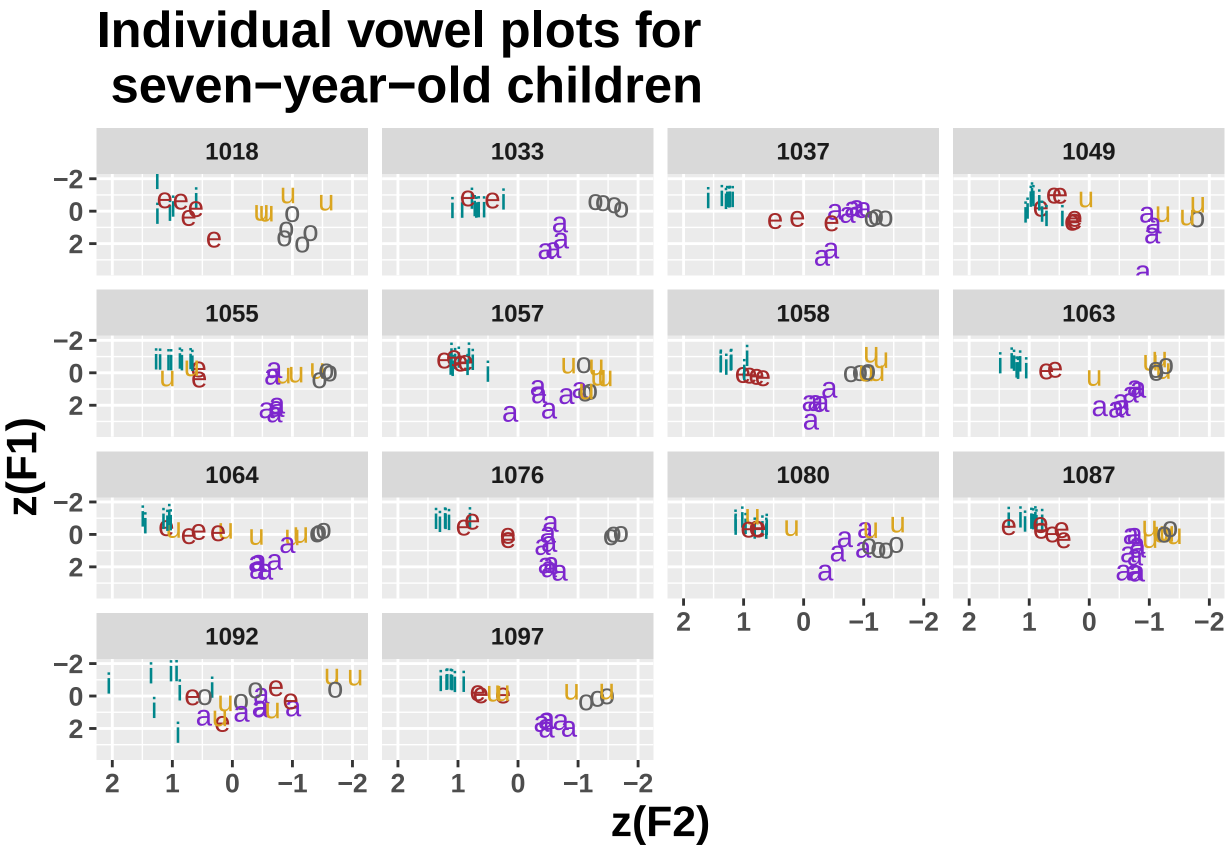


Figure 4.22: Individual vowel plots for seven-year-olds.

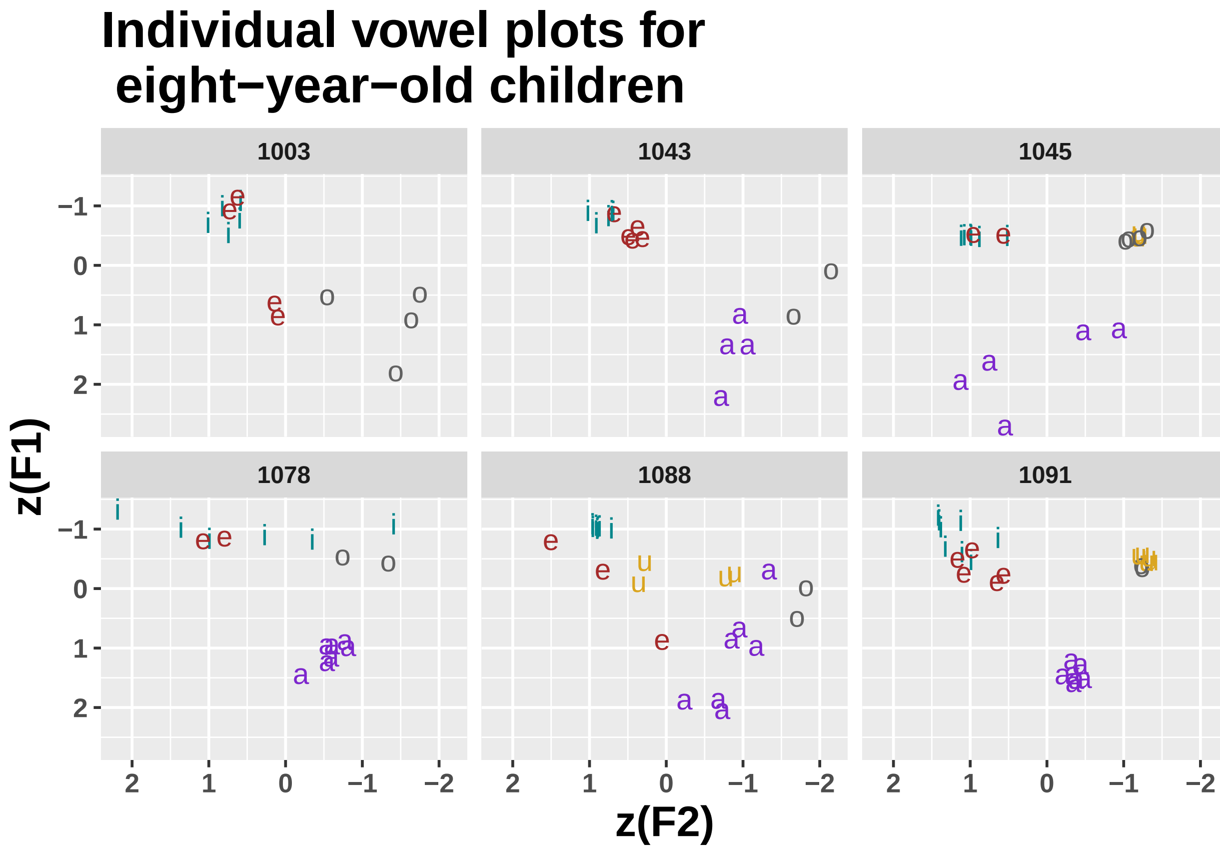


Figure 4.23: Individual vowel plots for eight-year-olds.

Morphology

Table 4.20: Model predicting coarticulation difference by maternal language profile

Intercept	0.91
	(−0.06, 1.87)
Biphone sequence:[ap]	1.99***
	(1.22, 2.76)
Lang. profile: Quechua-dominant	0.11
	(−1.32, 1.54)
Lang. profile: Bilingual Quechua-Spanish	−0.50
	(−1.56, 0.57)
Observations	62
Log Likelihood	−117.64
Akaike Inf. Crit.	247.28
Bayesian Inf. Crit.	260.05
Note:	*p<0.05; **p<0.01; ***p<0.001

Table 4.21: Model predicting durational difference by maternal language profile

Intercept	40.35^{**}
	(15.39, 65.31)
Biphone sequence:[ap]	58.26^{***}
	(45.90, 70.61)
Lang. profile: Quechua-dominant	-17.94
	(-56.93, 21.05)
Lang. profile: Bilingual Quechua-Spanish	-5.73
	(-34.79, 23.33)
Observations	62
Log Likelihood	-293.52
Akaike Inf. Crit.	599.04
Bayesian Inf. Crit.	611.80
Note:	*p<0.05; **p<0.01; ***p<0.001

