THE UNIVERSITY OF CHICAGO

TOWARDS AN ARTICULATORY MODEL OF HANDSHAPE: WHAT FINGERSPELLING TELLS US ABOUT THE PHONETICS AND PHONOLOGY OF HANDSHAPE IN AMERICAN SIGN LANGUAGE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE HUMANITIES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF LINGUISTICS

BY
JONATHAN KEANE

CHICAGO, ILLINOIS
DECEMBER 2014

for my loved one(s)

Table of Contents

List of Figures

List of Tables

Acknowledgments

I was extremely fortunate to have a number of people influence my development throughout my education. First, my committee individually, and together, shaped me in innumerable ways. Diane has such a breadth and depth of knowledge and experience with linguistics, especially the sign language world. She was the first person to formally expose me to sign language linguistics. And she pushed me in ways I did not think possible. I will never forget during one of our first one-on-one meetings, telling me that if I wanted to do this work I needed to be fluent and able to sign yesterday. Jason was there to lead me to the right solution for stumbling blocks, be they technical, personal, or political. The way that he never told me I was wrong, but made sure that I was in a situation to figure that out myself made me a stronger researcher and person. Karen always knew exactly how to ask the deep question that got me thinking about the core idea that was I presenting in a new way that I had not before. Finally, although not on my committee, and not working on related work, Karlos Arregi was always available to chat about linguistics, life, and has an infectious inquisitiveness that always left me invigorated after meetings.

My linguistics eduction did not start, or stay only at the University of Chicago. I have to thank my three undergraduate advisors, D. Gary Miller, Hana Filip, and Eric Potsdam. Each contributed to my linguistics upbringing in their own unique ways. I must thank Louis Goldstein for always informative and thought-provoking discussion about Articulatory Phonology. Martha Tyrone is a wonderful mentor, and opened my world to Articulatory Phonology and sign languages among many other things. More than that, she has always been a warm presence wherever we met. David Quinto-Pozos is another mentor that has a knack for asking questions that make me think about things in ways I had not thought of. Although, not frequent enough, it is always great when we run into each other. Robbin Battison, on top of his work, opened his home to me in Stockholm replete with wonderful Swedish food (Thank you, Christina!) and stimulating conversation.

This work simply could not have happened without the signers who graciously fingerspelled hundreds of words for this corpus: Drucilla Ronchen, Andy Gabels, Rita Mawr, and Robin Shay, I could not have completed this project without you. Drucie, especially, deserves thanks for being a wonderful teacher and opening up my world to ASL.

Additionally, the corpus would not have been useful without the annotation and programming work of everyone involved with the project: Susan Post Rizzo, Erin Dahlgren, Alice Fine, Julia Goldsmith-Pinkham, TJ Heins, Katie Henry, Rachel Hwang, Linda Liu, Rachel Perry, and Katina Vradelis.

I was fortunate to have many stimulating conversations with many people at the University of Chicago, but should mention Carissa Abrego-Collier, Rebekah Baglini, Emily Coppess, Josh Falk, Gallagher Flinn, Tim Grinsell, Peet Klecha, Martina Matrinović, Mike Pham, Adam Singerman, and Julia Thomas.

I would like to especially thank Katie Franich for always being there to listen to my crazy theories, and reassure me when I was down. I learned so much from Morgan Sonderegger. Not only is he a great colleague and teacher, but he is always there to tell me that I am wrong when I am doing something wrong and set me on the right track.

I was supported by and have learned from colleagues outside of the university as well. Leah Geer is a wonderful friend, colleague, and collaborator, who never gets tired of my messages that start with "QUESTION!". Thanks to Lynn Hou for many great conversations while walking around Austin, traveling in England, or making ravioli. Christopher Green showed me how to be a good linguist while still maintaining a balanced life. I learned so much from others, in long stretching conversations at conferences among other places, and I would not be where I am today without the insights of Cecily Whitworth, Julie Hochgesang, Jordan Fenlon, and Aaron Shield.

This work was also supported in part by a Doctoral Dissertation Research Improvement Grant: NSF BCS 1251807.

Many thanks to innumerable coffee shops and cafes throughout Chicago that provided caffeine and the ambient noise sometimes needed to get work done. Above all, I must thank Eddie and Rich at Chava cafe. Their wonderful coffee, sandwiches, and atmosphere sustained me through much of my dissertating and grad school in general.

Finally, there are a number of people in my life that kept me sane, happy, and engaged. First, my mother Ellen has been an inspiration throughout my life. She has never stopped supporting me, no matter how strange what I was doing seemed to be. Yaron McNabb grounded me throughout grad school. He became a best friend within minutes of us meeting when I visited UofC. I will never forget the long discussions about our respective research, the many dinners we had together and with others, and most importantly the strange encounters we had together in Chicago, and now throughout the world. Thanks to Molly Flaherty and Lauren Applebaum for expanding my horizons, and always being there to talk to when I needed it most. Jonathan Goya was a wonderful partner through some of the more stressful experiences of grad school. He challenged me to be a better researcher and person in subtle and not so subtle ways. Ted and Keith, you are both wonderful in different ways, and I would not have been able to finish without you two.

Abstract

This dissertation is, at its core, an exploration of the phonetics-phonology interface, through the lens of handshape in American Sign Language (ASL). This exploration is split into three areas: 1. the development and implementation of a theory of the phonetics-phonology interface for handshape, 2. a quantitative analysis of the temporal properties of ASL fingerspelling, 3. a quantitative analysis of pinky extension coarticulation.

Although the phonology of sign languages in general — and handshape specifically — has seen quite a bit of study, the phonetics-phonology interface has not been explored as much. Chapter 2 proposes a model of the phonetics-phonology interface called the Articulatory Model of Handshape. This model builds on both the articulatory phonology and sign language phonology literatures, extending it to cover handshape in sign languages. This model proposes a maximal set of possible phonological contrasts of handshapes, as well as a concrete method of turning phonological specifications into phonetic targets. This model has not only been proposed and characterized, but it has also been implemented computationally. Part of this implementation is a mapping from phonological features to phonetic targets. This implementation is important because it allows for precise understanding of how choices of phonological specification, as well as the mapping from phonological features to phonetic targets, affect the system of handshapes that have been proposed. Additionally, this implementation includes a method to synthesize 3D renderings of handshapes from either phonological or phonetic specifications. This is an important first step in a number of directions: 1. it will allow for models of coarticulation to be visually approximated and tested, 2. it will allow for further investigation into phonological specification and its phonetic consequences, 3. it is a step forward in the field of automatic sign synthesis.

The temporal properties of fingerspelling have seen quite a bit of study, although most studies have been quite limited in the number of tokens that they have analyzed. Chapter 3 explores a large corpus of fingerspelling from ASL, and analyzes the temporal properties of this corpus. Additionally, motion capture data from a larger number of signers was collected and analyzed. This

alternative methodology required the development and testing of automatic data classification and other techniques for analysis. The motion capture data produced similar results on an analysis of fingerspelled word duration and fingerspelling rate. There are a large number of factors that contribute to the temporal properties of a given fingerspelled word (as well as the letters that make up that word). A number of variables showed a large amount of variation (especially among signers), which could be one source of the wide range of rates that have been reported in the literature. This work is a critical step in understanding core phonetic properties of ASL fingerspelling.

These temporal properties are then used as predictors in an analysis of one aspect of handshape coarticulation in fingerspelling. Context-dependent phonetic variation (especially coarticulation) is seen broadly across segments in spoken languages. Chapter 4 concentrates first on three case studies that exhibit handshape variation and then looks at detailed quantification of pinky extension in a large corpus of fingerspelling data. Both the case studies and the deeper analysis support the hypotheses that are predicted given articulatory phonology models of the phonetic implementation of handshape in fingerspelling. The analysis of pinky extension here shows that there is clear contextually based variation: When a segment is close to another segment that has an extended pinky finger it is more likely to also have an extended pinky finger, even if it does not canonically have pinky extension. This pattern is mediated by a number of factors including the speed of fingerspelling, as well as certain phonological features of the segment of interest. This pattern is predicted by the articulatory model of handshape (discussed in detail in chapter 4). This provides evidence for an Articulatory Phonology account of coarticulation the relies on active and inactive articulators. Although it is commonly accepted that the inactive articulators are more susceptible to coarticulation, testing this directly in spoken languages can be difficult. Handshape in ASL provides a perfect test case for this because the articulators, when they are inactive, can take on configurations that are (on the surface) identical to configurations that they can take on when they are active. Because of these we can see a clear distinction: when a pinky is flexed and active it is much less likely to be extended when it is surrounded by an extended pinky; however, when a pinky is flexed and inactive, it is quite susceptible to coarticulatory extension.

This work contributes to articulatory phonology specifically, as well as theories of speech production broadly by studying the distinction between active and nonactive articulator gestures. Handshape in sign languages is especially well suited to study this phenomenon because there are many possible combinations of active and nonactive articulators (all five digits). Additionally, unlike most articulators for spoken languages, the articulators can be seen and tracked easily without the occlusion of the cheeks and neck.

Chapter 1

Introduction

At first glance, fingerspelling as a system seems easy to describe: there is a limited number of units (26), and these are just strung together sequentially, one unit after another. However, as with all language phenomena, actual productions of fingerspelling are not just a small number of discrete units strung together completely independent of each other; rather, the units will frequently influence the precise timing and configurations of each other in systematic ways. This dissertation develops a connection between the limited units (the phonology) and the variable production (the phonetics) that is actually articulated. In order to do this, we have developed and computationally implemented a model of the phonetics-phonology interface for handshape in signed languages. This model is then used to explore and analyze two phenomena of fingerspelling: the temporal properties, and handshape variation that is driven by coarticulation. This model, as well as the rest of this dissertation, explores the phonetics-phonology interface specifically. There has been much work on the phonology as well as the phonology-morphology interface in sign languages; however, the phonetics-phonology interface in sign languages has seen relatively little work until recently.

1.1 Fingerspelling

American Sign Language — ASL — is used by approximately 500,000 to 2 million people in the USA and Canada[1], the majority of whom are deaf. As with other sign languages, ASL makes use of the hands, arms, face, and body for communication.

Fingerspelling, while not the main method of communication, is an important part of ASL — used anywhere from 12 to 35 percent of the time in ASL discourse (Padden & Gunsauls, 2003). Fingerspelling is used more frequently in ASL than in other sign languages (Padden, 1991). Fingerspelling is a loanword system that has a form derived from the representation of English words through a

1. As was documented by Mitchell *et al.* (2006), these numbers range widely across sources.

series of handshapes[2], each of which maps to a letter in the word. Every letter used in English has a unique combination of handshape, orientation, and in a few cases movement path[3] (Cormier *et al.* (2008) among others). These are used sequentially to represent an English word. Figure 1.1 shows the handshapes for ASL fingerspelling. The orientation of each handshape is altered in this figure for ease of second language learning. In reality, all letters are articulated with the palm facing forward, away form the signer, except for -H-, -G- (in, towards the signer), -P-, -Q- (down) and the end of -J- (to the side).[4]

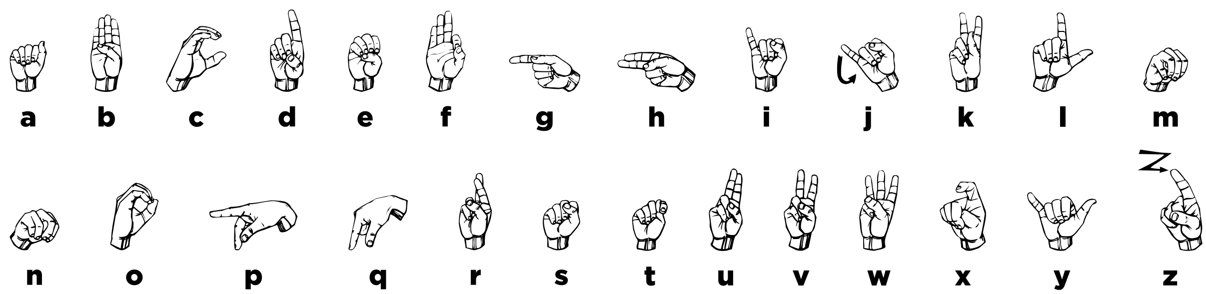


Figure 1.1: FS-letters for ASL fingerspelling

Throughout this dissertation, there is a clear focus on handshape. This is not to say that orientation is not important for fingerspelling (in fact the pairs -H- and -U- as well as -K- and -P- differ only in orientation), and orientation will be discussed in some parts. Rather, we concentrate on handshape because the coarticulatory process analyzed in chapter 4 (pinky extension) is specific to handshape alone; additionally, most letters are differentiated by handshape alone. The use of handshape (and orientation) to mark contrastive items is similar to the use that handshape has in core lexical items in other parts of the ASL lexicon, although in the core ASL lexicon there are

2. *Handshapes* is not quite the right word here, as will be explained in detail in the discussion of terminology in section 1.2

3. Traditionally movement is said to only be used for the letters -J- and -Z- as well as to indicate some instances of letter doubling.

4. This figure was generated using a freely available font created by David Rakowski. This figure is licensed under a Creative Commons Attribution-ShareAlike 3.0 Unported License and as such can be reproduced freely, so long as it is attributed appropriately. Contact jonkeane@uchicago.edu for an original file.

other, additional parameters that generate contrast: location, movement, and non-manual markers in addition to handshape and orientation. However, a sign segment will include a stable handshape (or two, if there is a handshape change in the sign), in the same way that is expected of segments in fingerspelling. Therefore, although the findings here are for fingerspelling specifically, we expect that they will for the most part generalize to the rest of ASL.

Fingerspelling is not used equally across all word categories. Fingerspelling is generally restricted to names, nouns, and to a smaller extent adjectives. These three categories make up about 77 percent of fingerspelled forms in data analyzed by Padden & Gunsauls (2003). This study analyzed the signing of 14 native signers for one study, and 36 native signers for another study. Both of these were a subset of signers from a larger sociolinguistics database that was compiled by Ceil Lucas, Robert Bayley, Clayton Valli, and their associates. In early research many situated fingerspelling as a mechanism to fill in vocabulary items that are missing in ASL. On further investigation, it has been discovered that this is not the whole story (Padden & Le Master, 1985). Fingerspelling can be used for emphasis as well as when the ASL sign for a concept is at odds with the closest English word, mainly in bilingual settings. One often cited example of the first is the use of Y-E-S-Y-E-S[5] and G-E-T-O-U-T. An example of the second is a teacher fingerspelling P-R-O-B-L-E-M as in a scientific problem in a science class, to clarify that what was intended here was not an interpersonal problem, but rather the setup for a scientific hypothesis. While fingerspelling is an integral part of ASL for all speakers of ASL, it is used more frequently by more educated signers, as well as more frequently by native signers when compared with non-native signers (Padden & Gunsauls, 2003).

Finally, there is already some literature on the nativization process from fingerspelled form to lexicalized sign (Brentari & Padden, 2001; Cormier *et al.*, 2008). The phonetics and phonology of fingerspelling are in many ways related to ASL in general, because it uses many of the same articulators, but there are important differences. One major difference is that because fingerspelling is

5. I am choosing to adopt the typographic conventions of Brentari & Padden (2001). Fingerspelled forms are written in small caps (an adaptation from Cormier *et al.* (2008)), with hyphens: A-T-L-A-N-T-I-C and ASL native signs are written in only small caps: GROUP. Single fingerspelled letters will be flanked by hyphens on either side (e.g. -T-).

comprised of rapid sequences of handshapes, it provides an excellent area to look at the effects of coarticulation on handshape. Thus it is important that we study the phonetics and phonology of fingerspelling as well as of ASL generally. With the exception of (Wilcox, 1992; Tyrone *et al.*, 1999; Emmorey *et al.*, 2010; Emmorey & Petrich, 2011; Quinto-Pozos, 2010) there is little literature on the phonetics of fingerspelling. Wilcox (1992) looks at a very small subset of words (~7) and attempts to describe the dynamics of movement in fingerspelling. Tyrone *et al.* (1999) looks at fingerspelling in Parkinsonian signers, and what phonetic features are compromised in Parkinsonian fingerspelling. Emmorey *et al.* (2010); Emmorey & Petrich (2011) studied the effects of segmentation on the perception of fingerspelling and compared it to parsing printed text. Finally Quinto-Pozos (2010) looks at the rate of fingerspelling in fluent discourse in a variety of social settings.

There has been a small amount of work on coarticulation in fingerspelling specifically. Jerde *et al.* (2003) mentions that there is coarticulation with respect to the pinky.[6] Tyrone *et al.* (1999) describes some Parkinsonian signers who blend letters together and gives an example of the first two FS-letters of P-I-L-L-S being blended together. Finally, Hoopes (1998) notes the existence of pinky extension coarticulation in fingerspelling but separates it from the pinky extension that he is interested in: the use of pinky extension in core lexical items as a sociolinguistic marker.

1.2 Terminology

The terminology used to describe fingerspelling is, for the most part, uncontroversial. Most people use the terms *word* and *letter* as they are applied to English (and other languages) and their orthographic representation. These are both fine for almost all descriptions of fingerspelling, although in the course of this work we have found the need to expand on these terms to ensure that we are making distinctions where they need to be made for theoretical and practical reasons. Figure 1.2 shows the general schema that follows: The term *FS-letter* (short for fingerspelled-letter) is one of

6. This study was of ASL interpreters. There was no discussion about the language backgrounds of these interpreters, so it is unclear if they are native signers, early learners, late second language learners of ASL; or if different language backgrounds might influence the results.

the 26 unique combinations of handshape and orientation that map onto the Latin alphabet used for English. This is what is traditionally described as the phonological level, which is abstract. Moving on to phonetic instantiations, an *apogee*[7] refers to a specific instance of an FS-letter in the data. The term *handshape* refers to the canonical (or phonological) configuration of the hand for each FS-letter. The term *hand configuration* (following others, including Whitworth (2011)) refers to the actual (phonetic) realization of handshape, which combined with (phonetic) orientation, forms an *apogee*).

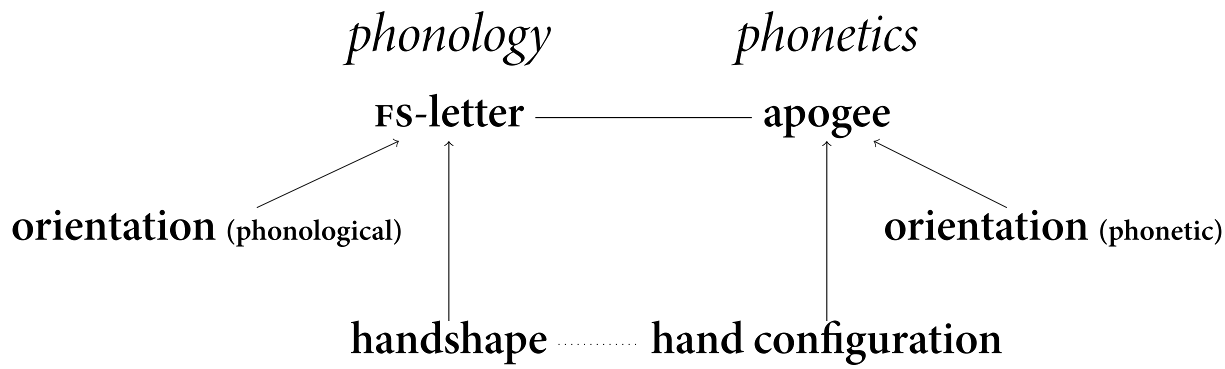


Figure 1.2: **Technical terms** A map of various terms used in this dissertation, and their relations to each other, as well as their situation in the traditional phonetics-phonology divide.

In the rest of this work, *apogee* will be used to refer to specific (phonetic) instances within specific fingerspelled words (e.g. "this apogee had a hand configuration hold of 93 milliseconds"), and *FS-letter* will be used to refer to one of the 26 possible canonical forms that make up the fingerspelling inventory (e.g. "the FS-letters that resist this phenomenon are -A- and -O-"). *Letter* will be used only to refer to orthographic letters within a written English word (which a fingerspelled word might be based on). The one exception to this is that in chapter 3 rates will be reported in letters per millisecond and durations frames per letter. This is to match previous literature which uses these terms, although more accurately they should be apogees per millisecond and frames per apogee.

7. This word was chosen so as to be neutral about what constitutes the boundaries of segments in fingerspelling.

1.3 Methodology

This work relies heavily on quantitative approaches to answer questions about the linguistic structure of fingerspelling. Quantitative methods are in no way new in linguistics, especially in phonetics; however recently it has seen a burgeoning in its use for phenomena especially in areas beyond phonetics. Throughout this work the methods and tools that are used for analysis will be described alongside the description of the results[8]. However, one model type is so central to every analysis, we have set it aside here for a brief introduction.

Much of the quantitative data in this work is modeled using hierarchical regression models. These are also more commonly known as mixed effects regressions. At their core, these models make a prediction (also known as the outcome, or dependent variable) that is either linear (analyzed with a hierarchical linear regression) or categorical (analyzed with a hierarchical logistic regression) in nature. This prediction is modeled on predictors (also known as inputs, or independent variables). There are a number of advantages to using hierarchical regression models. First, hierarchical models are more robust against unbalanced designs (for example, here, 2 signers fingerspelled words from 2 word lists and 2 signers fingerspelled words from only 1 yielding double the amount of data for 2 of the signers compared with the other 2). Even more importantly, hierarchical models were chosen because they account for the structure among the properties of the data that is being analyzed. In order to illustrate this, consider the structure of some of the data that will be analyzed here. Each production in our fingerspelling corpus has a number of properties about it that we want to include in our analysis:

- it has a word identity (which it shares with a few other productions),
- it was fingerspelled in a given trial (pair of word repetitions),
- within this trial it was either the first or second repetition,

8. Inspired, in part, by the use of instructions at the point of need (for a number of excellent examples and discussion, see (Tufte Forum Users, 2014)). Which puts the information needed to understand something as proximate that something as possible.

- these trials are ordered within each wordlist,
- each word has additional properties:
  - its length,
  - its type (i.e. name, noun, non-English word),
  - which wordlist it is a member of,
- finally, each production was fingerspelled by a specific signer.

Some of these properties are related, in that they are nested within each other: words are associated with a single wordlist. All of these properties may influence the rate of fingerspelling, and so need to be included in the model. There is a distinction between properties like these that are used as predictors (often called fixed effects), and properties that are used as grouping variables (often called random effects[9]), that is, those that define groups, and the structure of those groups, within the data.

The choice between what is used as a predictor and what is used as a grouping variable is not uncontroversial (Gelman & Hill, 2007; Barr *et al.*, 2013; r-sig-mixed-models listserv, 2010; glmm wiki, 2014) (further discussion on this point is included in chapter 3). However, the general division is as such: predictor variables are a finite set of variables (like conditions or treatments in a classical experiment), that we expect to have a direct effect on the outcome of the process that we are looking at. These are what would be included in a traditional regression, where each is associated with an effect of some magnitude (and sign). Grouping variables, on the other hand, are groups within the data that we are interested in generalizing over. Although there might be theoretically interesting results that come from the modeling of grouping variables (being able to describe the amount of inter-signer variation, for example), they are also used as a way of controlling for the fact that the data in a study is almost always a sample of the larger population (of people, words, etc.) and what we

9. Although the names fixed and random effects are fairly widespread in the linguistics literature, I am following Gelman & Hill (2007), among others, who describe them as predictors and grouping variables. This is for a number of reasons, the main being that the names fixed and random effects are not transparent to what they are doing in the model. Additionally they have a multiplicity of sometimes contradictory uses (Gelman & Hill, 2007, 245).

want to do is generalize across all of the individual levels, so as to predict as precisely as possible how a level that is out-of-sample (individual, word, item, etc.) would react given the predictor variables. Again, the distinction between these two levels is not always clear cut, and an exploration of this distinction is outside the scope of this dissertation, but see Gelman & Hill (2007) among others for much more detailed discussion.

The outputs of hierarchical models are as follows: there is an intercept for the outcome (the interpretation of which varies depending on the scales and types of predictors used) but is roughly equivalent to the mean response for default levels of categorical predictors (or, for all of the data under some contrast coding schemes) and values of zero for continuous predictors. Then, for each predictor (and interactions specified between predictors), the model generates a coefficient which is the magnitude and direction (sign) of the effect that the predictor has on the outcome. Grouping variables make adjustments to the intercept (also called random intercepts) or predictor coefficients (also called random slopes) based on group membership of a given data point.

Terms used here	also known as
hierarchical regression	mixed (effects) regression, multilevel regression
outcome	dependent variable
predictor (variable)	independent variable, input, fixed effect
grouping variable	random effect, level

Table 1.1: **Technical terms** Terms used in this work to describe hierarchical regression models, as well as other names used in the literature. See above for a discussion of why some of the terms used here were chosen, and (Gelman & Hill, 2007) for much more detail.

1.4 Roadmap

This dissertation is split into three large areas: 1. a model of the phonetics-phonology interface for handshape in sign languages, 2. a quantitative analysis of the temporal properties of fingerspelling, 3. a quantitative analysis of one phenomenon of handshape coarticulation in fingerspelling. Although these form the work as a whole, each chapter is relatively self-contained, and should be able to be read and understood on its own. The one notable exception to this is the discussion of the hierarchical logistic regression in chapter 4 will assume knowledge of hierarchical linear regressions (which will be used in chapter 3)

Although the phonology of sign languages in general, and handshape specifically, has seen quite a bit of study, the phonetics-phonology interface has not been explored as much. Chapter 2 proposes a model of the phonetics-phonology interface called the Articulatory Model of Handshape. This model builds on both the articulatory phonology and sign language phonology literatures, extending it to cover handshape in sign languages. This work is not only a proposal and characterization of a model, but it is accompanied with a computational implementation. Part of this implementation is a mapping from phonological features to phonetic targets. This implementation is important because it allows for precise understanding of how choices of phonological specification, as well as the mapping from phonological features to phonetic targets, affect the system of handshapes that has been proposed.

The temporal properties of fingerspelling have also seen quite a bit of study, although most studies have been quite limited in the number of tokens that they have analyzed. Chapter 3 explores a large corpus of fingerspelling from American Sign Language, and analyzes the temporal properties of this corpus. There are a large number of factors that contribute to the temporal properties of a given fingerspelled word (as well as the letters that make up that word). A number of variables showed a large amount of variation (especially intersigner variation), which could be one source of the wide range of rates that have been reported in the literature.

These temporal properties are then used as predictors in an analysis of one aspect of handshape coarticulation in fingerspelling. Context-dependent phonetic variation (especially coarticulation) is seen broadly across segments in spoken languages. Chapter 4 concentrates first on three case studies that exhibit handshape variation, and then looks at detailed quantification of pinky extension in a large corpus of fingerspelling data. Both the case studies and the deeper analysis support general and specific hypotheses that are predicted given articulatory phonology models of the phonetic implementation of handshape in fingerspelling. The analysis of pinky extension here shows that there is clear contextually-based variation: When a segment is close to another segment that has an extended pinky finger it is more likely to also have an extended pinky finger, even if it does not canonically have pinky extension. This pattern is mediated by a number of factors including the speed of fingerspelling and phonological features of the segment of interest. This pattern is predicted by the articulatory model of handshape (discussed in detail in chapter 4). Although it is commonly accepted that the inactive articulators are more susceptible to coarticulation, testing this directly in spoken languages can be difficult. Handshape in ASL provides a perfect test case for this because the articulators, when they are inactive, can take on configurations that are (on the surface) identical to configurations that they can take on when they are active. Because of these we can see a clear distinction: when a pinky is flexed and active it is much less likely to be extended when it is surrounded by an extended pinky; however, when a pinky is flexed and inactive, it is quite susceptible to coarticulatory extension.

This work contributes to articulatory phonology, as well as theories of speech production broadly by studying the distinction between active and nonactive articulator gestures. Handshape in sign languages is especially well-suited to study this phenomenon because there are many possible combinations of active and nonactive articulators (all five digits). Additionally, unlike most articulators for spoken languages, all of the articulators can be seen and tracked easily without the occlusion of the cheeks and neck. This work establishes general norms for fingerspelling in native ASL users. Having quantitative norms of specific features of fingerspelling allows for the development of metrics and tests for what types of productions fall outside of the range of typical signers. This has further impacts on diagnosing language disorders, which has been particularly understudied in ASL signers. There has been research showing a correlation between fingerspelling ability and literacy (Haptonstall-Nykaza & Schick, 2007; Emmorey & Petrich, 2011). Understanding basic phonetic facts about the production of fingerspelling will allow for more detailed future work on the perception of fingerspelling. Furthermore, understanding how fingerspelling is produced and perceived will enable the study of this correlation in more detail.

Chapter 2

The Articulatory Model of Handshape

Sign language phonology has been explored since the advent of sign language linguistic research (Stokoe, 1960; Mandel, 1981; Liddell & Johnson, 1989; Sandler, 1989; van der Hulst, 1995; Brentari, 1998). As with all languages, there is considerable variation within groups that are described as a single phonological category. One source of this variation is phonetic variation that is the result of the physical, articulatory implementation of abstract phonological categories by the signer. Most of this research has concentrated on the phonological structure, and how this structure interacts with morphology of signed languages. This chapter (and dissertation), in contrast, concentrates on the phonetics-phonology interface, an area that has seen relatively little research. This chapter develops a model of the phonetics-phonology interface for handshape that accounts for certain types of variation that is observed in sign languages (e.g. coarticulation). Section 2.1 gives an overview of sign language phonology. Section 2.2 describes theories of the phonetics-phonology interface for spoken languages. Section 2.3 describes gestures and articulatory phonology, proposing a new way to deal with inactive articulators. Section 2.4 describes the Articulatory Model of Handshape (AMOHS) and its consequences for variation in handshape in sign languages. Finally, section 2.5 details an implementation of the Articulatory Model of Handshape as a Python module, and its ability to translate between phonological and phonetic models of handshape.

2.1 Sign phonology

Although a number of different models have been proposed, most agree that all sign languages have five major parameters: handshape, movement, location, orientation, and non-manual markers (Battison, 1978; Mandel, 1981; Liddell & Johnson, 1989; Sandler, 1989; van der Hulst, 1995; Brentari, 1998; Eccarius, 2002; Sandler & Lillo-Martin, 2006). All sign languages have a number of different categories that are phonologically contrastive within each of these parameters. Every sign has at least one of each of the major categories. See figures 2.1–2.5 for examples of minimal pairs of handshape, location, movement, orientation, and non-manuals respectively.