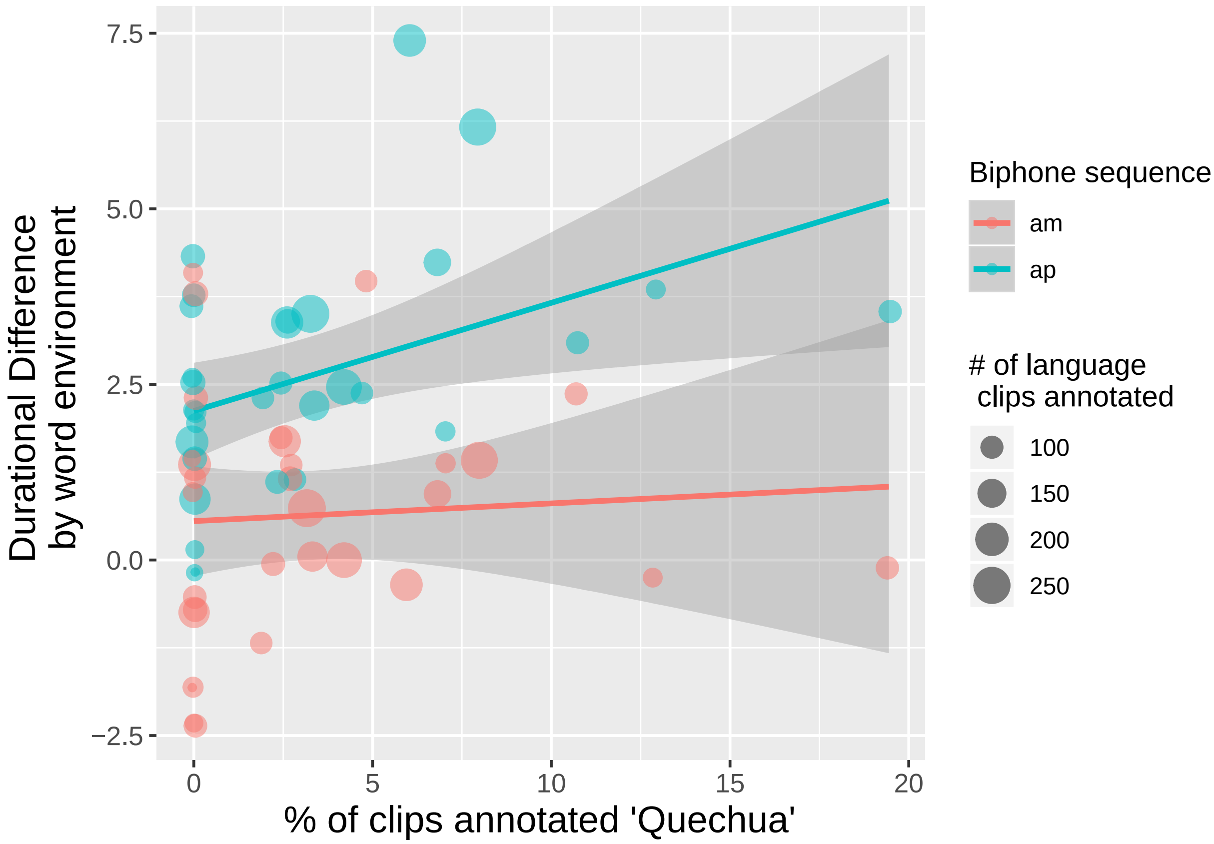


Figure 4.24: Coarticulation difference by the percentage of Quechua clips in the recording.

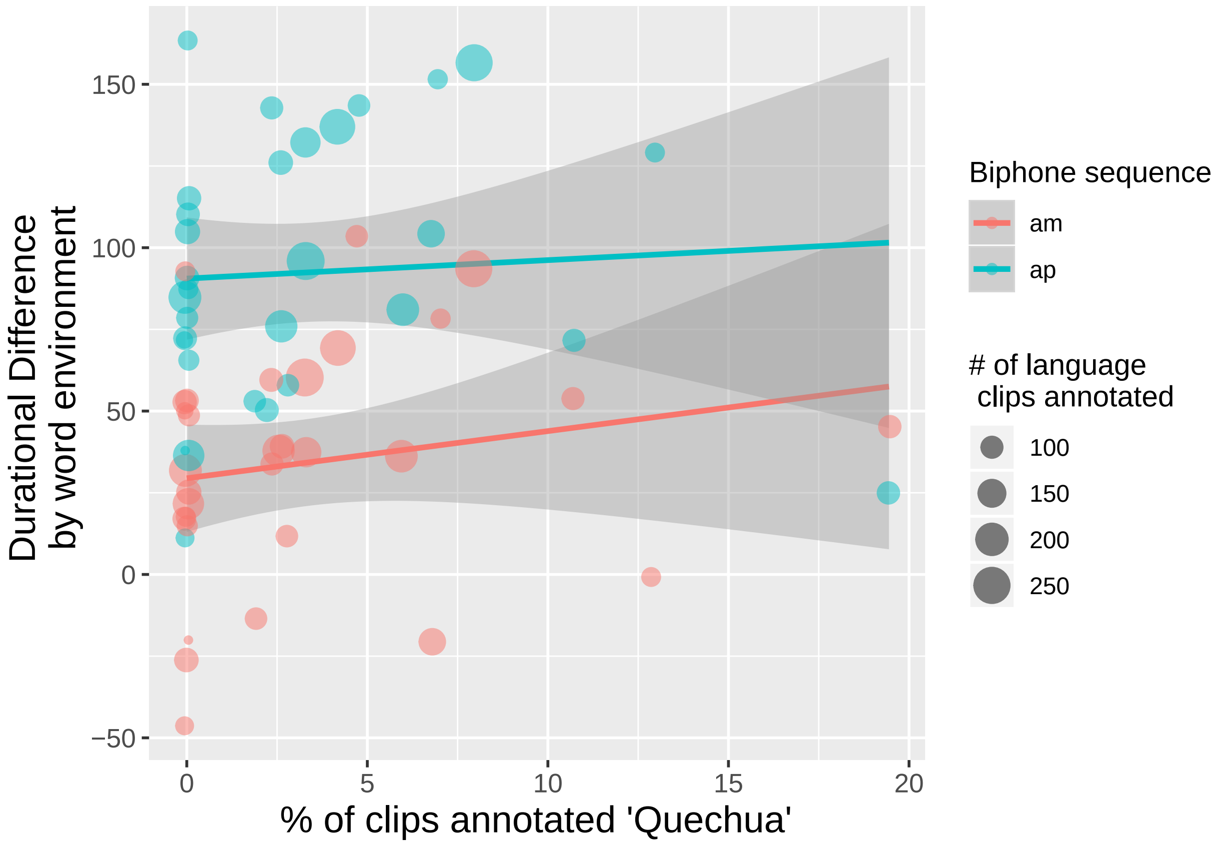


Figure 4.25: Durational difference by the percentage of monolingual Quechua clips in the recording.

Conclusion

This dissertation evaluated how children's ambient language environment may affect their spoken language development. To that end, the studies presented here measured the phonetic development of bilingual Quechua-Spanish children living in a mid-size town in Bolivia. This bilingual population presented a unique setting in which to test several outstanding developmental questions: namely, what is the role of linguistic structure - phonological inventory and morphological structure - and quantity of language exposure for children's language development?

Chapters 2 and 3 examined the role of linguistic structure. Chapter 2 evaluated how Quechua's phonological inventory may constrain children's vowel variability over the course of development for four- through ten-year-olds. It was found that in a language like Quechua, with three phonemic vowel contrasts, children as young as four are able to approximate adult-like levels of vowel variability. However, this chapter did not argue that children necessarily speak like adults. Rather, the vowel data in chapter 2 were additionally used to evaluate how well children articulatorily compensate for their distinct vocal tract morphologies. It was found that even some of the ten-year-olds studied did not approximate adult-like formant frequency ratios between the front and back cavities in the vocal tract, further suggesting that children are constantly - and at times inaccurately - updating their acoustic-articulatory mappings in light of non-uniform anatomical changes in vocal tract morphology.

Chapter 3 likewise evaluated the potential role of linguistic structure for children's speech development. Specifically, chapter 3 studied how Quechua's particular morphological structure interacts with children's known tendency to coarticulate more than adults. The results showed complex relationships between word structure and coarticulation, as both children and adults coarticulated more within morphemes than across morphemes. This, it was argued, indicated that children and adults may both decompose morphologically complex words. However, only children compensated for the words' prosodic structure; that is, only children, and not adults, shortened word duration as additional suffixes were appended to roots. Thus, it was unclear if the children's patterning by word environment reflected prosodic structure, morphological structure, or both. Furthermore, the difference between adults and children, it was suggested, could best be explained by adults' more frequent Quechua use and greater dominance in the language.

Finally, chapter 4 evaluated how the quantity of language that children are exposed to and use may influence their speech development. In the bilingual population studied here, different children are exposed to differing amounts of Quechua, Spanish, and mixed Quechua-Spanish speech. First, using the vowel data from chapter 2, it was found that children with monolingual Quechua mothers had tighter, less variable vowel categories than children with Quechua-dominant mothers or bilingual Quechua-Spanish mothers. Then, using daylong audio recordings of the children's linguistic environments, chapter 4 found that children who use more Quechua throughout the day distinguish between the within-morpheme and across-morpheme word environments studied in chapter 3. This indicates that children who use more Quechua are better able to analyze and break down the internal structure of morphologically complex words. Crucially, differences in the children's language usage were estimated by sampling randomly from the daylong recordings. As a result, it was only necessary to annotate an average of 90 minutes from each recording, suggesting that this ecologically-valid method of evaluating bilingual language dominance may be feasible for labs going forward.