(a) CANDY start (b) CANDY end (c) APPLE start (d) APPLE end

Figure 2.1: Handshape minimal pairs[1]

(a) DRY start (b) DRY end (c) SUMMER start (d) SUMMER end

Figure 2.2: Location minimal pairs

(a) NAME start (b) NAME end (c) EGG start (d) EGG end

Figure 2.3: Movement minimal pairs

2.1.1 *Handshape in depth*

Early work on sign language phonology treated handshape as a single holistic feature of signs (Stokoe, 1960; Stokoe *et al.*, 1965); however, more recent work on the phonology of signed languages does

1. These, and the following (figures 2.1–2.5) images are material courtesy of Bill Vicars and lifeprint.com

(a) CHILD start (b) CHILD end (c) THING start (d) THING end

Figure 2.4: Orientation minimal pairs

(a) LATE (b) NOT-YET

Figure 2.5: **Non-manual markers minimal pairs** The tongue protrudes for NOT-YET , but does not for LATE .

not take such a monolithic view of handshape (Mandel, 1981; Liddell & Johnson, 1989; Sandler, 1989; van der Hulst, 1995; Brentari, 1998; Eccarius, 2002; Sandler & Lillo-Martin, 2006). Each of these systems has mechanisms to account for handshape within signing, but rather than assuming that each handshape is entirely unique—where similarities or differences between them are accidental—they decompose each handshape into a number of (phonological) features allowing for relationships to be established between handshapes based on featural similarities. They all make use of a system of selected versus nonselected fingers to divide the hand into groups based on what fingers are active in the handshape. The only exception is Eccarius (2002), who splits selected fingers into two groups: primary and secondary selected fingers. Many of these models established the existence of the selected versus nonselected distinction by looking at the distribution of handshapes in signs with two-handshapes in sequence: although a sign can contain two handshapes, these two handshapes must have the same set of selected fingers (Brentari (1998) among others).

Referencing selected fingers has been argued to explain a number of phenomena in signed languages:

1. **Only a subset of fingers are able to move in handshape contours**

   In signs that have two handshapes both handshapes must have the same set of selected fingers (Mandel, 1981).

2. **Minimal pairs lexically (ASL: APPLE versus NERVE) and in classifiers**

   Some signs vary only in the number of, or which, fingers are selected (Brentari, 1998). Although this distinction could be accomplished without referencing selected fingers by specifying joint configurations for each finger separately, this fails to capture the generalization that what are called selected fingers all have the same joint configurations.[2]

3. **Handshape assimilation in compound signs**

   The prediction here is that when two handshapes are combined together into a single handshape it is the selected fingers of either handshape that will be preserved (Sandler, 1986). On the surface, this seems similar to coarticulation in fingerspelling (in that it is the blending of two handshapes together), but it is operating at a very different level of representation: in compounds the blend is the result of blending two lexical items at a very abstract level; in fingerspelling, coarticulatory blending is a property of the implementation of the phonetics-phonology interface.

Although it is widely assumed that nonselected fingers are inactive, in the work on selected versus nonselected fingers, up to this point, no one has explicitly linked the selected fingers with what are in spoken language linguistics called active articulators. Mandel compares the selected fingers to the active hand to describe the fact that they can have more complex configurations (Mandel, 1981, p. 82):

2. This is a bit too simple, as can be seen by the innovation of secondary selected fingers by Eccarius (2002), but there are at most two groups of selected fingers.

> The selected/other distinction of fingers on the Internal scale is comparable to the External distinction between active hand(s) and passive or uninvolved hand. The active hand, the dez, is the foreground hand. It can have any hand configuration in the inventory. [...] Similarly, in the hand, the selected fingers can take any position except the closed position, in which they would merge with the outline of the midhand and lose their identity as fingers.

But he does not discuss the implications this has for variation of the selected or nonselected fingers. In fact, Mandel might have imagined a model similar to this; however, this work is before the advent of articulatory phonology, and possibly the widespread use of the terms active and inactive articulators as they are used today. In the Articulatory Model of Handshape, which will be discussed in detail in section 2.4, I make the connection explicit: selected fingers are the active fingers, nonselected fingers are the inactive fingers.

The variation in ASL fingerspelling that will be discussed in chapter 4 contributes to this as another phenomenon that supports the existence of selected fingers (including (Keane *et al.*, 2012a, forthcoming)). Additionally fingerspelling is in many ways an ideal area in which to look for variation in handshape because: 1. Fingerspelling has a large number of individual handshape tokens. 2. These tokens are in a wide variety of contexts; in principle any handshape can precede or follow any other. and 3. Fingerspelling uses 72% of the possible handshapes in ASL (Brentari & Padden, 2001). As such, fingerspelling is a good phenomenon to analyze handshape variation in ASL generally. Moreover, because fingerspelling is more sequential than other types of signing, the resulting phonetic analyses will allow for more direct comparison with similar spoken language work in terms of assessing the effects of articulatory ease, gestural overlap, and gestural activation in fingerspelling production.

2.1.2 Variation in handshape in sign languages

Recently, there has been work that looks at variation within the phonological categories of handshape: Liddell & Johnson (1989) note the existence of handshape assimilation, giving the concrete example of the handshape of a first person pronoun that assimilates to the handshape of the following predicate. Hoopes (1998) notes the existence of pinky extension coarticulation in fingerspelling as well as in signing. He separates it from the pinky extension that he is interested in, which is pinky extension that is present for an entire sign. He argues that this latter type of pinky extension is a sociolinguistic marker. Cheek (2001) finds anticipatory and perseveratory coarticulation between the 1-handshape and 5-handshape in ASL. She finds that this is also dependent on rate where faster signing results in more coarticulation. Bayley *et al.* (2002) looks at coarticulation of the 1-handshape from a corpus of signers from a variety of regions across the United States. They find that multiple factors affect the realization of the 1-handshape, including "grammatical function and features of the preceding and following segments, as well as a range of social constraints including age, regional origin, and language background." They note that the grammatical category of the 1-handshape signs they analyze have a stronger effect than that of coarticulatory pressures. Parisot (2003) notes that the handshape of pronouns tends to assimilate to the surrounding context in Quebec Sign Language. Mauk (2003) found rate-conditioned coarticulation of 1-handshapes making them more like surrounding 5-handshapes. Additionally, he found that there was no significant coarticulation in 5-handshapes when they had surrounding 1-handshapes (which is the opposite of what was found for 1-handshapes). Jerde *et al.* (2003) found that there is both assimilatory and dissimilatory coarticulation for various parts of the hand. Finally, Fenlon *et al.* (2013), using methods similar to Bayley *et al.* (2002), found that 1-handshape signs varied with "the preceding and following phonological environment, grammatical category, indexicality, [and] lexical frequency". They also found no significant social factors, except for region. There has also been work on variation in location (Wilbur & Schick, 1987; Meier & Holzrichter, 2000; Crasborn, 2001; Lucas *et al.*, 2002; Mauk & Tyrone, 2008;

Grosvald & Corina, 2008; Tyrone & Mauk, 2010; Tyrone *et al.*, 2010; Mauk & Tyrone, 2012) although this dissertation will concentrate on handshape variation.

The above work has shown that there is coarticulatory variation in handshape in sign languages generally, and ASL specifically. The variation due to coarticulation is in addition to other kinds of variation seen across languages (e.g. dialect differences). The specifics of how this coarticulation is implemented or can be modeled have not (yet) been explored. Mauk notes that there is a lack of coarticulation on the 5-handshape: "It appears that fingers specified as selected in the phonological description of a handshape may have little flexibility in terms of their precise position within that handshape. As a result, unselected fingers may be more prone to rate dependent undershoot." (Mauk, 2003, p. 265) He connects the lack of coarticulation with the selected/nonselected distinction, but does not give an explanation of how this influences coarticulation. The later sections of this chapter are devoted to developing a model that makes predictions about what kinds of coarticulation we expect to see given phonological models of the phonetics-phonology interface, handshape, and motoric constraints of the articulators. The model developed here predicts what Mauk observed, as well as other details of coarticulation.

2.2 The phonetics-phonology interface

In the modern linguistic era, phonetics and phonology have frequently been completely separated where phonology operated on segments that are (temporally) discrete bundles of features, and phonetics is the physical instantiation of these segments. Some mid-century models (e.g. SPE (Chomsky & Halle, 1968)) made a distinction between systematic phonetics and physical phonetics. Systematic phonetics was the output of the phonological system, with phonetically grounded features that were still idealized, abstract, and segmental in nature. Physical phonetics took these, and then converted them into the physical form of the language via biomechanical and physical properties that are universal to all humans (Ladd, 2011). Where this division should be made has been increasingly called into question: where does phonology end and phonetics take over? That is, do processes like assimilation and coarticulation belong in phonology (are they operations on abstract, symbolic representations), or do they belong in phonetics (where they are operations that can apply gradiently in time or amount)?

2.2.1 *Separating phonology from phonetics*

At its simplest, the phonetics-phonology interface is a mapping from a discrete, symbolic phonological system, to a gradient, physical system. This distinction is manifested in statements like "phonology is categorical, phonetics is gradient". While on the surface this distinction seems right, this has been shown to be something of a false dichotomy in the later part of the century. First, there are numerous examples of perceptual cues for a specific segment extending well beyond the strict limits of time for a particular segment. This is a problem for the strictly segmental conceptualization of phonology (as well as systematic phonetics), labeled here *indivisibility*. Second, the implementation of the gradient, phonetic objects (be they sounds, handshapes, etc.) and their interaction is something that is not universal, as can be seen by cross-linguistic differences of coarticulation, phonetic details of specific segments, as well as resting states, labeled here *non-universality*.

Indivisibility

There are numerous examples of perceptual cues existing outside of the strict boundaries of a single segment. First, in many languages including English, the voicing of stops is not only cued by the vibration of vocal folds during the closure of the stop, but also by the time it takes for the vocal folds to start vibrating during the execution of the following vowel. In English specifically, for many speakers the voiced/voiceless distinction is not made by vibrating the vocal folds during the stop closure, but by the lag of voicing in the articulation of the following vowel. For voiced stops, the vocal folds start vibrating with a very short lag (0–10 msecs), whereas for voiceless stops the lag is longer, on the order of 50–70 msecs (Klatt (1975) among many others).

Another similar phenomenon in English is that the vowels preceding voiced stops are longer than those preceding voiceless stops. Raphael (1972) found that, "with one exception and regardless of the voicing cues used in their synthesis, all final consonants and clusters were perceived as voiceless when preceded by vowels of short duration and as voiced when preceded by vowels of long duration". In other words, the effect of the preceding vowel length outweighed other cues (including the presence of voicing during the consonant) about the voicing of final stops in English.

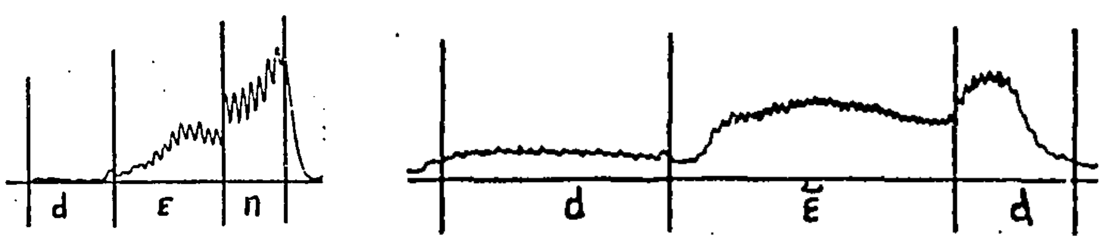
Finally, it has been shown that the vowel following a fricative can alter the category that the fricative is perceived as belonging to. When followed by an [u], English speakers responded with [s] more frequently than [ʃ] for stimuli that are ambiguous between [s] and [ʃ]. There have been a variety of studies (Mann & Repp, 1980; Repp, 1981; Nittrouer & Studdert-Kennedy, 1987), but the basic setup is, subjects are given stimuli on a continuum from [s] to [ʃ] in different vowel contexts. The identification curves are more biased toward [s] when the following vowel is an [u] than when it is an [a] (or [i] in (Nittrouer & Studdert-Kennedy, 1987)). This means that for the same point around the center on the continuum, speakers identify the sequence [(s|ʃ)u] as [su], but [(s|ʃ)i] as [ʃi]. The explanation given is that speakers perceptually correct for the anticipatory lip rounding from the [u], effectively erasing the rounding cues from an ambiguous [(s|ʃ)] token.

These three phenomena show that the perceptual cues for a particular segment are not always limited to the time period for that specific segment, but rather some properties of adjacent segments are important for the perception of a target segment. It is not always possible to draw such clear temporal boundaries between segments, as a phonologically segmented string, or even a narrow phonetic transcription would indicate.

Non-universality

Second, not all languages implement aspects traditionally described as phonetics in the same way. There is variation in the way that segments interact with one another (e.g. coarticulation). There are large variations in categories typically assumed to be the same. Finally, the articulatory setting (simplistically: rest position for speech) of languages varies.

Although many attribute coarticulation to limits on the speed of specific articulators, numerous examples show that this is not the sole reason for coarticulation. One often cited example is that nasal coarticulation and spread is the result of the velum being a slow(er) articulator than others (e.g. tongue, lips, jaw). This example, however, is simply not true: the speed of the velum is not the only source of coarticulatory pressure, but rather, different linguistic systems produce different outcomes (Louis Goldstein, personal communication). Vowel nasalization in French and English show distinct patterns that cannot be attributed to the physiological limits of velum movement. Cohn (1990) shows that English has large amounts of anticipatory nasal coarticulation, where (oral) vowels preceding nasal segments are (at least partially) nasalized. French, on the other hand has contrastive oral and nasal vowels. In English there is a cline where nasalization increases at a steady pace throughout (the latter half of) the vowel; whereas in French, nasalization is nearly instantaneous during a nasal vowel segment, with nasalization starting within the first 10–20 msecs of the vowel starting, and staying stable (at a plateau) throughout the duration of the vowel. See figure 2.6 that illustrate this.



(a) English: den /dɛn/ (b) French: daim (deux) /dɛ̃#d/

Figure 2.6: **Nasal airflow for English and French words with nasals**, from (Cohn, 1990)

There are large variations in categories typically assumed to be the same. One example of this is where a particular vowel falls within a language's vowel space. Although a vowel from two different languages may sound similar (even to trained phoneticians), and thus be transcribed with the same IPA symbol, they have very different articulatory (and thus acoustic) properties. Bradlow (1995) found differences between English and Spanish speakers' vowels, specifically that English speakers have higher F2s than Spanish speakers. Additionally English speakers' vowel spaces are slightly larger than Spanish speakers. The tightness of vowel categories is not significantly different across English and Spanish, even though Spanish has a smaller vowel inventory. Chung *et al.* (2012) expanded on Bradlow's work with more languages (comparing 5 languages: Cantonese, American English, Greek, Japanese, and Korean). They found that, even when normalizing for variation of speakers' vocal tract length, there are still differences in vowel spaces for the vowels /i/, /e/, /o/, and /u/.

Wilson (2006) describes the history of articulatory settings[3]: The idea that there are different default positions of the articulators (sometimes called the rest state, and more recently articulatory setting) for different languages was described as early as Wallis (1653/1972). Although this idea has been around and discussed for decades, only recently has there been available technology to measure these differences instrumentally (Gick *et al.*, 2004; Wilson, 2006; Wilson & Gick, 2006). This line of work has found that English and French speakers have different inter-speech postures, which they claim are evidence of differences in the default articulatory settings. They find, for example, that English speakers have a higher tongue tip, more protruded lips, and more narrowed lips from maximum spread (Wilson & Gick, 2006). These differences are even retained to some extent in bilingual speakers. Wilson (2006) tested English/French bilinguals, and found that for bilinguals (who are perceived as native) in both languages, their articulatory settings while speaking English were significantly different from their articulatory settings while speaking French, in the same direction as the differences between monolingual speakers of each language. Thus, it is not the case that these differences are attributable to physiology, or some universal property of the body. Rather, these articulatory settings must be learned by speakers as they acquire the language.

3. We use the term *articulatory settings*, because that is what is used in previous literature cited here. Others have called this *neutral position* or *inter-speech postures*.

These three phenomena show that there are parts of what has, in the past, been described as phonetic details, that must be learned by children as they are acquiring the language. The amount of nasal coarticulation in English cannot be explained solely by the slowness of the velum, because speakers of French are able to transition from a nasal to a non-nasal without an intervening segment. The distribution of the same set of phonologically distinct vowels varies across languages in a way that is not predicted if children are just learning the same categories (or categories that are based on the same set of features like [±high], [±low], [±front], [±back], etc.). Finally, the differences in articulatory settings of different languages cannot be explained by universal physiological properties, but rather is something that children must learn when they are acquiring their native language (or languages). There is even literature showing that children learn (at least some) of these properties by 11 months (Seidl *et al.*, 2009).

Some phenomena are easy to label as clearly phonetic, and others are easy to label clearly phonological. However, both of the problems of indivisibility and non-universality show that it cannot be the case that all parts of phonetics are universals based on human physiology and all parts of phonology are completely segmented and abstract. There is an interaction between the phonetics and phonology that makes drawing a clear division between the two difficult (or impossible).

2.2.2 *Models of coarticulation*

Coarticulation is one of the main areas where people have tried to draw a dividing line between phonetics and phonology. As alluded to above, there are clear patterns of contextual dependence for the articulation of segments based on what surrounds them. There have been a variety of proposals as to which area coarticulation rightly belongs to: is it a phonological or phonetic phenomenon? Initial conceptions of the divide held that phonological phenomena are learned and categorical, whereas phonetic phenomena are universal and continuous. For many reasons, including those discussed above as well as phenomena discussed later in this dissertation, this distinction between phonetics and phonology cannot be drawn so explicitly. Some newer research has taken to modeling phonetics and phonology together: Poeppel *et al.* (2008); Poeppel & Idsardi (2011) for example develop a model of speech perception that relies on processing and categorization using both categorical and continuous approaches.

A complete review of the coarticulation literature is beyond the scope of this dissertation, but can be found in (Farnetani & Recasens, 1999), among others. There are three main themes of models: 1. target-undershoot, 2. feature spreading, and 3. coproduction.

The target-undershoot model by (Lindblom, 1963; Moon & Lindblom, 1994) hypothesized that coarticulation was the result of the most economic transition between two sounds. The relationships would thus be extremely local, allowing for only neighboring segments to interact. In order to account for vowel-to-vowel coarticulation, Öhman (1966, 1967) proposed that consonant and vowel gestures are programmed separately and overlaid on top of each other. This allows for the vowels to interact across a consonant (as well as be influenced by some aspects of the consonantal articulation). Here coarticulation is a result of the interpolation between two segments, and is driven only by constraints on the motor system. It predicts that possible values of coarticulation will also be only between the points specified by adjacent segments. For example: a mid vowel followed by a high vowel will have a cline in height that goes from mid to high (so long as an intervening consonant does not have a height feature associated with it). This model is arguably phonetic in nature, in that the articulators transition in the most articulatorily economical way, which is a universal property of humans and the articulators being used.

The next major model type is that of feature spreading, including the look-ahead model (Daniloff & Hammarberg, 1973; Hammarberg, 1976). In this model features spread right to left, from specified to unspecified, which explains anticipatory coarticulation. Left to right, or carry-over coarticulation is deemed a passive response of the articulators. This model predicts that coarticulation will be categorical in both time (across segments) and activation (amount of the feature being coarticulated). This is demonstrably false for at least some coarticulation phenomena, namely nasal coarticulation in English (as discussed above) is gradient in both time and activation. One solution proposed to this that fit with the feature spreading model was to add coarticulatory resistance (Bladon & Al-Bamerni, 1976). This adapts the look ahead model by giving each feature a weighting for how much it is allowed to coarticulate (or vary). Although this explains the activation variance, it does not explain the time dependence seen in this type of coarticulation. The feature spreading models so far are very much phonological in nature, in that coarticulation is an operation on the discrete, abstract segments. Keating (1980/1990) proposed the window model, which gives each feature a window of acceptable articulations. These windows vary if the feature is [+feature], [-feature], or unspecified. For unspecified features the window encompasses the entire range of motion from [+feature] to [-feature]. The articulators go between specified features with a minimal effort interpolation between them but must pass through an acceptable window. In this model, coarticulation can be either phonetic, where it is gradient in time and activation; or phonological, where it is more categorical in time and activation. Languages differ in which instances of coarticulation are which. Cohn (1990) found that although not phonologically contrastive, Sundanese has phonological coarticulation of nasals where vowels are nasal when they precede a nasal stop, but nasal airflow reaches a stable plateau for most of the vowel (similar to a language that has phonemic nasal vowels like French). English, on the other hand, has phonetic coarticulation with nasalization operating on more of a cline throughout the vowel preceding a nasal. Additionally, the width of the window in this model is language-dependent. This is an early indication, in the coarticulation literature, that something that is traditionally below the level of phonology (non-contrastive nasalization), and thus a part of phonetics, is a pattern that needs to be learned.

Finally, models of coproduction take a slightly different approach: instead of going directly from abstract features to articulatory targets, abstract features are converted into articulatory gestures which have not only targets, but also a time dimension. These gestures are then allowed to overlap temporally. When there is a conflict between them (that is, they overlap and have different specifications), two gestures are added together, producing gradience in activation. In addition, because each gesture has distinct periods within it: an activation period (onset), a plateau (with a target and release at the beginning and end respectively), and an offset; an individual gesture will be gradient in time over both the onset and offset (and consequently over any periods where these overlap with other gestures). A more detailed description of articulatory phonology, which is one such coproduction model, will be given, along with a novel adaptation, in section 2.3.

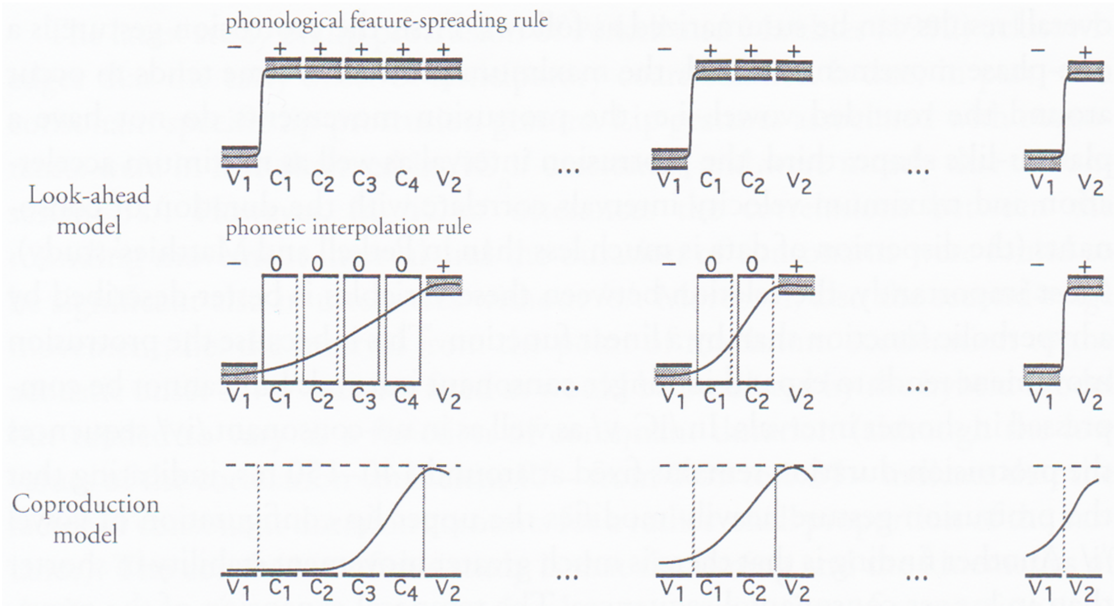


Figure 2.7: **Differences between different model types of coarticulation**, from (Farnetani & Recasens, 1999, 62)

There are a few additional models of the phonetics-phonology interface (notably the BiPhon model (Boersma, 2009)) using an OT-style set of ranked constraints to implement targets. These models need to rely on large numbers of constraints to approximate the same time and activation gradience seen by the models discussed above. Because of the lack of ability to predict and model gradience without large numbers of ad hoc constraints, they will not be explored in detail here.

In conclusion, the phonetics-phonology interface has, in the past half century, seen a great deal of research. As models of coarticulation have evolved over time it has become clear that in order to account for patterns of coarticulation, models must allow for both temporal and activation gradience. Additionally, this gradience is not something that is a universal property of the motor systems of humans, but rather is bound by parameters that children must learn as they are acquiring a language. Because of this, whether coarticulation (or a particular kind of coarticulation) is phonetic or phonological in nature is not in itself an important question, because either way a speaker must learn the parameters needed to generate the pattern appropriate for the language they are acquiring.

2.3 Gestures and articulatory phonology

2.3.1 A brief overview of articulatory phonology

Articulatory phonology is a theory of phonetics and phonology that assumes the basic units of speech are articulatory gestures (beginning with (Browman & Goldstein, 1986, 1992), with many others following). These gestures are dynamic and unfold over time, which allows them to overlap and interact. This interaction can produce both (phonologically) meaningful contrasts as well as phonetic variation (e.g. coarticulation). Individual articulators come together to act on *tract variables*, which are the units of gestures that are phased with respect to each other to produce language (see figure 2.8 for the tract variables and articulators in the vocal tract). Recent work on the cognitive neuroscience of speech perception (Poeppel *et al.*, 2008; Poeppel & Idsardi, 2011) supports a gesturally based theory like articulatory phonology: "The final, featurally specified representation [...] constitutes the format that is both the endpoint of perception – but which is also the set of instructions for articulation." (Poeppel & Idsardi, 2011, 8)

Gestural scores (like that in figure 2.9) can be created that consist of tract variables that are activated (either constricted or opened) over periods of time (the boxes in the figure). Given this general specification, tract variable trajectories (the lines in the figure) can then be computed. Note that this score is explicitly underspecified: "not every tract variable is specified at every point in time" (Browman & Goldstein, 1992, 28) where articulators that are not specified assume a default state. It has been observed that this cannot be the whole story. In particular Browman (1994) found that in order to predict the movement of the jaw after certain stop closures there must be an *active*

	tract variable	articulators involved
LP	lip protrusion	upper & lower lips, jaw
LA	lip aperture	upper & lower lips, jaw
TTCL	tongue tip constrict location	tongue tip, tongue body, jaw
TTCD	tongue tip constrict degree	tongue tip, tongue body, jaw
TBCL	tongue body constrict location	tongue body, jaw
TBCD	tongue body constrict degree	tongue body, jaw
VEL	velic aperture	velum
GLO	glottal aperture	glottis

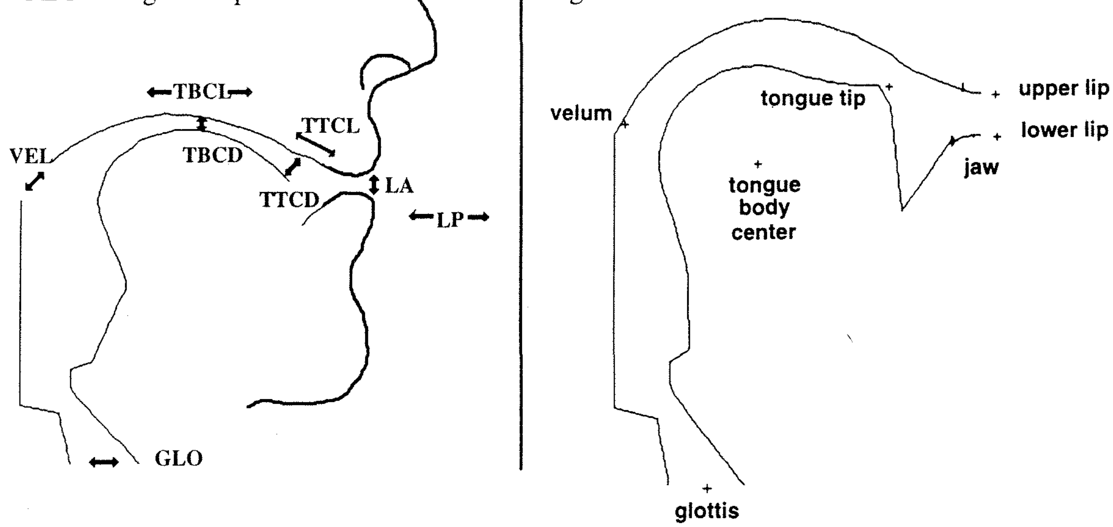


Figure 2.8: **Tract variables and articulators involved with each for the vocal tract**, from (Browman & Goldstein, 1992, pp24)

release, that is the articulator cannot be left to passively return to a neutral state, but rather there must be a gesture of some sort to propel it away form the closure point:

> While it is possible that a lowering 'kick' should be used, i.e. a 'Gesture' with no target, that is not possible to test with the current version of the task-dynamic model. Therefore, a consonant release with constriction-degree and constriction-location targets was added to the computational Gestural model, basically in order to test whether the tongue body could remain in a constant position even in the presence of an active release.
>
> (Browman, 1994, pp345)

Articulatory phonology is a connection between very abstract segmental phonological specifications and the phonetic implementation of speech (including signing)[4]. It explicitly describes how phonological features are translated into physiological reality. Through this, it makes predictions about what kind of variation (e.g. coarticulation) will be found in language vis-a-vis gestural overlap. The way that articulatory phonology describes articulator gestures as well as their timing, is particularly well-suited to describe how different parts of the hand are configured over time to produce the handshapes that are necessary for signing. The division of articulators into subparts that are each associated with specific gestures that are allowed to interact makes predictions about what kinds of contextually dependent variation should be observed in fluid signing, which would not be easily represented at a phonological level.

4. I use the term *speech* to refer to language production generally, both spoken languages and signed languages. This is because the underlying phonetic processes that generate the language signal are the same: at their most fundamental they are motor plans to move a set of articulators to targets in a structured way in order to form language.

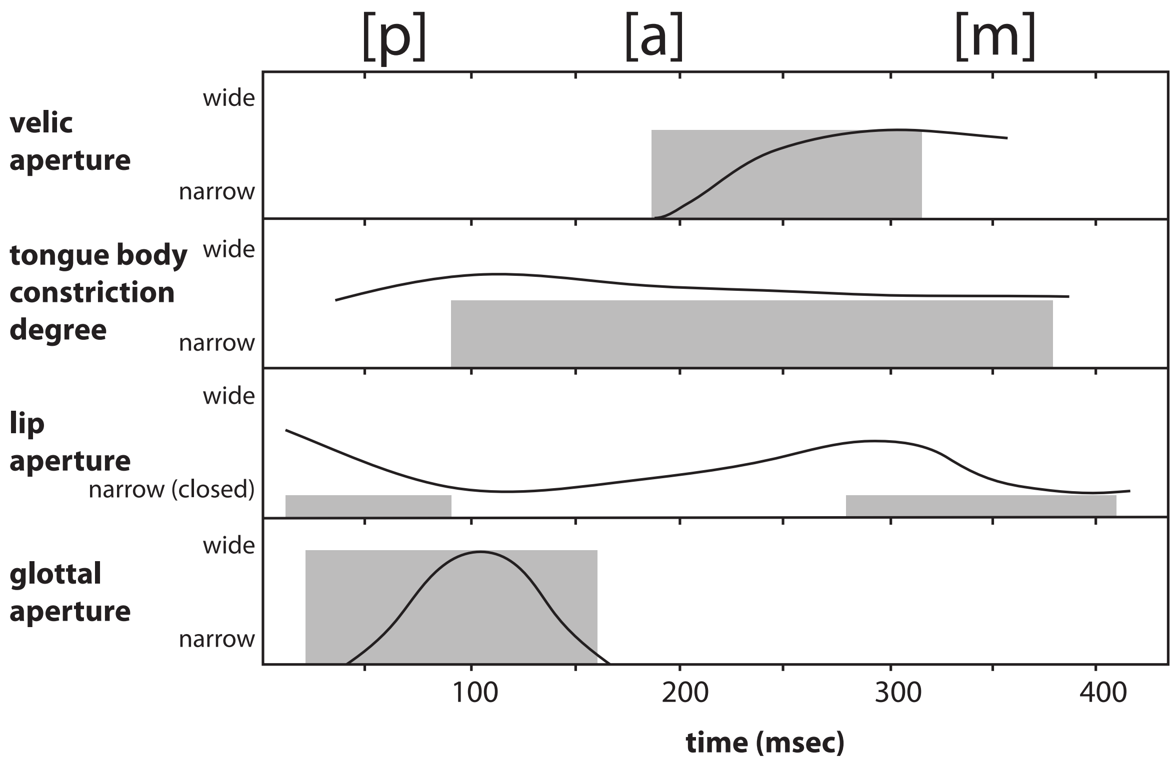


Figure 2.9: "Gestural score for the utterance 'palm' (pronounced [p^{h}am]), with boxes and tract variable motions as generated by the computational model. The input is specified in ARPAbet, so IPA [pam] = ARPAbet [paam]. The boxes indicate gestural activation, and the curves the generated tract variable movements. Within each panel, the height of the box indicates the targeted degree of opening (aperture) for the relevant constriction: the higher the box (or curve), the greater the amount of opening." (Browman & Goldstein, 1992, pp28) Figure recreated with clarifying labels on the y-axis.