In all, this dissertation has evaluated the role of two environmental factors - linguistic structure and quantity of language exposure - for children's spoken language development. Linguistic structure appears to interact a great deal with speech development, at least in the population studied here. A large takeaway from this work is that developmental assumptions, such as children's high acoustic variability and tendency to coarticulate more than adults, may need to be considered in line with the language(s) children are acquiring. Another takeaway is that the child's particular learning environment, and quantity of language exposure, influences their phonetic development. Bilingual children who use more of one of their languages exhibit more adult-like speech patterning in that language. How long these discrepancies by child language dominance last, and how widespread these developmental patterns may be for other languages, remain open avenues for future research.

Bibliography

Adank, P., Smits, R., and van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, 116(5):3099–3107.

Adriaans, F. and Swingley, D. (2012). Distributional Learning of Vowel Categories is Supported by Prosody in Infant-Directed Speech. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*, Sapporo, Japan.

Agwuele, A., Sussman, H. M., and Lindblom, B. (2008). The effect of speaking rate on consonant vowel coarticulation. *Phonetica*, 65(4):194–209.

Ambridge, B., Kidd, E., Rowland, C. F., and Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, 42(02):239–273.

Anderson, H. and Fausey, C. (2019). Modeling non-uniformities in infants' everyday speech environments. Paper presented at the 2019 Biennel Meeting of the Society for Research on Child Development.

Anderson, J. D. (2007). Phonological neighborhood and word frequency effects in the stuttered disfluencies of children who stutter. *Journal of Speech Language and Hearing Research*, 50(1):229–247.

Arnon, I. (2010). *Starting Big: The Role of Multiword Phrases*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Arnon, I. and Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1-L2 differences. *Topics in Cognitive Science*, 9(3):621–636.

Arnon, I. and Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, 56(3):349–371.

Atal, B. and Schroeder, M. (1974). Recent advances in predictive coding—applications to speech synthesis. In *Proceedings of Speech Communication Symposium 1974*, pages 27–31.

Aust, F. and Barth, M. (2018). *Papaja: Create APA Manuscripts with R Markdown*. Retrieved from: `https://crsh.github.io/papaja_man/`.

Aylett, M. and Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1):31–56.

Baayen, H. R. (1992). Quantitative aspects of morphological productivity. In Booij, G. and van Marle, J., editors, *Yearbook of Morphology 1991*, pages 109–149. Kluwer Academic Publishers, Dordrecht.

Baayen, H. R., McQueen, J. M., Dijkstra, T., and Schreuder, R. (2003). Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In Baayen, H. R. and Schreuder, R., editors, *Morphological Structure in Language Processing*, pages 355–390. De Gruyter Mouton, Berlin, New York.

Bannard, C. and Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, 19(3):241–248.

Barbier, G., Perrier, P., Ménard, L., Payan, Y., Tiede, M. K., and Perkell, J. S. (2013). Speech planning as an index of speech motor control maturity. In *Proceedings of Interspeech 2013*, Lyon, France.

Barbier, G., Perrier, P., Ménard, L., Payan, Y., Tiede, M. K., and Perkell, J. S. (2015). Speech planning in 4-year-old children versus adults: Acoustic and articulatory analyses. In *Proceedings of Interspeech 2015*, Dresden, Germany.

Barbier, G., Perrier, P., Payan, Y., Tiede, M. K., Gerber, S., Perkell, J. S., and Ménard, L. (2020). What anticipatory coarticulation in children tells us about speech motor control maturity. *PLoS ONE*, 15(4):e0231484.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Beddor, P. S., Coetzee, A. W., and Styler, W. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language*, 94(4):931–968.

Benson, C. (2004). Bilingual schooling in Mozambique and Bolivia: From experimentation to implementation. *Language Policy*, 3(1):47–66.

Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., and Tor, S. (2019a). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, 22(1):e12715.

Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., and Amatuni, A. (2019b). What Do North American Babies Hear? A large-scale cross-corpus analysis. *Developmental Science*, 22(1):e12724.

Bergmann, C., Tsuji, S., and Cristia, A. (2017). Top-down versus bottom-up theories of phonological acquisition: A big data approach. In *Interspeech 2017*, pages 2103–2107, Stockholm, Sweden. International Speech and Communication Association.

Bergmann, T., Dale, R., and Lupyan, G. (2016). Socio-demographic influences on language structure and change: Not all learners are the same (commentary on Christiansen & Chater). *Behavioral and Brain Sciences*, 39:22–23.

Berko, J. (1958). The child's learning of English morphology. *Word*, 14(2-3):150–177.

Bernhardt, B. H. and Stemberger, J. P., editors (1998). *Handbook of Phonological Development: From the Perspective of Constraint-Based Nonlinear Phonology.* Academic Press, New York, NY.

Best, C. T. and Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In Bohn, O.-S. and Munro, M. J., editors, *Language Learning & Language Teaching*, volume 17, pages 13–34. John Benjamins Publishing Company, Amsterdam.

Bijeljac-Babic, R., Serres, J., Höhle, B., and Nazzi, T. (2012). Effect of bilingualism on lexical stress pattern discrimination in French-learning infants. *PLoS ONE*, 7(2):e30843.

Birdson, D., Gertken, L. M., and Amengual, M. (2012). *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism.* COERLL, University of Texas at Austin. Retrieved from: `https://sites.la.utexas.edu/bilingual/`.

Boersma, P. and Weenik, D. (2019). Praat: Doing phonetics by computer. Retrieved from: `http://www.fon.hum.uva.nl/praat/`. Version 6.1.03.

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *The Journal of the Acoustical Society of America*, 97(3):1916–1924.

Bradlow, A. R. (2002). Confluent talker- and listener-oriented forces in clear speech production. In Gussenhoven, C. and Warner, N., editors, *Laboratory Phonology*, volume 7, pages 241–273. Mouton de Gruyter, Berlin and New York.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., and Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9(2):378–400.

Browman, C. P. and Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6(2):201–251.

Browman, C. P. and Goldstein, L. (1992). Articulatory Phonology: An overview. *Phonetica*, 49(3-4):155–180.

Bruderer, A. G., Danielson, D. K., Kandhadai, P., and Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences*, 112(44):13531–13536.

Bybee, J. L. and Slobin, D. (1982). Why small children cannot change language on their own: Suggestions from the English past tense. In Alqvist, A., editor, *Papers from the Fifth International Conference on Historical Linguistics*, pages 29–37, Amsterdam. John Benjamins.

Byers-Heinlein, K. (2013). Parental language mixing: Its measurement and the relation of mixed input to young bilingual children's vocabulary size. *Bilingualism: Language and Cognition*, 16(1):32–48.

Carbajal, M. J. and Peperkamp, S. (2019). Dual language input and the impact of language separation on early lexical development. *Infancy*, 25(1):22–45.

Casillas, M., Brown, P., and Levinson, S. (2019). Early language experience in a Tseltal Mayan village. *Child Development*, 0:1–17.

Caudrelier, T., Ménard, L., Perrier, P., Schwartz, J.-L., Gerber, S., Vidou, C., and Rochet-Capellan, A. (2019). Transfer of sensorimotor learning reveals phoneme representations in preliterate children. *Cognition*, 192:103973.

Charles-Luce, J. and Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language*, 17(1):205–215.

Chen, W.-R., Whalen, D. H., and Shadle, C. H. (2019). F0-induced formant measurement errors result in biased variabilities. *The Journal of the Acoustical Society of America-Express Letters*, 145(5):EL360–EL366.

Chen, Y.-p., Keen, R., Rosander, K., and Hofsten, C. V. (2010). Movement planning reflects skill level and age changes in toddlers. *Child Development*, 81(6):1846–1858.

Cho, T. (2001). Effects of morpheme boundaries on intergestural timing: Evidence from Korean. *Phonetica*, 58(3):129–162.

Chomsky, N. (1959). A review of BF Skinner's Verbal Behavior. *Language*, 35(1):26–58.

Cohen, C. (2014). Probabilistic reduction and probabilistic enhancement: Contextual and paradigmatic effects on morpheme pronunciation. *Morphology*, 24(4):291–323.

Colé, P., Segui, J., and Taft, M. (1997). Words and morphemes as units for lexical access. *Journal of Memory and Language*, 37(3):312–330.

Cooper, A., Fecher, N., and Johnson, E. K. (2018). Toddlers' comprehension of adult and child talkers: Adult targets versus vocal tract similarity. *Cognition*, 173:16–20.

Cournane, A. (2017). In defence of the child innovator. In Mathieu, E. and Truswell, R., editors, *Micro-Change and Macro-Change in Diachronic Syntax*, pages 10–23. Oxford University Press, Oxford, UK, first edition.

Cournane, A. (2019). A developmental view on incrementation in language change. *Theoretical Linguistics*, 45(3-4):127–150.