2.3.2 *Implementing inactivity*

It is widely assumed in the articulatory phonology literature that when an articulator is not active (through being unspecified in the gestural score) it assumes a neutral state. One example of this is that the velum, when not active, assumes a closed position; only when it is actively opened does it deviate from that position. This assumption makes predictions about spoken languages that seem to be fairly robust: nasal sounds are more marked than non-nasal, and nasalization spreads from nasal sounds, etc. This neutral position, however, is at odds with the position that the velum assumes naturally when people are at rest (e.g. not speaking), which is open[5], allowing for air to be drawn into the respiratory system from the nose or mouth. Originally, these positions were assumed to be filled in at a low level in the task-dynamic system (Louis Goldstein, personal communication). This being the case, there must be some muscular activity on the velum articulator during periods that have previously been described as inactivity in order to keep it closed. One solution to this apparent problem is to specify gestures for periods previously assumed to have no activity.[6] These gestures are weaker than those that operate on the active articulators, but this can be accomplished using stiffening or dampening as has already been described to differ between tract variable gestures, speaking rates, etc. This will make similar predictions that the current assumption of non-control of the passive articulators makes: namely that coarticulation is seen spreading from surrounding active gestures more on articulators that are nonactive, than those that are active. Additionally, by explicitly assigning nonactive gestures for articulators that are unspecified in the gestural score, we no longer need to specify a special active release, but rather all releases are necessarily active because the articulator is being pulled by a nonactive gesture to a position that is specified for each articulator.

5. In fact, the rest state is more open than even nasal segments (Bell-Berti, 1993).

6. These gestures are being tentatively called *nonactive*. Although this name is not as transparent as it could be, it is better than inactive or passive, which suggest a complete inertness. One possible better solution would be something like *lessactive*, although that implies that there might be more than a binary contrast between active and nonactive.

Although these default states might seem arbitrary, they might be motivated by other aspects of language production. Returning to the example of the velum being by default closed: we know that closing the velum during spoken language production has a number of acoustic and articulatory benefits. First, a closed velum allows stops to have a complete cessation of airflow, and thus have a large sonority differential in the acoustic signal that would not be possible if the velum were open either partially or fully. Second, by keeping the velum closed except for when it is explicitly specified as open, the contrast between nasal and non-nasal sounds is enhanced allowing for easier perception (and thus learning).

There are three major predictions that come from the fact that the targets associated with nonactive gestures are not a physiologically neutral state. First, it is possible that the targets for nonactive gestures will differ cross linguistically with different languages having different default states. Second, it is possible that the targets for nonactive gestures will vary depending on the targets of the active gestures. The first is supported in the work in spoken languages looking at default targets, or what are described as articulatory settings which vary from language to language (Wilson & Gick, 2006; Wilson, 2006; Gick *et al.*, 2004). The second will be used in the Articulatory Model of Handshape for the configuration of the nonactive (nonselected) fingers. Additionally, it makes a prediction that is not able to be accounted for with previous models of speech production (e.g. target-undershoot or feature spreading discussed in section 2.2.2): even when not active, an articulator could be pulled from a position between those specified on either side of the target segment. For an example in spoken languages: the tongue position for a consonant that is flanked by a mid vowel and then a high vowel, such as the [m] in [sɛmi] might actually be lower than the mid-to-high window that is specified in previous models of coarticulation.

Although this does not appear to have been proposed before, it makes explicit a widely used assumption with the same machinery that is already in use to describe other phenomena. This proposal makes explicit the underlying default states which were used in the task dynamic implementation. Making this explicit has a number of advantages: 1. it allows for an easy explanation of languages varying in their default articulatory settings, as discussed above. 2. It accounts for the active release of gestures that required an additional stipulation about the system before. 3. It makes predictions about variation with respect to coarticulation. Exactly where the gestures for the unspecified articulators are implemented is beyond the scope of this work, but in principle, they could either explicitly be added to the gestural scores or be specified by some sort of default rule at a lower level, but before the general task dynamic implementation. Finally, this proposal is intended to be language-general and independent of modality. Testing this proposal on signed languages is convenient because the articulators are visible as opposed to spoken languages, where beyond the lips and the very front of the oral cavity, the articulators are hidden from easy view.

2.4 The Articulatory Model of Handshape

At first glance signed languages and spoken languages would seem to be irreconcilably different with respect to their phonetics and phonology. Many phonological models of signed languages use features that are based on the articulators in use (i.e. hands, arms, body, and face) to the exclusion of features that are based on the articulators used for spoken languages (although there is some work attempting to unify the underlying features to be the same (Peter Jurgec, personal communication)). The phonetics of signed and spoken languages, however, show remarkable similarity: even though the articulators themselves are different, in both cases the final result is a person using muscles to move articulators to positions, defined by a language, in sequence to convey a message to a perceiver. There has been a small amount of work looking at ASL phonology from an articulatory phonology perspective (Tyrone *et al.*, 2010), although no one has yet attempted to model handshape using articulatory phonology. As a first step in integrating the models of handshapes discussed in section 2.1 into an articulatory phonology framework, we must identify what the articulators and tract variables are. This work looks at fingerspelling specifically because fingerspelling consists of dense, rapidly changing information. Moreover, and as discussed above, fingerspelling is more sequential than other types of signing, so the resulting phonetic analyses will allow for more direct comparison with similar spoken language work in terms of assessing the effects of articulatory ease, frequency, and phonological processes in sign production. This is not meant to imply that this sequentiality means that the production of each unit of fingerspelling is discrete, but rather, just as with spoken language, the production stream of fingerspelling is actually the confluence of gestures from a number of independent and semi-independent articulators to produce a stream of language. By quantifying variation in handshape as well as specific articulators, it will be possible to start to develop a gestural score for the hand during fingerspelling.

The articulators that make up the hand are the fingers (index, middle, ring, and pinky) and the thumb. Each finger can be flexed at each of three joints: the metacarpophalangeal (MCP), proximal interphalangeal (PIP), and distal interphalangeal (DIP) joints. Each finger can be spread from its neighbors, a process called abduction; and each finger pair can be abducted independently from the other finger pairs. The thumb has two joints which can be flexed: the metacarpophalangeal (MCP) and distal interphalangeal (DIP) joints. Additionally the thumb can be opposed (also known as palmar abduction away from the palm) and abducted (also known as radial-ulnar abduction in (roughly) the same plane as the palm) at the carpometacarpal joint (CM). Finally, the wrist and elbow joints can act together to change the orientation of the hand. See figure 2.10 for a diagram of joint locations.

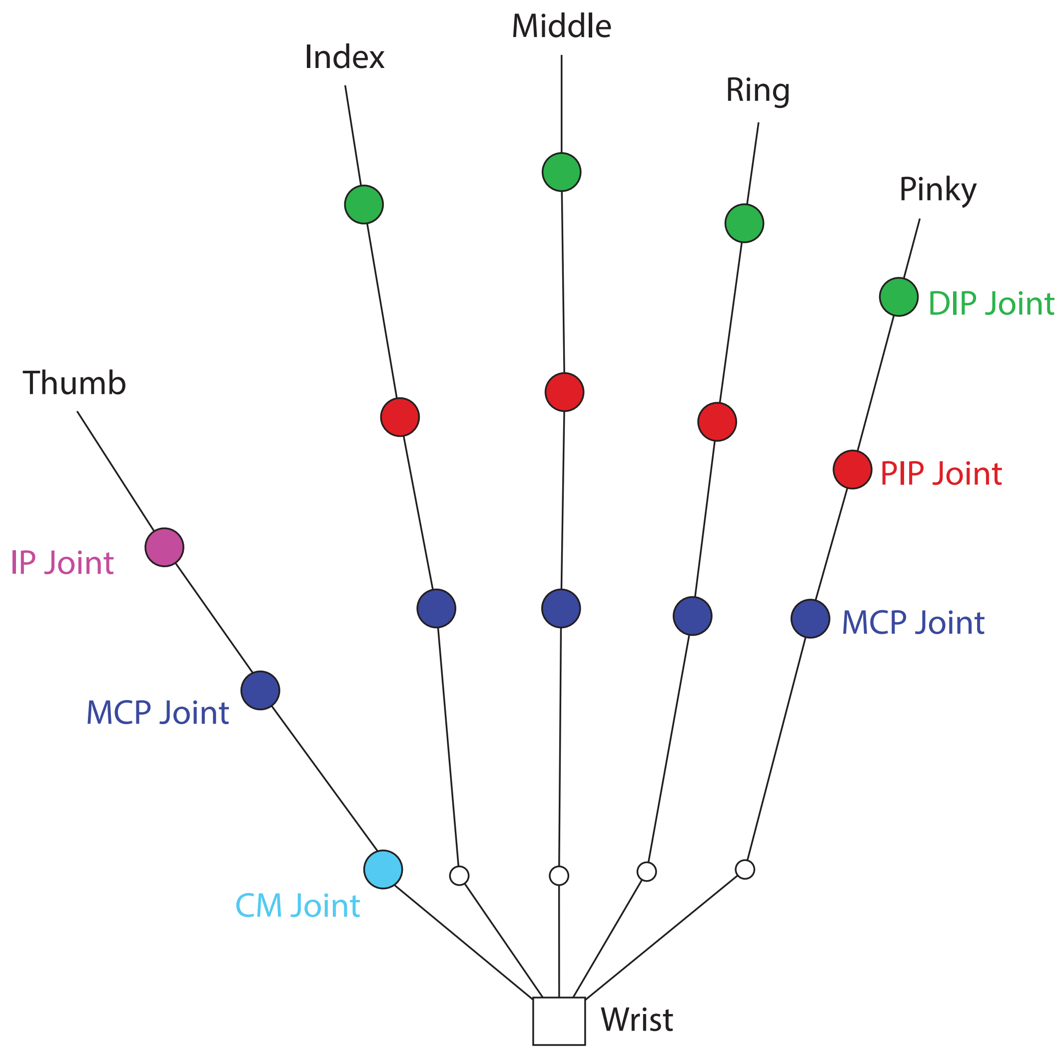


Figure 2.10: **Diagram of joints on the hand** The joint abbreviations are: DIP distal interphalangeal joint; PIP proximal interphalangeal joint; MCP metacarpophalangeal joint; IP interphalangeal joint (on the thumb only); and CM carpometacarpal joint

Most of the joints can be configured independently by combining muscle activity that extends or flexes each digit. There are some configurations which are either not physiologically possible, or are extremely difficult to articulate spontaneously (Ann, 1993). For example: fingers cannot be abducted and flexed at the MCP joint at the same time. Additionally there are tendencies of specific joints to assume the same configuration. For example, for the fingers, the DIP almost always assumes the same configuration as the PIP when the finger is not pressed against a rigid body (Whitworth (2011) showed that PIP flexion alone predicts 85% of DIP flexion, although she notes that this relationship is not absolute).

Although for spoken language all of the articulators conspire to form constrictions for each tract variable, the articulators of the hand do not all map easily onto a set of constrictions. Instead, because every joint is able to be defined in terms of flexion (the only exception being abduction, but for purposes here abduction can be thought to be the same as extension, with adduction being flexion), the tract variables for handshape ought to be specified in terms of flexion, rather than constriction.

For the vocal tract, tract variables are defined such that they are made up of at least one, but possibly multiple articulators; for example, Lip Aperture (LA) involves the upper lips, lower lips, and jaw. When looking at handshape what should constitute tract variables is a little bit more complicated. As many have noted, and as has been captured in many models of handshape, at least the MCP and PIP joints are able to be specified independently for each finger.[7] One possibility is that the MCP and PIP for each finger are independent tract variables. This is not optimal for two reasons: it fails to capture the generalization that fingers within each selected finger group all have the same configuration, and it does not account for the patterns seen when looking at certain pairs of fingers (i.e. the index and the middle fingers tend to assume the same or similar configurations, and the pinky and ring fingers tend to pattern together). For this reason we propose that the tract vari-

7. This is only mostly true. Some handshapes (e.g. full extension of all joints in the middle and ring fingers, with full flexion at all in the pinky and index fingers without the thumb holding down the pinky and index fingers) are hard (or impossible) to articulate, and are very rare cross-linguistically (Ann, 1993). But for our purposes these particular restrictions are not important.

ables for handshape are the MCP and the PIP configurations of each of the selected fingers groups. This allows different specifications for the MCP and PIP joints, but it constrains the number of different configurations of each to a maximum of two: one for the selected fingers, and one for the secondary selected fingers. Not only does this reduce the number of tract variables by (at least) half when compared to the every-joint-is-a-tract-variable model, it also captures the generalization that all selected fingers assume the same configuration. In other words, for any given handshape it is never the case that every finger can assume a different (phonologically contrastive) configuration; rather, there are a limited number of groups of configurations. This limited number of groups is precisely the selected/nonselected finger distinction. Within either of these groups, all of the fingers are configured in the same way (e.g. all of their MCP joints are extended, and all of their PIP and DIP joints are flexed). The consequence of this is that each tract variable is not explicitly associated with a specific set of articulators, but rather the articulators involved are determined by which fingers are selected or secondary selected. The values for each tract variable are given here as a range of angles, which are the possible values for the given joint. Most sign languages have 3 possible configurations that are (phonologically) contrastive: fully flexed (~90°), fully extended (~180°), and something in the middle (often called bent, ~135°) (Eccarius, 2008; Brentari & Eccarius, 2010). Future, detailed cross-linguistic studies will illuminate if these categories are exhaustive for all sign languages, or need to be altered.

For now, the secondary selected fingers group has the same set of tract variables as the selected fingers group. This might not be quite right, because the secondary selected fingers are only able to assume a small set of configurations (extended, flexed, and looped) (Eccarius, 2002), but for this work this question will not be critical. Given the variation observed, the secondary selected finger group should be no more complicated (i.e. contain more tract variables) than the selected finger group, although a simpler or possibly more abstract set of tract variables might be better than those given here. We set this particular issue aside for future work.