Courtney, E. H. and Saville-Troike, M. (2002). Learning to construct verbs in Navajo and Quechua. *Journal of Child Language*, 29(3):623–654.

Crelin, E. S. (1987). *The Human Vocal Tract: Anatomy, Function, Development, and Evolution*. Vantage Press, New York.

Cristia, A. (2020). Language input and outcome variation as a test of theory plausibility: The case of early phonological acquisition. *Developmental Review*.

Cristia, A., Dupoux, E., Gurven, M., and Stieglitz, J. (2017). Child-directed speech is infrequent in a forager-farmer population: A time allocation study. *Child Development*, 90(3):759–773.

Cristia, A. and Seidl, A. (2014). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language*, 41(4):913–934.

Cychosz, M. (2018). *Cychosz HomeBank Corpus*. DOI: 10.21415/YFYW-HE74.

Cychosz, M. (2019). Holistic lexical storage: Coarticulatory evidence from child speech. In *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia.

Cychosz, M., Cristia, A., Bergelson, E., Casillas, M., Baudet, G., Warlaumont, A. S., Scaff, C., Yankowitz, L., and Seidl, A. (under review). Canonical babble development in a large-scale, crosslinguistic corpus.

Cychosz, M., Edwards, J. R., Munson, B., and Johnson, K. (2019). Spectral and temporal measures of coarticulation in child speech. *The Journal of the Acoustical Society of America-Express Letters*, 146(6):EL516–EL522.

Cychosz, M., Erskine, M., Munson, B., and Edwards, J. R. (2020a). A lexical advantage in four-year-old children's word repetition. *Journal of Child Language*.

Cychosz, M. and Kalt, S. (2018). Language-specific sources of acoustic stability in phonological development. In *Proceedings from the 42nd Boston University Conference on Language Development*, Boston, MA.

Cychosz, M., Munson, B., and Edwards, J. R. (2020b). Children talking also matters: Language practice and experience predict speech production patterns. Paper to be presented at the 2021 International Association for the Study of Child Language.

Cychosz, M., Villanueva, A., and Weisleder, A. (2020c). Efficient estimation of bilingual children's language exposure from daylong audio recordings. *Manuscript in progress*.

Davis, M. and Redford, M. A. (2019). The emergence of discrete perceptual-motor units in a production model that assumes holistic phonological representations. *Frontiers in Psychology*, 10(2121):1–19.

de Boysson-Bardies, Sagart, L., and Durand, C. (1984). Discernible differences in the babbling of infants according to target language. *Journal of Child Language*, 11(1):1–15.

de Boysson-Bardies, B., Vihman, M. M., and de Boysson-Bardies, B. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, 67(2):297–319.

Demuth, K. (2006). Crosslinguistic perspectives on the development of prosodic words. *Language and Speech*, 49(2):129–135.

Denny, M. and McGowan, R. S. (2012a). Implications of peripheral muscular and anatomical development for the acquisition of lingual control for speech production: A review. *Folia Phoniatrica et Logopaedica*, 64(3):105–115.

Denny, M. and McGowan, R. S. (2012b). Sagittal area of the vocal tract in young female children. *Folia Phoniatrica et Logopaedica*, 64(6):297–303.

DePaolis, R. A., Vihman, M. M., and Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior and Development*, 34(4):590–601.

Di Benedetto, M.-G. (1994). Acoustic and perceptual evidence of a complex relation between F1 and F0 in determining vowel height. *Journal of Phonetics*, 22(3):205–224.

Diamanti, V., Mouzaki, A., Ralli, A., Antoniou, F., Papaioannou, S., and Protopapas, A. (2017). Preschool phonological and morphological awareness as longitudinal rredictors of early reading and spelling development in Greek. *Frontiers in Psychology*, 8(2039):1–12.

Diessel, H. (2011). Grammaticalization and language acquisition. In Heine, B. and Narrog, H., editors, *Handbook of Grammaticalization*, page 130–141. Oxford University Press, Oxford.

Drager, K. K. (2011). Sociophonetic variation and the lemma. *Journal of Phonetics*, 39(4):694–707.

Dunn, L. M. and Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test*. Pearson Assessments.

Edwards, J. and Beckman, M. E. (2008). Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language Learning and Development*, 4(2):122–156.

Edwards, J., Beckman, M. E., and Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in novel word repetition. *Journal of Speech Language and Hearing Research*, 57:421–436.

Edwards, J., Beckman, M. E., and Munson, B. (2015). Frequency effects in phonological acquisition. *Journal of Child Language*, 42(2):306–311.

Eguchi, S. and Hirsh, I. (1969). Development of speech sounds in children. *Acta Otolaryngolica*, (suppl 257):21–51.

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

Fant, G. (1966). A note on vocal tract size factors and non-uniform F-pattern scalings. *Speech Transactions Laboratory Quarterly Progress and Status Report*, 1:22–30.

Fant, G. (1975). Non-uniform vowel normalization. *Speech Transactions Laboratory Quarterly Progress and Status Report*, 2-3:1–19.

Feldman, N. H., Griffiths, T. L., Goldwater, S., and Morgan, J. L. (2013). A role for the developing lexicon in phonetic category acquisition. *Psychological Review*, 120(4):751–778.

Fenson, L., Marchman, V., Thal, D. J., Dale, P., Reznick, J., and Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories User's Guide and Technical Manual*. Singular, San Diego, CA, 2nd edition edition.

Ferguson, C. and Farwell, C. (1975). Words and sounds in early language acquisition. *Language*, 51(2):419–439.

Ferjan Ramírez, N., Lytle, S. R., Fish, M., and Kuhl, P. K. (2019a). Parent coaching at 6 and 10 months improves language outcomes at 14 months: A randomized controlled trial. *Developmental Science*, 22(3):e12762.

Ferjan Ramírez, N., Lytle, S. R., Fish, M., and Kuhl, P. K. (2019b). Parent coaching at 6 and 10 months improves language outcomes at 14 months: A randomized controlled trial. *Developmental Science*, 22(3):e12762.

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

Fernald, A. and Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. 20(1):104–113.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., and Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16(3):477–501.

Fikkert, P. and Levelt, C. (2008). How does Place fall into Place? The lexicon and emergent constraints in children's developing phonological grammar. In Dresher, B. E. and Rice, K., editors, *Contrast in Phonology: Theory, Perception, and Acquisition*, pages 231–270. Walter de Gruyter.

Fitch, W. T. and Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 106(3):1511–1522.

Flege, J. E. (1988). Anticipatory and carry-over nasal coarticulation in the speech of children and adults. *Journal of Speech Language and Hearing Research*, 31:525–536.

Fletcher, S. (1973). Maturation of the speech mechanism. *Folia Phoniatrica et Logopaedica*, 25(3):161–172.

Fourakis, M. (1991). Tempo, stress, and vowel reduction in American English. *The Journal of the Acoustical Society of America*, 90(4):1816–1827.

Fowler, C. A. (1981). A relationship between coarticulation and compensatory shortening. *Phonetica*, 38(1-3):35–50.

Fricke, M. and Johnson, K. (2012). Measuring coarticulation in spontaneous speech: A preliminary report. *UC Berkeley PhonLab Annual Report*, 8:306–320.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, 84(3):474–496.

Gallagher, G. (2016). Vowel height allophony and dorsal place contrasts in Cochabamba Quechua. *Phonetica*, 73(2):101–119.

Gavrilov, I. and Pusev, R. (2014). *Normtest: Tests for Normality*. `https://CRAN.R-project.org/package=normtest`.

Gay, T. (1981). Mechanisms in the control of speech rate. *Phonetica*, 38:148–158.

Gerosa, M., Lee, S., Giuliani, D., and Narayanan, S. (2006). Analyzing children's speech: An acoustic study of consonants and consonant-vowel transition. In *2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings*, volume 1, pages 393–396, Toulouse, France. IEEE.

Goffman, L., Smith, A., Heisler, L., and Ho, M. (2008). The breadth of coarticulatory units in children and adults. *Journal of Speech Language and Hearing Research*, 51(6):1424–1437.

Goldstein, U. G. (1980). *An Articulatory Model for the Vocal Tracts of Growing Children*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Goodell, E. W. and Studdert-Kennedy, M. (1992). Acoustic evidence for the development of gestural coordination in the speech of 2-year-olds: A longitudinal study. *Haskins Laboratories Status Report on Speech Research*, SR-111/1123:63–88.

Green, J. R., Moore, C. A., Higashikawa, M., and Steeve, R. W. (2000). The physiologic development of speech motor control: Lip and jaw coordination. *Journal of Speech Language and Hearing Research*, 43(1):239–255.

Greenwood, C. R., Thiemann-Bourque, K., Walker, D., Buzhardt, J., and Gilkerson, J. (2011). Assessing children's home language environments using automatic speech recognition technology. *Communication Disorders Quarterly*, 32(2):83–92.

Grigos, M. I. (2009). Changes in articulator movement variability during phonemic development: A longitudinal study. *Journal of Speech, Language, and Hearing Research*, 52(1):164–177.

Grossberg, S. (1978). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. *Progress in Theoretical Biology*, 5:233–274.

Guy, G. R. (1991). Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change*, 3(1):1–22.

Hale, M. and Reiss, C. (1998). Formal and empirical arguments concerning phonological acquisition. *Linguistic Inquiry*, 29(4):656–683.

Hanique, I. and Ernestus, M. (2012). The role of morphology in acoustic reduction. *Lingue e linguaggio*, 11(2):147–164.

Harrington, J., Kleber, F., Reubold, U., and Siddins, J. (2015). The relationship between prosodic weakening and sound change: Evidence from the German tense/lax vowel contrast. *Laboratory Phonology*, 6(1):87–117.

Hart, B. and Risley, T. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Paul H. Brookes Publishing, Baltimore, MD.

Hay, J. (2003). *Causes and Consequences of Word Structure*. Routledge, New York and London.

Hay, J. B., Pierrehumbert, J. B., Walker, A. J., and LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition*, 139:83–91.

Hazan, V. and Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, 28(4):377–396.

Hillenbrand, J., Getty, L. A., Wheeler, K., and Clark, M. J. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, 97(5):3099–3111.

Hlavac, M. (2018). *Stargazer: Well-Formatted Regression and Summary Statistics Tables*. Central European Labour Studies Institute (CELSI), Bratislava, Slovakia. `https://cran.r-project.org/web/packages/stargazer/stargazer.pdf`.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, 74(5):1368–1378.

Hoole, P., Nguyen-Trong, N., and Hardcastle, W. (1993). A comparative investigation of coarticulation in fricatives: Electropalatographic, electromagnetic, and acoustic data. *Language and Speech*, 36(2-3):235–260.

Hornberger, N. H. (2009). Multilingual education policy and practice: Ten certainties (grounded in Indigenous experience). *Language Teaching*, 42(2):197–211.

Hornberger, N. H. and King, K. A. (1996). Language revitalisation in the Andes: Can the schools reverse language shift? *Journal of Multilingual and Multicultural Development*, 17(6):427–441.

Huttenlocher, J., Vasilyeva, M., Cymerman, E., and Levine, S. (2002). Language input and child syntax. *Cognitive Psychology*, 45:337–374.

Icht, M. and Mama, Y. (2015). The production effect in memory: A prominent mnemonic in children. *Journal of Child Language*, 42(5):1102–1124.

Imbrie, A. (2005). *Acoustical Study of the Development of Stop Consonants in Children*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Inkelas, S. and Rose, Y. (2007). Positional neutralization: A case study from child language. *Language*, 83(4):707–736.

Iskarous, K., Mooshammer, C., Hoole, P., Recasens, D., Shadle, C. H., Saltzman, E., and Whalen, D. H. (2013). The coarticulation/invariance scale: Mutual information as a measure of coarticulation resistance, motor synergy, and articulatory invariance. *The Journal of the Acoustical Society of America*, 134(2):1271–1282.

Jarosz, G., Calamaro, S., and Zentz, J. (2017). Input frequency and the acquisition of syllable structure in Polish. *Language Acquisition*, 24(4):361–399.

Johnson, K. (1988). *Processes of Speaker Normalizaton in Vowel Perception*. Unpublished doctoral dissertation, Ohio State University, Columbus, OH.

Johnson, K. (2004). Massive reduction in conversational American English. In Yoneyama, K. and Maekawa, K., editors, *Proceedings of the 1st Session of the 10th International Symposium*, pages 29–54, The National International Institute for Japanese Language.

Johnson, K. and Sjerps, M. (2018). Speaker normalization in speech perception. *UC Berkeley Phonetics and Phonology Lab Annual Report*, pages 32–64.

Johnson, K., Strand, E. A., and D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics*, 27(4):359–384.

Katz, W. F., Kripke, C., and Tallal, P. (1991). Anticipatory coarticulation in the speech of adults and young children: Acoustic, perceptual, and video data. *Journal of Speech Language and Hearing Research*, 34:1222–1232.

Kemps, R. J. J. K., Wurm, L. H., Ernestus, M., Schreuder, R., and Baayen, H. R. (2005). Prosodic cues for morphological complexity in Dutch and English. *Language and Cognitive Processes*, 20(1-2):43–73.

Kent, R. D. (1983). The segmental organization of speech. In MacNeilage, P. F., editor, *The Production of Speech*, pages 57–89. Springer New York, New York, NY.

Klein, R., Lasky, R. E., Yarbrough, C., Habicht, J., and Sellers, M. J. (1977). Relationship of infant/caretaker interaction, social class and nutritional status to developmental test performance among Guatemalan infants. In Leiderman, P., editor, *Culture and Infancy: Variations in the Human Experience*, pages 385–403. Academic Press, Cambridge, MA.

Kleinschmidt, D. F. (2017). *Perception in a Variable but Structured World: The Case of Speech Perception*. Unpublished doctoral dissertation, University of Rochester, Rochester, NY.

Koenig, J.-P. and Jurafsky, D. (1995). Type underspecification and on-line type construction in the lexicon. In Aranovich, R., Byrne, W., Preuss, S., and Senturia, M., editors, *Proceedings of the 13th West Coast Conference on Formal Linguistics*, pages 270–285, Stanford, CA. CSLI Publications.

Konner, M. (1977). Infancy among the Kalahari desert San. In Leiderman, P., editor, *Culture and Infancy: Variations in the Human Experience*, pages 287–328. Academic Press, Cambridge, MA.

Koopmans-Van Beinum, F. J. (1980). *Vowel Contrast Reduction, an Acoustic and Perceptual Study of Dutch Vowels in Various Speech Conditions*. PhD thesis, University of Amsterdam, Amsterdam, Netherlands.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50(2):93–107.

Kuhl, P. K., Andruski, J. E., Chistovich, I., Chistovich, L., Kozhevnikova, E., Ryskina, V. L., Stolyarova, E., Sundberg, U., and Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277(5326):684–686.

Kuznetsova, A., Brockhoff, P., and Christensen, R. (2017). lmerTest Package: Tests in linear mixed-effects models. *Journal of Statistical Software*, 82(13):1–26.

Lahiri, A. and Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38:245–294.

Lammert, A. C. and Narayanan, S. S. (2015). On short-time estimation of vocal tract length from formant frequencies. *PLoS ONE*, 10(7):e0132193.

Lee, S., Potamianos, A., and Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, 105(3):1455–1468.

Lee-Kim, S.-I., Davidson, L., and Hwang, S. (2013). Morphological effects on the darkness of English intervocalic /l/. *Laboratory Phonology*, 4(2):475–511.

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *The Journal of the Acoustical Society of America*, 51(6B):2018–2024.

Liberman, A. M., Cooper, F. S., Shankweiler, D., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6):431–461.

Liberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1):1–36.

Lidz, J. and Gagliardi, A. (2015). How nature meets nurture: Universal Grammar and statistical learning. *Annual Review of Linguistics*, 1(1):333–353.

Lieberman, P. and Crelin, E. S. (1972). On the speech of neanderthal man. *Linguistic Inquiry*, 2(2):203–222.

Lieven, E. V. M. (1994). Crosslinguistic and crosscultural aspects of language addressed to children. In Gallaway, C. and Richards, B. J., editors, *Input and Interation in Language Acquisition*, pages 56–73. Cambridge University Press, Cambridge.

Lieven, E. V. M., Pine, J. M., and Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, 24(1):187–219.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H Theory. In Hardcastle, W. J. and Marchal, A., editors, *Speech Production and Speech Modelling*, pages 403–439. Springer Netherlands, Dordrecht.

Liu, H.-M., Tsao, F.-M., and Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology*, 43(4):912–917.

Lo, S. and Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, 6(1171):1–16.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, 49(2B):606–608.

Lorge, I. and Chall, J. (1963). Estimating the size of vocabularies of children and adults: An analysis of methodological issues. *The Journal of Experimental Education*, 32(2):147–157.

Losiewicz, B. L. (1995). Word frequency effects on the acoustic duration of morphemes. *Journal of the Acoustical Society of America*, 97(5):3243–3243.

Lupyan, G. and Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE*, 5:e8559.

Mahr, T. and Edwards, J. (2017). Lexical processing of children with cochlear implants. In *42nd Annual Boston University Conference on Language Development*, Boston, MA.

Mahr, T. and Edwards, J. (2018). Using language input and lexical processing to predict vocabulary size. *Developmental Science*, 21(6):1–14.

Mann, V. and Repp, B. (1980). Influence of vocalic context on perception of the [sh]-[s] distinction. *Perception & Psychophysics*, 28:213–228.

Manuel, S. and Krakow, R. (1984). Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Report on Speech Research*, SR-77/78:69–78.