As with spoken language, the articulators that are in neither the selected nor secondary selected fingers groups (often called nonselected fingers) are considered to be in some way in- or less-active. These nonselected fingers do not assume a single default position as most articulators do in spoken language, but rather there is one of two options: either completely flexed or completely extended. Which configuration is chosen is generally predictable, where the nonselected fingers are extended if the selected fingers are (more) flexed, and the nonselected fingers are flexed if the selected fingers are (more) extended (van der Hulst, 1995; Brentari, 1998).[8] The idea that the nonselected articulators would oscillate between two extremes initially seems odd, but is intuitive if it is a method of easing the overall perception of handshape. The nonactive articulators assume a configuration that is maximally different from the active articulators to make the task of identifying which fingers belong to what group and then identifying what configuration the selected fingers are in easier. Although this perception argument has not been tested rigorously, it is appealing because it explains what looks idiosyncratic, as a regular process. Additionally it allows us to account for handshape with much of the same machinery that is used for spoken language. Table 2.1 shows all of the tract variables so far. The thumb has a separate variable for its carpometacarpal joint (CM). Figure 2.11 shows a visualization of the model for a 5 handshape. This joint is specified as a pair of angles because it has two degrees of freedom. Its MCP and DIP will be controlled by the *-MCP or *-PIP variable if it is a selected or secondary selected finger,[9].

8. The most cited example in ASL is when the index and the thumb are selected and not completely flexed or completely extended similar to the fingerspelled -F- handshape, used for signs like FIND and PRINT. The nonselected fingers are fully extended or fully flexed, respectively.

9. Remember, that the thumb only has a single interphalangeal joint, which is by convention described as the distal one (hence DIP).

group	joint	tract variable	values
selected fingers	MCP	SF-MCP	[flexed, mid, extended]
	PIP	SF-PIP	[flexed, mid, extended]
	MCP	SF-ABDUCTION	[abducted, adducted, negative abducted]
secondary selected fingers	MCP	SSF-MCP	[flexed, mid, extended]
	PIP	SSF-PIP	[flexed, mid, extended]
thumb opposition	CM	CM	[opposed, unopposed]
thumb abduction	CM	CM	[abducted, adducted, negative abducted]
nonselected fingers	all	NSF	[flexed, extended]

Table 2.1: **Tract variables for all fingers** The Articulatory Model of Handshape describes each handshape with a limited number of tract variables. The tract variable values are given as targets that are converted to joint-angles during motor planning. The phonetic realization of joint angles is continuous, although the phonology of any given sign language will divide that continuous range into targets of a small (circa 3) number of categories, labeled here as flexed, mid, and extended. Further cross-linguistic study is needed to see if 3 categories of flexion is sufficient for all contrasts.

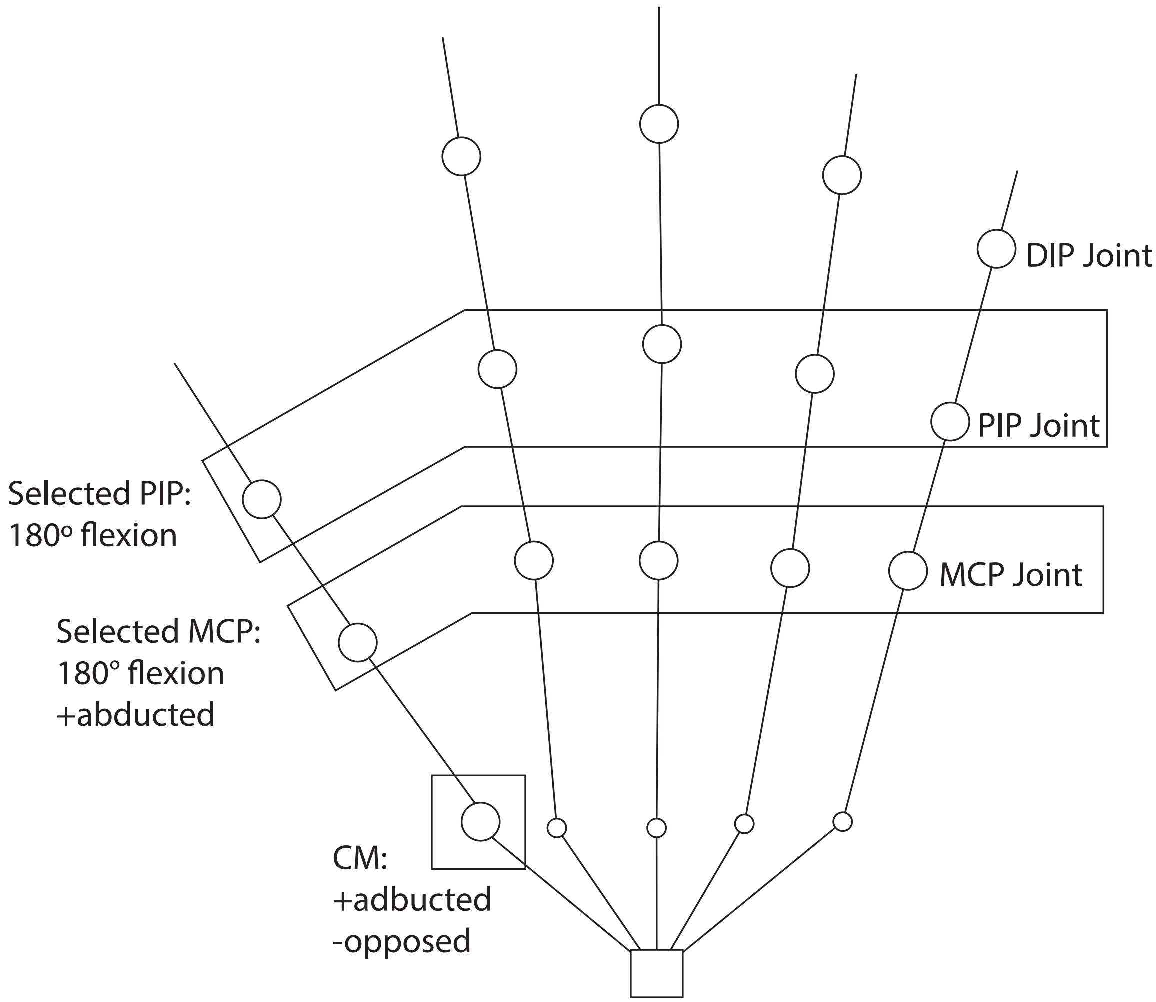


Figure 2.11: A visual representation of the Articulatory Model of Handshape for a 5 handshape.

Using this Articulatory Model of Handshape, with the predictions made by articulatory phonology generally (along with the small addition of nonactive gestures, described in section 2.3.2), we can make a few general hypotheses about handshape variation in ASL fingerspelling:

A. Because gestures are dynamic, individual handshapes and the articulators that make up the hand will not be static, sequential elements (i.e., discrete FS-letters[10]), but rather individual articulator gestures will overlap across several hand configurations (apogees).
B. The hand configuration of a specific instance of a given FS-letter will vary in predictable ways based on the surrounding context.

Chapter 4 shows a quantification of contextual handshape variation which confirms that variation exists and is context-dependent (confirming hypotheses A and B respectively). This leads to specific hypotheses about how this variation is constrained by phonetic and phonological representations of handshape[11]:

1. The nonselected (nonactive) fingers are more frequently the targets of coarticulatory pressure (vs. selected (active) fingers).

2. The selected fingers are the sources of coarticulatory pressure.

3. Finger configuration that is due to (phonetic) coarticulatory pressure will differ from configuration due to phonological specification.

Most of these hypotheses are in sync with what is commonly seen in research on coarticulation in spoken languages; however, research on fingerspelling with this perspective is new. Although everyone who has looked at fingerspelling has noted that the beads on a string model of static FS-letter after static FS-letter is not accurate, this research models the deviations from this, using articulatory phonology as a base to predict what should and should not constrain coarticulation in handshape.

10. As a reminder: *FS-letter* (short for fingerspelled-letter) is one of the 26 unique combinations of handshape and orientation that map on to the latin alphabet used for English.

11. These hypotheses are not meant to be inviolable or obligatory constraints, but rather theoretically grounded tendencies in the observed variation.

2.5 Implementation of the Articulatory Model of Handshape

Although there have been a number of models of the phonology of handshape (most also including other parameters) for signed languages (Battison, 1978; Mandel, 1981; Liddell & Johnson, 1989; Sandler, 1989; van der Hulst, 1995; Brentari, 1998; Eccarius, 2002; Sandler & Lillo-Martin, 2006) as well as one model of handshape phonetics (Johnson & Liddell, 2011a,b; Liddell & Johnson, 2011a,b), there has yet to be a concrete connection from phonological specification to phonetic implementation. The Articulatory Model of Handshape is exactly this connection. The Articulatory Model of Handshape is intended to account for, and be able to distinguish all of the possible phonological contrasts (for handshape) in any sign language (although, any individual sign language is not excepted to make all of the possible contrasts). The model then makes specific predictions for what the phonetic targets of each of these contrastive handshapes should look like. In addition to the theory as proposed in section 2.4, the Articulatory Model of Handshape has been implemented computationally as a Python module. This Python module additionally includes the ability to synthesize 3D renderings of handshapes from either phonological or phonetic specification. This is the first step in creating a model that can predict and describe the types of coarticulation that we hypothesize will occur above, and that we will investigate in chapter 4. The version of the software included in this work is v.0.1.0 (Keane, 2014) and is printed in appendix A. The source code (along with continued development) is available at `github.com/jonkeane/amohs`.

2.5.1 *From Prosodic Model specification to articulatory model to joint angles*

The AMOHS module uses a number of different custom classes[12]:

12. Throughout this section the terms frequently used to describe object oriented programming are being adopted. An introduction to object oriented programming is far beyond the scope of this dissertation, and is readily available from many sources (one example for Python being (Phillips, 2010)). For the discussion here, it is important to know that *classes* are templates for *objects* that all have the same (or similar) properties. *Objects* are individual instances of a class. A rough analogy (and one that is implemented in the AMOHS module) is that a class is the construct of handshape, and an object is a specific handshape. The construct of handshape has a set of features (selected fingers, joint configuration, etc.), but does not have specific values for any of those features. A specific handshape (like the 1-handshape) is a particular instance that has the set of features, with values specified for each.

pmHandshape[13] is a representation of handshapes that match that of the Prosodic Model from (Brentari, 1998; Eccarius, 2002). This class has the properties primary selected fingers (`SF`), secondary selected fingers (`SSF`), and nonselected fingers (`NSF`). `SF` and `SSF` further have the properties: abduction (`abd`) specifying the abduction for the group, fingers (`fing`) specifying the members of the group, joints (`joint`) specifying the configuration for both the base (MCP) and non-base (PIP and DIP) joints, thumb (`thumb`) specifying if the thumb is in the group, and opposition (`oppos`) specifying if the thumb is opposed. `NSF` only has one property: joints (`joint`) which specifies if the nonselected fingers are flexed or extended.

Objects of this class can be generated from strings of handshape codes as described in (Eccarius & Brentari, 2008). A few modifications to the notation are accepted: the secondary selected fingers accept the same range of values as the primary selected fingers.[14] Each character of an input notation string is placed in the appropriate property in an object of the `pmHandshape` class. Where the absence of a character in the prosodic notation string is meaningful (for example, when the selected (or secondary selected) fingers are extended there is simply no joint configuration character), the feature is filled in with `None`.

The articulatory model specification for handshapes is represented in the **handshape** class. Objects of `handshape` class can be created by specifying selected fingers (`selectedFingers`), secondary selected fingers (`secondarySelectedFingers`), thumb specifications (`thumb`), and nonselected fingers (`nonSelectedFingers`) directly. These properties are similar to those for `pmHandshapes`, although instead of using characters from the prosodic notation system to represent configurations, they use full character strings to represent categorical contrasts of handshape configurations (as opposed to the short single-character notation that the prosodic notation uses). Additionally, no

13. For ease of reference, we use the class name as it is used in the code to reference the class.

14. As will be discussed later, a few additions to the secondary selected fingers group were necessary. As a first step for that, the full range of possible values for primary selected fingers were ported over. This is likely to be more values than are attested in the world's languages, but detailed cross-linguistic study is needed to know what the limits are.

features are left unspecified to be filled in later by redundancy rules (e.g. all joint specifications have value).

Selected fingers (`selectedFingers`) and secondary selected fingers (`secondarySelectedFingers`) have the following features: members (`members`), which is a list of which digits are members of the group (i.e. index, middle, pinky, ring, thumb); configuration for the MCP joint (`MCP`), which is extended (`ext`), mid-flexed[15] (`mid`), and flexed (`flex`); and abduction (`abd`) which is either abducted (`abducted`), adducted (`adducted`), or negatively abducted (`negativeAbducted`).

It should be noted that the decomposition of overall finger configuration (e.g. extended (AKA fully open), closed, bent-closed) into configurations for each joint (e.g. MCP: ext PIP: ext (for extended), MCP: flex PIP: flex (for closed), MCP: ext PIP: flex (for bent-closed)) is meant to make all of the same contrasts as the Prosodic Model (Brentari, 1998)[16]. The 9 categories here, again, are intended to be the 9 possible phonological contrasts of selected fingers configurations across all sign languages. Any individual sign language might not (and probably will not) have all of the 9 possible categories. Although Eccarius (2008); Eccarius & Brentari (2008) found that at least 8 categories are needed to account for some sign languages, other models of handshape phonology have fewer categories: Sandler (1989) has 4 extension categories (5 if spread is included), van der Hulst (1995) cites at least 7 categories; "There is a certain consensus that at least the configurations in (10) must be recognized as being potentially distinctive." and Van der Kooij (2002) proposed that there are is no phonological contrast for flexion at the MCP joint, rather differences are for articulatory reasons (due to orientation), or iconically motivated. For the PIP and DIP (non-base) joints, they propose three categories of flexion like those found in the Articulatory Model of Handshape.

For now, the translation between Prosodic Model finger configurations and Articulatory Model features is accomplished with a lookup table (from the psf, ssf, and nsf columns to the base (MCP) and

15. This could also be called mid-extended, but based on the Prosodic Model's assumption of extended being underlying, mid-flexed is used here.

16. Although given three features in two slots, there are 9 combinatorial possibilities, where as the Prosodic Model has 7 configuration groups, with one group later being split into two for a total of 8 contrastive configurations (Eccarius & Brentari, 2008). We have added the last possibility in table 2.4

nonbase (PIP)[17] columns in the table 2.4). It is possible that one could translate from the Prosodic Model features to the Articulatory Model features in a more principled way, but that is beyond the scope of this work. Although table 2.4 shows all of the relationships between Prosodic Model configurations and Articulatory Model features, we have reproduced most of the details of the Prosodic Model (Brentari, 1998, pp.107), as well as their Articulatory Model equivalents, below.

Here, each configuration along with the structure proposed in the Prosodic Model has been paired with the joint configurations from the Articulatory model. By and large, the Prosodic Model structures fit neatly with the Articulatory Model features: when the joints are not present in the structure they are always extended in the Articulatory Model features. When the joints are present in the structure, but there is no [flexed] in the structure, they are almost always mid in the Articulatory Model features. When the joints are present in the structure, and there is [flexed] in the structure, they are almost always flexed in the Articulatory Model features.

The last three (h, i, and j) were not found in (Brentari, 1998, pp.107). The first two of these represent logical possibilities of the structure proposed in the Prosodic Model. We tentatively propose that one (h) is the phonological specification of the curved open (wide) handshape that was described by Eccarius & Brentari (2008), and what was traditionally described just as curved open (b) is actually the phonological specification for curved open (narrow). This leaves just one structure that has not been previously described (i), and one set of Articulatory Model specifications (MCP: mid, PIP: flex) left to complete the combinatorial possibilities of both systems. Unfortunately, the last possible Prosodic Model structure (i) does not bear much similarity to the final possible Articulatory Model feature set (j). Future work is needed to investigate the relationship between the Prosodic Model structures proposed, and the features of the Articulatory Model. Going forward in this work, the Prosodic Model configurations are linked to the Articulatory Model features through a look up table (table 2.4), and through this, they can be considered to be near notational variants of each other.