Martin, A., Peperkamp, S., and Dupoux, E. (2013). Learning phonemes with a proto-lexicon. *Cognitive Science*, 37(1):103–124.

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., and Cristia, A. (2015). Mothers Speak Less Clearly to Infants Than to Adults: A Comprehensive Test of the Hyperarticulation Hypothesis. *Psychological Science*, 26(3):341–347.

Martinet, A. (1952). Function, structure, and sound change. *Word*, 8(1):1–32.

Martland, P., Whiteside, S., Beet, S., and Baghai-Ravary, L. (1996). Estimating child and adolescent formant frequency values from adult data. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, volume 2, pages 626–629, Philadelphia, PA, USA. IEEE.

Mastin, J. D. and Vogt, P. (2016). Infant engagement and early vocabulary development: A naturalistic observation study of Mozambican infants from 1;1 to 2;1. *Journal of Child Language*, 43(02):235–264.

Matthies, M., Perrier, P., Perkell, J., and Zandipour, M. (2001). Variation in anticipatory coarticulation with changes in clarity and rate. *Journal of Speech Language and Hearing Research*, 44(2):340–353.

Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3):B101–B111.

McAllister Byun, T. (2012). Positional velar fronting: An updated articulatory account. *Journal of Child Language*, 39(05):1043–1076.

McAllister Byun, T., Harel, D., Halpin, P. F., and Szeredi, D. (2016a). Deriving gradient measures of child speech from crowdsourced ratings. *Journal of Communication Disorders*, 64:91–102.

McAllister Byun, T., Inkelas, S., and Rose, Y. (2016b). The A-map model: Articulatory reliability in child-specific phonology. *Language*, 92(1):141–178.

McAllister Byun, T. and Tessier, A.-M. (2016). Motor influences on grammar in an emergentist model of phonology: Motor and grammar in emergentist phonology. *Language and Linguistics Compass*, 10(9):431–452.

McAuliffe, M., Socolof, M., Mihuc, A., Wagner, M., and Sonderegger, M. (2017). Montreal Forced Aligner. `http://montrealcorpustools.github.io/Montreal-Forced-Aligner/`. Version 1.0.0.

McBride-Chang, C., Tardif, T., Cho, J.-R., Shu, H., Fletcher, P., Stokes, S. F., Wong, A., and Leung, K. (2008). What's in a word? Morphological awareness and vocabulary knowledge in three languages. *Applied Psycholinguistics*, 29(3):437–462.

McCune, L. and Vihman, M. M. (2001). Early phonetic and lexical development: A productivity approach. *Journal of Speech Language and Hearing Research*, 44:670–684.

McFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., and Nieto, O. (2015). Librosa: Audio and music signal analysis in Python. In *Proceedings of the 14th Python in Science Conference*, pages 18–24, Austin, Texas.

McGowan, R. W., McGowan, R. S., Denny, M., and Nittrouer, S. (2014). A longitudinal study of very young children's vowel production. *Journal of Speech, Language, and Hearing Research*, 57(1):1–15.

McMurray, B., Danelz, A., Rigler, H., and Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Developmental Psychology*, 54(8):1472–1491.

Mehl, M. R. (2017). The electronically activated recorder (EAR): A method for the naturalistic observation of daily social behavior. *Current Directions in Psychological Science*, 26(2):184–190.

Mehl, M. R. and Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology*, 84(4):857–870.

Ménard, L. and Boë, L.-J. (2000). Exploring vowel production strategies from infant to adult by means of articulatory inversion of formant data. In *The Sixth International Conference on Spoken Language Processing*, Beijing, China.

Ménard, L., Perrier, P., Aubin, J., Savariaux, C., and Thibeault, M. (2008). Compensation strategies for a lip-tube perturbation of French [u]: An acoustic and perceptual study of 4-year-old children. *The Journal of the Acoustical Society of America*, 124(2):1192–1206.

Ménard, L., Schwartz, J.-L., Boë, L.-J., and Aubin, J. (2007). Articulatory–acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model. *Journal of Phonetics*, 35(1):1–19.

Menn, L., Schmidt, E., and Nicholas, B. (2013). Challenges to theories, charges to a model: The Linked-Attractor model of phonological development. In Vihman, M. M. and Keren-Portnoy, T., editors, *The Emergence of Phonology: Whole-Word Approaches and Cross-Linguistic Evidence*, pages 460–502. Cambridge University Press.

Metsala, J. and Walley, A. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precurors to phonemic awareness and early reading ability. In Metsala, J. L. and Ehri, L. C., editors, *Word Recognition in Beginning Literacy*, pages 89–120. Lawrence Erlbaum Associates, Mahwah, NJ.

Meylan, S. C., Frank, M. C., Roy, B. C., and Levy, R. (2017). The emergence of an abstract grammatical category in children's early speech. *Psychological Science*, 28(2):181–192.

Mooshammer, C., Hoole, P., and Geumann, A. (2006). Interarticulator cohesion within coronal consonant production. *Journal of the Acoustical Society of America*, 120(2):1028–1039.

Morrison, C. M., Chappell, T. D., and Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology Section A*, 50(3):528–559.

Morrison, C. M., Ellis, A. W., and Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition*, 20(6):705–714.

Mousikou, P., Strycharczuk, P., Turk, A., Rastle, K., and Scobbie, J. M. (2015). Morphological effects on pronunciation. In *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, Scotland.

Munhall, K., Fowler, C., Hawkins, S., and Saltzman, E. (1992). "Compensatory shortening" in monosyllables of spoken English. *Journal of Phonetics*, 20(2):225–239.

Munson, B., Edwards, J., and Beckman, M. E. (2005). Relationships Between Nonword Repetition Accuracy and Other Measures of Linguistic Development in Children With Phonological Disorders. *Journal of Speech, Language, and Hearing Research*, 48(1):61–78.

Muysken, P. (2012a). Contacts between indigenous languages in South America. In Campbell, L. and Grondona, V., editors, *The Indigenous Languages of South America: A Comprehensive Guide*, pages 235–258. Walter de Gruyter, Berlin, Germany.

Muysken, P. (2012b). Spanish affixes in the Quechua languages: A multidimensional perspective. *Lingua*, 122(5):481–493.

Nearey, T. (1977). *Phonetic Feature Systems for Vowels*. Unpublished doctoral dissertation, University of Alberta, Edmonton, CA.

Ngon, C., Martin, A., Dupoux, E., Cabrol, D., Dutat, M., and Peperkamp, S. (2013). (Non)words, (non)words, (non)words: Evidence for a protolexicon during the first year of life. *Developmental Science*, 16(1):24–34.

Nijland, L., Maassen, B., der Meulen, S. V., Gabreëls, F., Kraaimaat, F. W., and Schreuder, R. (2002). Coarticulation patterns in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 16(6):461–483.

Nittrouer, S. (1993). The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. *Journal of Speech Language and Hearing Research*, 36:959–972.

Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech Language and Hearing Research*, 32:120–132.

Nittrouer, S., Studdert-Kennedy, M., and Neely, S. T. (1996). How children learn to organize their speech gestures: Further evidence from fricative-vowel syllables. *Journal of Speech Language and Hearing Research*, 39:379–389.

Noiray, A., Abakarova, D., Rubertus, E., Krüger, S., and Tiede, M. (2018). How do children organize their speech in the first years of life? Insight from ultrasound imaging. *Journal of Speech, Language, and Hearing Research*, 61(6):1355–1368.

Noiray, A., Ménard, L., and Iskarous, K. (2013). The development of motor synergies in children: Ultrasound and acoustic measurements. *Journal of the Acoustical Society of America*, 133(1):444–452.

Noiray, A., Popescu, A., Killmer, H., Rubertus, E., Krüger, S., and Hintermeier, L. (2019a). Spoken language development and the challenge of skill integration. *Frontiers in Psychology*, 10(2777):1–17.

Noiray, A., Wieling, M., Abakarova, D., Rubertus, E., and Tiede, M. (2019b). Back from the future: Nonlinear anticipation in adults' and children's speech. *Journal of Speech, Language, and Hearing Research*, 62(8S):3033–3054.

Nordstrom, P. E. and Lindblom, B. (1975). A normalization procedure for vowel formant data. In *Proceedings of the 8th International Congress of Phonetic Sciences*, Leeds, England.

Ochs, E. (1988). *Culture and Language Development: Language Acquisition and Language Socialization in a Samoan Village*. Cambridge University Press, Cambridge, UK.

O'Donnell, T. J. (2015). *Productivity and Reuse In Language: A Theory of Linguistic Computation and Storage*. The MIT Press, Cambridge, MA.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America*, 39(1):151–168.

Orena, A. J., Byers-Heinlein, K., and Polka, L. (2019). Reliability of the Language Environment Analysis (LENA) in French-English bilingual speech. *Journal of Speech Language and Hearing Research*, 67(2):2491–2500.