17. Although both the PIP and the DIP are nonbase joints, we will sometimes refer to nonbase joints as PIP, this is out of expedience, and not making the claim that the DIPs are not a part of nonbase joints.

a. extended (fully open)

PM code: ∅

PM structure:

SF

AMOHS MCP: ext

AMOHS PIP: ext

b. curved open

possibly only (narrow)

PM code: c

PM structure:

SF

joints

base nonbase

AMOHS MCP: mid

AMOHS PIP: mid

c. curved closed

PM code: o

PM structure:

SF

joints

[flexed]

base nonbase

AMOHS MCP: flex

AMOHS PIP: mid

d. flat open

PM code: <

PM structure:

SF

joints

base

AMOHS MCP: mid

AMOHS PIP: ext

e. flat closed

PM code: >

PM structure:

SF

joints

[flexed]

base

AMOHS MCP: flex

AMOHS PIP: ext

f. bent closed

PM code: [

PM structure:

SF

joints

[flexed]

nonbase

AMOHS MCP: ext

AMOHS PIP: flex

g. (fully) closed

PM code: @

PM structure:

SF
|
joints
[flexed]

AMOHS MCP: flex

AMOHS PIP: flex

h. (not found in the PM)

possibly curved open (wide)

PM code: (

PM structure:

SF
|
joints
|
nonbase

AMOHS MCP: ext

AMOHS PIP: mid

i. (not found in the PM)

PM code: ?

PM structure:

SF
|
joints

AMOHS MCP: ?

AMOHS PIP: ?

j. (not found in the PM)

PM code: e

PM structure: ?

AMOHS MCP: mid

AMOHS PIP: flex

Nonselected fingers (`nonSelectedFingers`) have only two properties: members (`members`) like selected and secondary selected fingers, and joints (`joints`) for all of the joints, which is either extended (`ext`) or flexed (`flex`).

Finally, the thumb (`thumb`) has a single property opposition (`opposition`) which determines if the thumb is rotated such that it has palmar abduction[18]. There are only two possible values unopposed (`unopposed`) with no palmar abduction (radial abduction[19] will be determined by the abduction value for the selected finger group that the thumb is associated with) and opposed (`opposed`)

18. Palmar abduction is abduction perpendicular to the palmar plane

19. Radial abduction contrasts with palmar abduction as being abduction in the same plane as the palm (and thus parallel with it).

with full palmar abduction if the thumb is in a selected finger group with abduction, and only slight palmar abduction if the thumb is in a selected finger group with adduction.

Additionally, objects of `pmHandshape` class can be automatically converted to `handshape` classes with the function `toAMhandshape`. This function takes the strings that are in the properties of `pmHandshape` and translates each into the appropriate value for the articulatory model representation (an object of class `handshape`). This includes features that have type `None` which resulted from parts of the prosodic notation string that are unwritten (e.g. if there is no joint specification the joints are extended). Tables that map the parsed strings to the features are given below (tables 2.2–2.4).

base symbol	fingers
B	index, middle, ring, pinky
M	index, middle, ring
D	middle, ring, pinky
U	index, middle
H	index, pinky
A	middle, ring
P	middle, pinky
2	ring, pinky
1	index
8	middle
7	ring
J	pinky
T	thumb

Table 2.2: **Finger membership symbols** The prosodic notation system (and properties of the `pmHandshape` class) uses the base symbol column, the articulatory model (and properties of the `handshape` class) uses the fingers column.

PSF	joint config	pmfeatures	abd
empty	adducted		adducted
x	crossed	cross	negativeAbducted
k	stacked	stack	adducted?
^	spread	spread	abducted

Table 2.3: **Abduction symbols** Note that stacked is not currently implemented in AMOHS. The prosodic notation system (and properties of the `pmHandshape` class) uses the base PSF column (this column is called PSF because it is used in place of the joint configurations used in the joint symbols table (table 2.4), where the base and nonbase joints are always extended), the articulatory model (and properties of the `handshape` class) uses the abd column.

PSF	SSF	NSF	joint config	pmfeatures	base (MCP)	nonbase (PIP)
empty	empty	/	extended		ext	ext
c	c	None	curved-open (narrow)	nonbase+base	mid	mid
(	(	None	curved-open(wide)	nonbase+base	ext	mid
o	o	None	curved-closed	[flex]+nonbase+base	flex	mid
<	<	None	flat-open	base	mid	ext
>	>	None	flat-closed	[flex]+base	flex	ext
[	[	None	bent	[flex]+nonbase	ext	flex
@	@	#	closed	[flex]	flex	flex
e	e	None	?	?	mid	flex

Table 2.4: **Joint symbols** As discussed by Eccarius & Brentari (2008), the difference between `c` and `(` configurations were not represented in the Prosodic Model, which only has a single extended feature (with flexed being the underlying specification). Here, we have proposed that the difference between them can be made if we propose that this contrast can be made by using three levels of extension: flexed, mid, and extended that are associated with both the base (MCP) and nonbase (PIP and DIP) joints. With this additional level, there is one additional symbol we have added (e) that was not already documented, but exhausts the combinatorial possibilities of three features across two joints (and, as we will see later, is needed for a few handshapes). The prosodic notation system (and properties of the `pmHandshape` class) use the psf, ssf, and nsf columns (depending on which selection category the joints are being specified for), the articulatory model (and properties of the `handshape` class) use the base (MCP) and nonbase (PIP) columns. The pmfeatures column comes from (Eccarius & Brentari, 2008, pp.81, table 2), and is intended as a short-hand for describing the categories of finger configuration that come from the Prosodic Model (Brentari, 1998, pp.107).

Objects of the `handshape` class can further be translated from their form which represents (articulatory) phonological specification to a form which represents phonetic targets for each of the joints in the hand. This phonetic target of hand configuration is represented in the handconfiguration class. In this class, each digit is its own property, and each has properties for each joint that exists on it. For the index, middle, ring, and pinky fingers those joints are the metacarpophalangeal (MCP) joint (`MCP`), proximal interphalangeal (PIP) joint (`PIP`), and distal interphalangeal (DIP) joint (`DIP`). Each of these has at least one degree of freedom which is mapped from flexion-extension. Additionally, the MCP has one more degree of freedom which is mapped from abduction-adduction. The thumb has a carpometacarpal (CM) joint (`CM`), metacarpophalangeal (MCP) joint (`MCP`), and (a single) interphalangeal (IP) joint (`IP`). The CM joint has three degrees of freedom: flexion-extension, palmar abduction-adduction, and radial abduction-adduction. Both the MCP and the IP joints have a single degree of freedom: flexion-extension.

The function `toHandconfigTarget` converts from `handshape` class to `handconfiguration` class. To do this, it generates a joint angle for each joint on the hand from the selected finger group the finger the joint is on is part of, as well as the its flexion-extension value, and if applicable the abduction and opposition values (see tables 2.5–2.7 for feature to value conversions). In this implementation the distal interphalangeal (DIP) joint target value is copied from the PIP. Although this is approximately correct in most cases (Whitworth (2011) showed that PIP flexion alone predicts 85% of DIP flexion), it should be refined in the future. The thumb IP takes on the same value as the PIP joint in the selected finger group that the thumb is part of. This is also only approximately correct: there seems to be a tendency for the thumb's IP joint to be extended if any of the finger's DIP joints are extended, although this particular relationship needs more investigation, which is set aside for future work.

feature	joint angle target
ext	180°
mid	135°
flex	90°

Table 2.5: **Flexion-extension features and joint targets** Note that the joint angle targets are the joint angles with respect to the bone that is immediately proximate on the body, so for the PIP joint, that is the angular difference of the intermediate phalanx with respect to the proximate phalanx.

	joint angle target			
feature	index	middle	ring	pinky
abducted	20°	0°	-10°	-20°
neutral abducted	10°	5°	-5°	-10°
adducted	0°	0°	0°	0°
negative abducted (crossed)	-10°	10°	10°?	10°?

Table 2.6: **Abduction-adduction features and joint targets for fingers** Note that each finger has different values for each feature because this specifies the relationship of the proximal phalanx to the metacarpal (via the MCP joint). The negative abduction values for the ring and pinky fingers need refining, although not used for fingerspelling handshapes (or any handshapes documented for ASL in general), and so are outside of the scope of this dissertation.

feature		joint angle target for CM		
abduction	opposition	flexion	abduction	rotation
abducted	opposed	NA	NA	NA
abducted	unopposed	15°	27°	9°
adducted	opposed	-22°	13°	-27°
adducted	unopposed	23°	8°	0°
negative abducted	opposed	-34°	-24°	-53°
negative abducted	unopposed	NA	NA	NA

Table 2.7: **Abduction-adduction as well as opposition features and joint targets for the thumb** The angles here are not specified based on the relation to the previous joint, but rather, were tuned using the rendering functions of `libHand` which will be discussed later. Additionally, cells marked as NA are those that are considered physiologically impossible.

Objects that are of the `handconfiguration` class can also be compared using a subtraction method, which gives you an object of the **`armconfigurationDelta`** class which represents the difference between between the two handshapes that are being compared. This class has the same features as the `handconfiguration` class, but the properties are the differences between the joint angles for each joint on the hand. Additionally, the differences can be weighted based on how proximate the joints are (this is to represent the intuition that more proximal joints will be more visually salient.) For now, the weights simply increase by one as the joints are more proximal: (DIPs and IPs have a weight of 1, PIPs have a weight of 2, finally MCPs and CMs have a weight of 3). More work is needed on the perception of handshape contrasts to map these weights onto weights that represent actual perceptual differences between more or less proximal joints.

Although not strictly part of handshape, the wrist (and arm) contribute to different orientations that are used contrastively in fingerspelling (as well as in the rest of ASL). To account for this, the beginnings of whole arm specifications have been added at both the phonetic level and the phonological (at least via the articulatory model[20]). The articulatory model representation of the arm is represented by the `arm` class. This has the properties of a handshape (which is an object of the `handshape` class, described above), and an orientation. The phonetic level representation is represented by the `armconfiguration` class, like the phonological representation of the arm, it takes an object representing the handshape (of the `handconfiuration` class), as well as a property representing the wrist. The wrist joint has three degrees of freedom: flexion-extension, rotation (AKA ulnar and radial flexion), and pronation-supination. Of course, the wrist joint itself does not actually have all three of these degrees of freedom, pronation-supination is a property that is controlled by the elbow and wrist together, but for the present purposes combining it with the wrist is fine. The mapping from orientation to wrist angles is given in table 2.8. Like the `handconfiguration` class, objects of the `armconfiguration` class can be compared via subtraction (where the handconfigurations are compared along with the wrist configurations). The weight for the wrist is 4.

20. Of course, the Prosodic Model can also represent orientation, however, it was not included in the prosodic notation system from (Eccarius & Brentari, 2008)

	joint angle target		
orientation	flexion	rotation	pronation
neutral	0°	0°	0°
default fingerspelling	-10°	0°	0°
palm in	-75°	0°	80°
palm down	-75°	0°	0°

Table 2.8: **Palm orientation features and joint targets for the wrist** Note that rotation is also called radial-ulnar flexion.

2.5.2 *Visualizing handshapes using the articulatory model*

Once we have joint angle targets in the form of objects of the `handconfiguration` class[21], visualizing the hand configuration is just a matter of that appropriate anatomical model, and existing 3-dimensional rendering software. `LibHand` is an open source library that renders hands based on joint angles (Šarić, 2011)[22]. Using this library, we developed a simple C++ binary `imageGen`[23], that can be run on the command line, and accepts as arguments: a scene specification file[24], a pose file (where the joint angles are specified), and an output image file where an image of the hand is written. `ImageGen` in its current form is simple and only outputs a single view. It is possible to render the hand from multiple views, or even in an interactive or video form using additional features of `LibHand`. For the purposes of checking if the articulatory model features were approximately correct, we only needed the one view. An example rendered hand is given in figure 2.12 below.

21. Or, more accurately the `armconfiguration` class including information about the wrist.

22. In order for `LibHand` to compile on recent versions of OS X, some changes needed to be made in the source, as well as in satisfying dependencies. The updated source code can be downloaded from github.com/jonkeane/libhand. In order to compile it on OS X 10.9 and later, one must install `boost`, `opencv`, and `ogre` which are readily available from package managers including homebrew (brew.sh). A patched version of `ogre` compatible with `LibHand` can be found at github.com/jonkeane/homebrew-libhand

23. The source is available at github.com/jonkeane/imagegen

24. A scene specification file is included within the AMOHS module and is used as the default.

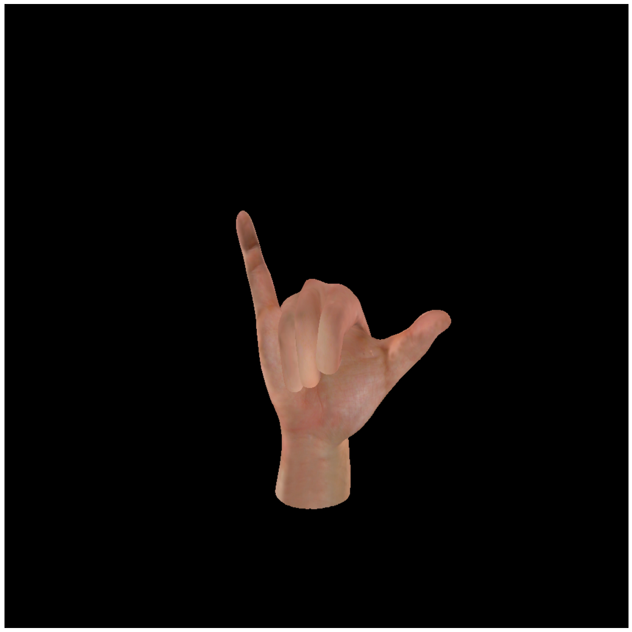


Figure 2.12: A rendering of the hand for the -Y- handshape

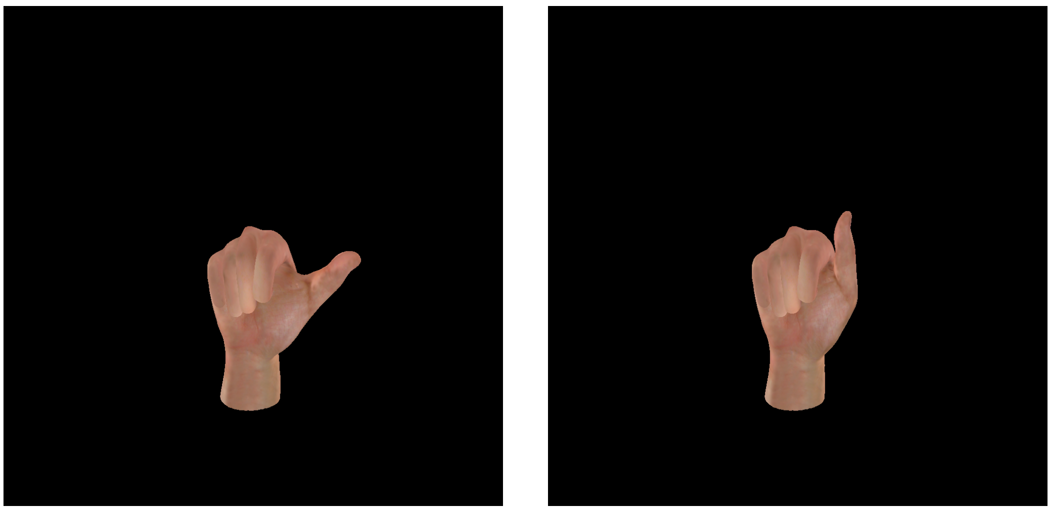
2.5.3 *Fingerspelled letter forms, using the articulatory model*