Orena, A. J., Byers-Heinlein, K., and Polka, L. (2020). What do bilingual infants actually hear? Evaluating measures of language input to bilingual-learning 10-month-olds. *Developmental Science*, 23(2):1–14.

Pace, A., Luo, R., Hirsh-Pasek, K., and Golinkoff, R. M. (2017). Identifying pathways between socioeconomic status and language development. *Annual Review of Linguistics*, 3(1):285–308.

Pearson, B. Z., Fernandez, S. C., Lewedeg, V., and Oller, D. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, 18(1):41–58.

Perfetti, C. A., Beck, I., Bell, L. C., and Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly*, 33(3):283–319.

Perkell, J. S. and Klatt, D., editors (2014). *Invariance and Variability in Speech Processes*. Psychology Press, New York and London.

Peterson, G. E. and Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24(2):175–184.

Pettinato, M., Tuomainen, O., Granlund, S., and Hazan, V. (2016). Vowel space area in later childhood and adolescence: Effects of age, sex and ease of communication. *Journal of Phonetics*, 54:1–14.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46(2-3):115–154.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, 2:33–52.

Pinker, S. and Prince, A. (1994). Regular and irregular morphology and the psychological status of rules of grammar. In Lima, S. D., Corrigan, R. L., and Iverson, G. K., editors, *The Reality of Linguistic Rules*, pages 321–351. John Benjamins Publishing Company, Amsterdam/Philadelphia.

Pinker, S. and Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, 6(11):456–463.

Place, S. and Hoff, E. (2011). Properties of dual language exposure that influence 2-year-olds' bilingual proficiency: Dual language exposure and bilingual proficiency. *Child Development*, 82(6):1834–1849.

Place, S. and Hoff, E. (2016). Effects and noneffects of input in bilingual environments on dual language skills in 2 -year-olds. *Bilingualism: Language and Cognition*, 19(5):1023–1041.

Plag, I. (2014). Phonological and phonetic variability in complex words: An uncharted territory. *Rivista di Linguistica*, 26(2):209–228.

Plag, I. and Baayen, H. R. (2009). Suffix ordering and morphological processing. *Language*, 85(1):109–152.

Plag, I., Homann, J., and Kunter, G. (2017). Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*, 53(1):181–216.

Pluymaekers, M., Ernestus, M., Baayen, H. R., and Booij, G. (2010). Morphological effects on fine phonetic detail: The case of Dutch -igheid. In Fougeron, C., Kühnert, B., D'Imperio, M., and Vallée, N., editors, *Laboratory Phonology*, volume 10, pages 511–532. de Gruyter Mouton, Berlin.

Potter, C. E., Fourakis, E., Morin-Lessard, E., Byers-Heinlein, K., and Lew-Williams, C. (2019). Bilingual toddlers' comprehension of mixed sentences is asymmetrical across their two languages. *Developmental Science*, 22(4):e12794.

Ramírez-Esparza, N., García-Sierra, A., and Kuhl, P. K. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, 17(6):880–891.

Ramírez-Esparza, N., García-Sierra, A., and Kuhl, P. K. (2017a). The impact of early social interactions on later language development in Spanish-English bilingual infants. *Child Development*, 88(4):1216–1234.

Ramírez-Esparza, N., García-Sierra, A., and Kuhl, P. K. (2017b). Look who's talking NOW! Parentese speech, social context, and language development across time. *Frontiers in Psychology*, 8(1008):1–12.

Ratner, N. B. (1984). Patterns of vowel modification in mother–child speech. *Journal of Child Language*, 11(3):557–578.

Recasens, D. (1985). Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences. *Language and Speech*, 28(2):97–114.

Recasens, D. and Espinosa, A. (2006). Dispersion and variability of Catalan vowels. *Speech Communication*, 48(6):645–666.

Recasens, D. and Espinosa, A. (2009a). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *Journal of the Acoustical Society of America*, 125(4):2288–2298.

Recasens, D. and Espinosa, A. (2009b). Dispersion and variability in Catalan five and six peripheral vowel systems. *Speech Communication*, 51(3):240–258.

Recasens, D., Pallarès, M. D., and Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America*, 102(1):544–561.

Redford, M. A. (2018). Grammatical word production across metrical contexts in school-aged children's and adults' speech. *Journal of Speech, Language, and Hearing Research*, 61(6):1339–1354.

Redford, M. A. (2019). Speech production from a developmental perspective. *Journal of Speech, Language, and Hearing Research*, 62(8S):2946–2962.

Reidy, P. F., Kristensen, K., Winn, M. B., Litovsky, R. Y., and Edwards, J. R. (2017). The acoustics of word-initial fricatives and their effect on word-level intelligibility in children with bilateral cochlear implants:. *Ear and Hearing*, 38(1):42–56.

Repp, B. H. (1986). Some observations on the development of anticipatory coarticulation. *Journal of the Acoustical Society of America*, 79(5):1616–1619.

Roberts, I. and Roussou, A. (1999). A formal approach to "grammaticalization". *Linguistics*, 37(6):1011–1041.

Ronson, I. (1976). Word frequency and stuttering: The relationship to sentence structure. *Journal of Speech Language and Hearing Research*, 19(4):813–819.

Rose, Y. and Inkelas, S. (2011). The Interpretation of Phonological Patterns in First Language Acquisition: The Interpretation of Phonological Patterns in First Language Acquisition. In van Oostendorp, M., Ewen, C. J., Hume, E., and Rice, K., editors, *The Blackwell Companion to Phonology*, pages 1–25. John Wiley & Sons, Ltd, Oxford, UK.

RStudio Team (2020). RStudio: Integrated Development for R. https://rstudio.com/. Version 1.2.5033.

Rubertus, E., Abakarova, D., Ries, J., and Noiray, A. (2013). Anticipatory V-to-V coarticulation in German preschoolers. In Draxler, C. and Kleber, F., editors, *Tagungsband der 12. Tagung 9 Phonetik und Phonologie im deutschsprachigen Raum*, pages 12–14, Ludwig-Maximilians-Universitat Munchen.

Rubertus, E. and Noiray, A. (2018). On the development of gestural organization: A cross-sectional study of vowel-to-vowel anticipatory coarticulation. *PLoS ONE*, 13(9):1–21.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, 12(4):110–114.

Schenk, B. S., Baumgartner, W.-D., and Hamzavi, J. S. (2003). Changes in vowel quality after cochlear implantation. *ORL*, 65(3):184–188.

Schneider, B. A., Trehub, S. E., Morrongiello, B. A., and Thorpe, L. A. (1985). Auditory Sensitivity in Preschool Children. 79(2):447–452.

Schneider, W., Eschman, A., and Zuccolotto, A. (2012). E-Prime. Psychology Software Tools, Inc.

Seidl, A., Onishi, K. H., and Cristia, A. (2014). Talker variation aids young infants' phonotactic learning. *Language Learning and Development*, 10(4):297–307.

Sereno, J. A., Baum, S. R., Marean, G. C., and Lieberman, P. (1987). Acoustic analyses and perceptual data on anticipatory labial. *Journal of the Acoustical Society of America*, 81(2):512–519.

Sereno, J. A. and Lieberman, P. (1987). Developmental aspects of lingual coarticulation. *Journal of Phonetics*, 15:247–257.

Seyfarth, S., Garellek, M., Gillingham, G., Ackerman, F., and Malouf, R. (2018). Acoustic differences in morphologically-distinct homophones. *Language, Cognition and Neuroscience*, 33(1):32–49.

Shapiro, S. S., Wilk, M. B., and Chen, H. J. (1968). A comparative study of various tests for normality. *Journal of the American Statistical Association*, 63(324):1343–1372.

Shneidman, L. and Woodward, A. L. (2016). Are child-directed interactions the cradle of social learning? *Psychological Bulletin*, 142(1):1–17.

Shneidman, L. A., Arroyo, M. E., Levine, S. C., and Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language*, 40(3):672–686.

Shneidman, L. A. and Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: How important is directed speech? *Developmental Science*, 15(5):659–673.

Skinner, B. F. (1957). *Verbal Behavior*. Copley Publishing Group, Acton, MA.

Smith, A. and Goffman, L. (1998). Stability and Patterning of Speech Movement Sequences in Children and Adults. *Journal of Speech Language and Hearing Research*, 41(1):18–30.

Smith, A. and Zelaznik, H. N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology*, 45(1):22–33.

Smith, B. L. (1992). Relationships between duration and temporal variability in children's speech. *The Journal of the Acoustical Society of America*, 91(4):2165–2174.

Smith, L. B. and Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, 7(8):343–348.

Smith, R., Baker, R., and Hawkins, S. (2012). Phonetic detail that distinguishes prefixed from pseudo-prefixed words. *Journal of Phonetics*, 40(5):689–705.

Soli, S. D. (1981). Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *The Journal of the Acoustical Society of America*, 70(4):976–984.

Song, J. Y., Demuth, K., Evans, K., and Shattuck-Hufnagel, S. (2013a). Durational cues to fricative codas in 2-year-olds' American English: Voicing and morphemic factors. *Journal of the Acoustical Society of America*, 133(5):2931–2946.

Song, J. Y., Demuth, K., Shattuck-Hufnagel, S., and Ménard, L. (2013b). The effects of coarticulation and morphological complexity on the production of English coda clusters: Acoustic and articulatory evidence from 2-year-olds and adults using ultrasound. *Journal of Phonetics*, 41:281–295.

Sosa, A. V. and Stoel-Gammon, C. (2012). Lexical and Phonological Effects in Early Word Production. *Journal of Speech Language and Hearing Research*, 55(2):596.

Stanovich, K. E., Nathan, R. G., and Vala-Rossi, M. (1986). Developmental changes in the cognitive correlates of reading ability and the developmental lag hypothesis. *Reading Research Quarterly*, 21(3):267–283.

Stokes, S. F. and Surendran, D. (2005). Articulatory complexity, ambient frequency, and functional load as predictors of consonant development in children. *Journal of Speech Language and Hearing Research*, 48(3):577–591.

Storkel, H. L. (2002). Restructuring of similarity neighbourhoods in the developing mental lexicon. *Journal of Child Language*, 29(2):251–274.

Storkel, H. L. (2004). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics*, 25:201–221.

Story, B. H., Vorperian, H. K., Bunton, K., and Durtschi, R. B. (2018). An age-dependent vocal tract model for males and females based on anatomic measurements. *The Journal of the Acoustical Society of America*, 143(5):3079–3102.

Strycharczuk, P. (2019). Phonetic detail and phonetic gradience in morphological processes. In *Oxford Research Encyclopedia of Linguistics*, pages 1–25. Oxford University Press, Oxford, UK.

Strycharczuk, P. and Scobbie, J. M. (2016). Gradual or abrupt? The phonetic path to morphologisation. *Journal of Phonetics*, 59:76–91.

Sugahara, M. and Turk, A. (2009). Durational correlates of English sublexical constituent structure. *Phonology*, 26(03):477–524.

Sussman, H. M., Minifie, F. D., Buder, E. H., Stoel-Gammon, C., and Smith, J. (1996). Consonant-vowel interdependencies in babbling and early words: Preliminary examination of a locus equation approach. *Journal of Speech Language and Hearing Research*, 39:424–433.

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1536):3617–3632.

Swingley, D. and Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13(5):480–484.

Swingley, D. and Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology*, 54(2):99–132.

Taft, M. and Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior*, 15(6):607–620.

The Many Babies Consortium (2020). Quantifying sources of variability in infancy research using the infant-directed speech preference. *Advances in Methods and Practices in Psychological Science*, 3(1):24–52.

Thelen, E. and Smith, L. B. (1996). *A Dynamic Systems Approach to the Development of Cognition and Action*. MIT Press, Cambridge, MA.

Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, 15(4):426–445.

Tilsen, S. (2016). Selection and coordination: The articulatory basis for the emergence of phonological structure. *Journal of Phonetics*, 55:53–77.

Tingley, B. and Allen, G. (1975). Development of speech timing control in children. *Child Development*, 46(1):186–194.

Tomaschek, F., Tucker, B. V., Ramscar, M., and Baayen, H. R. (under review). How is anticipatory coarticulation of suffixes affected by lexical proficiency? Retrieved from: `https://psyarxiv.com/gv89j/`.

Torero, O. (1964). Los dialectos Quechuas. In *Anales Científicos de La Universidad Agraria*, volume 2, pages 446–478. Lima, Peru.

Tsao, Y.-C., Weismer, G., and Iqbal, K. (2006). The effect of intertalker speech rate variation on acoustic vowel space. *Journal of the Acoustical Society of America*, 119(2):1074–1082.

Tsuji, S., Fikkert, P., Minagawa, Y., Dupoux, E., Filippin, L., Versteegh, M., Hagoort, P., and Cristia, A. (2017). The more, the better? Behavioral and neural correlates of frequent and infrequent vowel exposure. *Developmental Psychobiology*, 59(5):603–612.

Turner, R. E., Walters, T. C., Monaghan, J. J. M., and Patterson, R. D. (2009). A statistical, formant-pattern model for segregating vowel type and vocal-tract length in developmental formant data. *The Journal of the Acoustical Society of America*, 125(4):2374–2386.

Umesh, S., Kumar, S. V. B., Vinay, M. K., Sharma, R., and Sinha, R. (2002). A simple approach to non-uniform vowel normalization. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages I–517–I–520, Orlando, USA.

Usoltsev, A. (2015). Voice Activity Detector-Python. Retrieved from: `https://github.com/marsbroshok/VAD-python`.

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

van Gelderen, E. (2004). *Grammaticalization as Economy*. John Benjamins, Amsterdam.

Vihman, M. and Croft, W. (2007). Phonological development: Toward a "radical" templatic phonology. *Linguistics*, 45(4):683–725.

Vihman, M. M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology*, 108(1):1–27.

Vihman, M. M. and Keren-Portnoy, T., editors (2013). *The Emergence of Phonology: Whole-Word Approaches and Cross-Linguistic Evidence*. Cambridge University Press, Cambridge, MA.

Vogt, P., Mastin, J. D., and Schots, D. M. A. (2015). Communicative intentions of child-directed speech in three different learning environments: Observations from the Netherlands, and rural and urban Mozambique. *First Language*, 35(4-5):341–358.

Vorperian, H. K. and 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.

Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., and Yandell, B. S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *The Journal of the Acoustical Society of America*, 117(1):338–350.

Vorperian, H. K. and Wang, S. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study). *Journal of the Acoustical Society of America*, 125(3):13.

Vorperian, H. K., Wang, S., Schimek, E. M., Durtschi, R. B., Kent, R. D., Gentry, L. R., and Chung, M. K. (2011). Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of Speech, Language, and Hearing Research*, 54(4):995–1010.

Wagner, R. K., Torgesen, J. K., and Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology*, 30(1):73–87.

Warlaumont, A. S. (2015). Modeling the emergence of syllabic structure. *Journal of Phonetics*, 53:61–65.

Warner, N., Good, E., Jongman, A., and Sereno, J. (2006). Orthographic vs. morphological incomplete neutralization effects. *Journal of Phonetics*, 34(2):285–293.

Watanabe, A. (2001). Formant estimation method using inverse-filter control. *IEEE Transactions on Speech and Audio Processing*, 9(4):317–326.

Wedel, A., Kaplan, A., and Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, 128(2):179–186.

Weisleder, A. and Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11):2143–2152.

Werker, J. F. and Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, 1(2):197–234.

Werker, J. F. and Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7(1):49–63.

Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua*, 106:23–79.

Whalen, D. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]-[S] boundary. *The Journal of the Acoustical Society of America*, 69(1):275–282.

Whalen, D. H. (1990). Coarticulation is largely planned. *Journal of Phonetics*, 18:3–35.

Whiteside, S. and Hodgson, C. (2000). Speech patterns of children and adults elicited via a picture-naming task: An acoustic study. *Speech Communication*, 32(4):267–285.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. `https://ggplot2.tidyverse.org`.

Yang, C., Crain, S., Berwick, R. C., Chomsky, N., and Bolhuis, J. J. (2017). The growth of language: Universal Grammar, experience, and principles of computation. *Neuroscience & Biobehavioral Reviews*, 81:103–119.

Yang, C. D. (2004). Universal Grammar, statistics or both? *Trends in Cognitive Sciences*, 8(10):451–456.

Zamuner, T. S., Gerken, L., and Hammond, M. (2005). The acquisition of phonology based on input: A closer look at the relation of cross-linguistic and child language data. *Lingua*, 115(10):1403–1426.

Zamuner, T. S., Strahm, S., Morin-Lessard, E., and Page, M. P. (2018). Reverse production effect: Children recognize novel words better when they are heard rather than produced. *Developmental Science*, 21(4):e12636.

Zharkova, N., Hardcastle, W. J., and Gibbon, F. E. (2018). The dynamics of voiceless sibilant fricative production in children between 7 and 13 years old: An ultrasound and acoustic study. *The Journal of the Acoustical Society of America*, 144(3):1454–1466.

Zharkova, N., Hewlett, N., and Hardcastle, W. J. (2011). Coarticulation as an indicator of speech motor control development in children: An ultrasound study. *Motor Control*, 15(1):118–140.

Zharkova, N., Hewlett, N., Hardcastle, W. J., and Lickley, R. J. (2014). Spatial and temporal lingual coarticulation and motor control in preadolescents. *Journal of Speech Language and Hearing Research*, 57(2):374–388.

Zingler, T. (2018). Reduction without fusion: Grammaticalization and wordhood in Turkish. *Folia Linguistica*, 52(2):415–447.