Now that we have a full understanding of the Articulatory Model, and its relationship with the Prosodic Model, we can explore phonological specifications for the handshapes used in ASL fingerspelling. Table 2.9 has full specifications for articulatory model features as well as Prosodic Model notations. The Prosodic Model codes were taken from (Eccarius & Brentari, 2008). The images produced by the `amohs` module are meant to be examples of hypercanonical handshapes, that is, those that are not in any way affected by contextual, signer, dialectal, etc. variation. In a few cases the handshapes that were rendered from the Prosodic Model codes did not match the expected handshape for a canonical version of that letter. In those cases, new articulatory model representations (and then Prosodic Model notations) were generated (and refined through visualizations) that better conformed to the expected canonical handshapes. The old PM notation letters are labeled with *traditional* following the letter for easy comparison. Further, some letters defy easy decision of a canonical form; in these cases multiple feature sets (and PM codes) have been given. Overall, the visualizations are strikingly close to the expected canonical forms based solely on the simple set of joint angle target mappings discussed in section 2.5.1. Additionally, although some corrections to phonological specifications are necessary for some FS-letters, the majority of them work out fine: going directly from phonological specification, to a synthesized canonical handshape that looks remarkably like what is observed in fingerspelling.

letter	psf-members	psf-mcp	psf-pip	psf-abd	ssf-members	ssf-mcp	ssf-pip	ssf-abd	t-oppos	nsf-members	nsf-joints	orientation	pmCode
-A- traditional	i,m,r,p	flex	flex	add	None	None	None	None	None	t	ext	defaultFS	B@;/
-A-	i,m,r,p	flex	flex	add	t	mid	ext	add	unopposed	None	None	defaultFS	B@;T-<
-B- traditional	i,m,r,p	ext	ext	add	None	None	None	None	None	t	flex	defaultFS	B;#
-B-	i,m,r,p	ext	ext	add	t	mid	ext	neg abd	opposed	None	None	defaultFS	B;Tx<
-C- traditional	i,m,r,p,t	mid	mid	add	None	None	None	None	opposed	None	None	defaultFS	BTc
-C-	i,m,r,p,t	ext	mid	add	None	None	None	None	opposed	None	None	defaultFS	BT(
-D- big	m,r,p,t	flex	mid	add	None	None	None	None	opposed	i	ext	defaultFS	DTo;/
-D- traditional	m,t	flex	mid	add	i	ext	ext	add	opposed	r,p	flex	defaultFS	8To;1;#
-D-	m,t	mid	mid	add	i	ext	ext	add	opposed	r,p	flex	defaultFS	8Tc;1;#
-E- open	i,m,r,p,t	ext	flex	add	None	None	None	None	opposed	None	None	defaultFS	BT[
-E-	i,m,r,p,t	mid	flex	add	None	None	None	None	opposed	None	None	defaultFS	BTe
-E- traditional	i,m,r,p	ext	flex	add	None	None	None	None	None	t	flex	defaultFS	B[;#
-E- alternate	i,m,r,p	mid	flex	add	None	None	None	None	None	t	flex	defaultFS	Be;#
-F- traditional	i,t	flex	mid	add	None	None	None	None	opposed	m,r,p	ext	defaultFS	1To;/
-F-	i,t	mid	mid	add	None	None	None	None	opposed	m,r,p	ext	defaultFS	1Tc;/
-G- traditional	i	ext	ext	add	t	ext	ext	add	opposed	m,r,p	flex	palmIn	1;T;#
-G-	i,t	ext	ext	add	None	None	None	None	opposed	m,r,p	flex	palmIn	1T;#
-G- thumb	i,t	ext	ext	add	None	None	None	None	unopposed	m,r,p	flex	palmIn	1T-;#
-H-	i,m	ext	ext	add	None	None	None	None	None	r,p,t	flex	palmIn	U;#
-I-	p	ext	ext	add	None	None	None	None	None	i,m,r,t	flex	defaultFS	J;#
-J-	p	ext	ext	add	None	None	None	None	None	i,m,r,t	flex	defaultFS	J;#
-K-	i	ext	ext	add	m	flex	ext	add	None	r,p,t	flex	defaultFS	1;8>;#
-L- traditional	i,t	ext	ext	add	None	None	None	None	unopposed	m,r,p	flex	defaultFS	1T-;#
-L-	i	ext	ext	add	t	ext	ext	abd	unopposed	m,r,p	flex	defaultFS	1;T-^;#
-M- closed	i,m,r	flex	flex	add	None	None	None	None	None	p,t	flex	defaultFS	M@;#
-M-	i,m,r	flex	ext	add	None	None	None	None	None	p,t	flex	defaultFS	M>;#
-N- closed	i,m	flex	flex	add	None	None	None	None	None	r,p,t	flex	defaultFS	U@;#
-N-	i,m	flex	ext	add	None	None	None	None	None	r,p,t	flex	defaultFS	U>;#
-O- traditional	i,m,r,p,t	flex	mid	add	None	None	None	None	opposed	None	None	defaultFS	BTo
-O-	i,m,r,p,t	mid	mid	add	None	None	None	None	opposed	None	None	defaultFS	BTc
-P-	i	ext	ext	add	m	flex	ext	add	None	r,p,t	flex	palmDown	1;8>;#
-Q-	i,t	ext	ext	add	None	None	None	None	opposed	m,r,p	flex	palmDown	1T;#
-R-	i,m	ext	ext	neg abd	None	None	None	None	None	r,p,t	flex	defaultFS	Ux;#
-S-	i,m,r,p,t	flex	flex	add	None	None	None	None	opposed	None	None	defaultFS	BT@
-T- traditional	i,t	flex	flex	neg abd	None	None	None	None	opposed	m,r,p	flex	defaultFS	1Tx@;#
-T-	i	flex	mid	add	None	None	None	None	None	m,r,p,t	flex	defaultFS	1o;#
-U-	i,m	ext	ext	add	None	None	None	None	None	r,p,t	flex	defaultFS	U;#
-V-	i,m	ext	ext	abd	None	None	None	None	None	r,p,t	flex	defaultFS	U^;#
-W-	i,m,r	ext	ext	abd	None	None	None	None	None	p,t	flex	defaultFS	M^;#
-X-	i	ext	flex	add	None	None	None	None	None	m,r,p,t	flex	defaultFS	1[;#
-Y-	p,t	ext	ext	abd	None	None	None	None	unopposed	i,m,r	flex	defaultFS	JT-^;#
-Z-	i	ext	ext	add	None	None	None	None	None	m,r,p,t	flex	defaultFS	1;#

Table 2.9: **Full articulatory model feature specifications for all fingerspelled letters** Abbreviations in the members column are i for index, m for middle, r for ring, p for pinky, t for thumb. traditional variants are those that were identified based on the photos in (Eccarius & Brentari, 2008), but when rendered with the amohs software did not look like the canonical forms.

The traditional notation for -A- has the thumb in the nonselected fingers group, and the nonselected fingers extended. This is problematic because, as can be seen in figure 2.13, the MCP on the thumb should be at most mid-extneded and not fully extended. For this reason, the new proposed -A- has the thumb in the secondary selected fingers group, with the joint configuration mid-ext for base and non-base joints respectively). Additionally, the thumb is unopposed. See table 2.9 for full specifications.



(a) -A- traditional (b) -A-

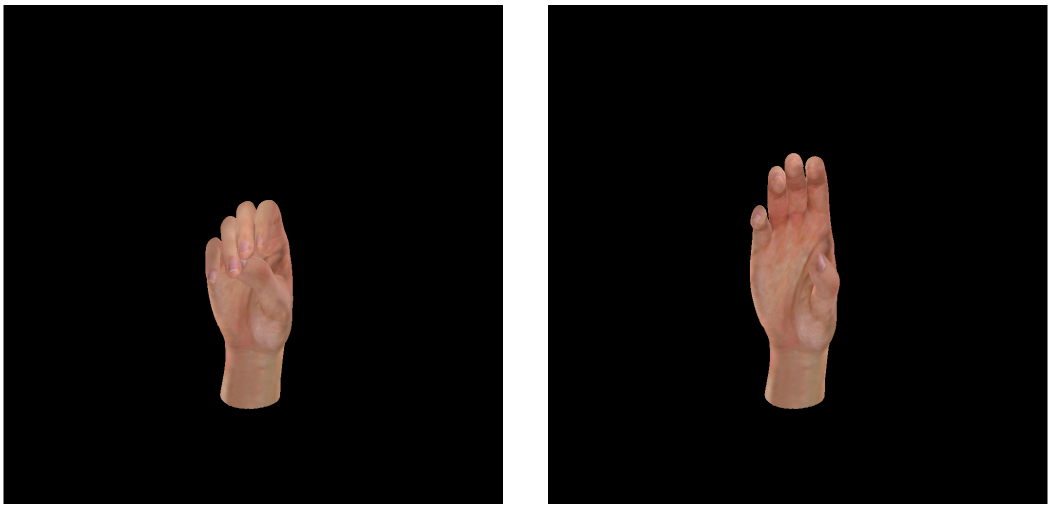
Figure 2.13: Traditional versus new -A-

The traditional notation for -B- has the thumb in the nonselected fingers group, and the non-selected fingers flexed. This is problematic because, as can be seen in figure 2.14, the IP on the thumb should be fully extended. For this reason, the new proposed -B- has the thumb in the secondary selected fingers group, with the joint configuration mid-ext for base and non-base joints respectively). Additionally, to account for the full radial (negative) abduction the thumb has the crossed feature. See table 2.9 for full specifications.

(a) -B- traditional (b) -B-

Figure 2.14: Traditional versus new -B-

The traditional notation for -C- has the joint configuration mid-mid for base and non-base joints respectively. This is problematic because, as can be seen in figure 2.15, the fingers are too closed. For this reason, the new proposed -C- has the joint configuration ext-mid for base and non-base joints respectively). See table 2.9 for full specifications.



(a) -C- traditional   (b) -C-

Figure 2.15: Traditional versus new -C-

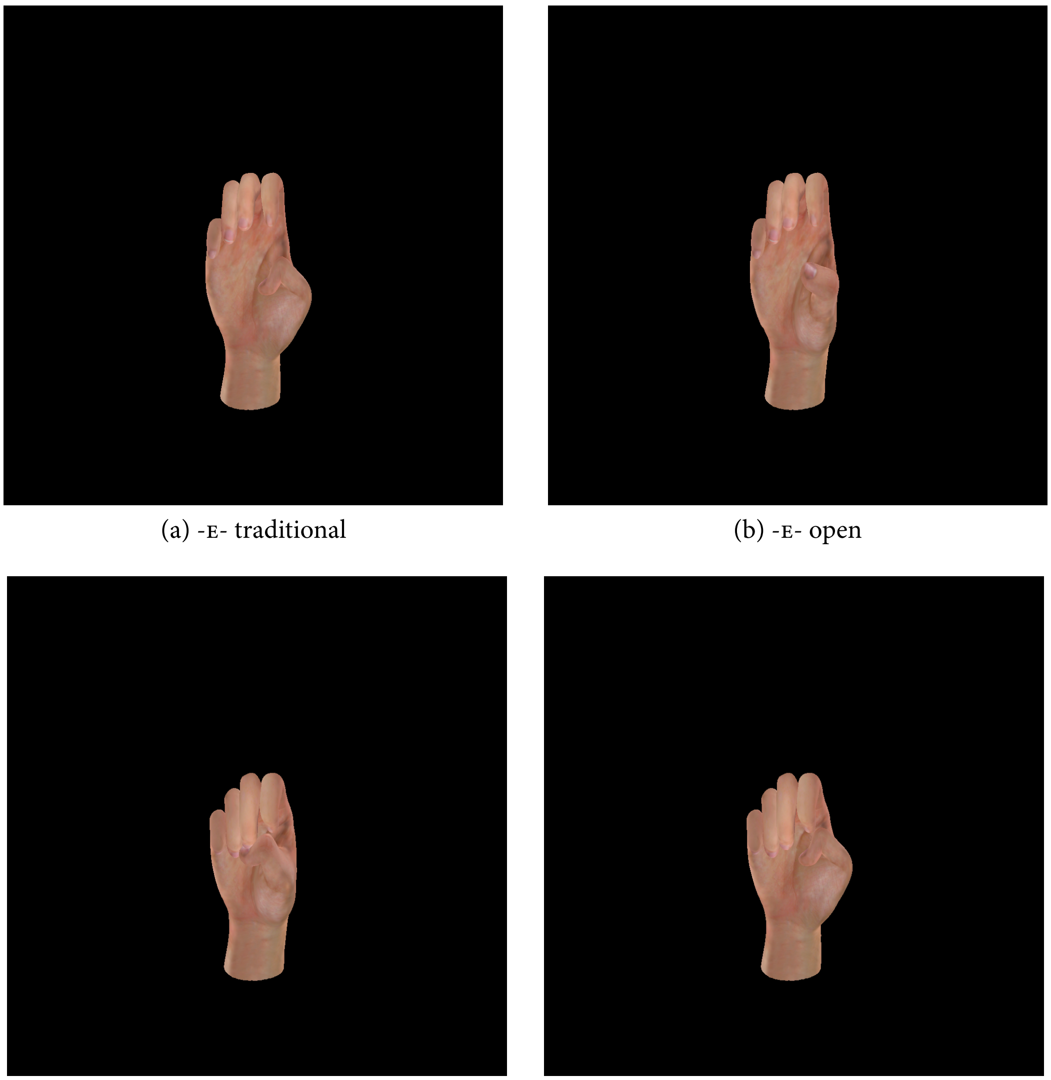
In other work, it has been shown that of the two variants of -D- (big-D-, where the index is extended, and the middle, ring, pinky, and thumb are all half flexed to form a ring versus little-D- that is identical to big-D-, except that the ring and pinky fingers are fully flexed) it is little-D- that is used nearly 100% of the time (Keane *et al.*, 2012a). For this reason we will only discuss the little-D- variant, and when we refer to the FS-letter -D-, that is the variant we are referring to. The traditional notation for -D- has the primary selected finger joint configuration flex-mid for base and non-base joints respectively. This is problematic because, as can be seen in figure 2.16, the fingers are too closed. For this reason, the new proposed -D- has the primary selected finger joint configuration mid-mid for base and non-base joints respectively). See table 2.9 for full specifications. The index-thumb contact is not exactly perfect — although it is quite good — additional refining of joint angles to allow for true contact is needed in future work.

(a) -D- traditional

(b) -D-

Figure 2.16: Traditional versus new -D-

The handshape -E- is controversial. There are two variants that show up in normal fingerspelling: a closed variant where the tips of the fingers touch the thumb, and an open variant. Although there are prescriptive rules proscribing the open variant, it is found in the fingerspelling of native Deaf signers (Keane *et al.*, 2013). For this reason, forms for both are proposed (see figure 2.17). The traditional notation for -E- has the thumb in the nonselected fingers group, and the joint configuration in the selected fingers group that are those of the open variant. This is similar to one variant that is seen in fingerspelling. Additionally, a more claw-like open variant is seen, labeled as -E- open, where the thumb is in the selected finger group. The closed variant, like the open variant, has the thumb in the primary selected fingers group, with the joint configuration of mid-flex for the base and non-base joints respectively. An alternate possibility (which is also seen in fingerspelling), has the thumb in the nonselected fingers group, and flexed. Because handshapes close to all of these have been observed, we are not choosing one as the only canonical form, further study is needed to see how each behaves in context. Additionally, the question of whether the variation observed in this handshape is phonological in nature (where the handshapes have different underlying phonological specifications) or if the variation is subphonemic is an open one. Either is possible, and either could be implemented in a model like the articulatory model here. See table 2.9 for full specifications.



(a) -E- traditional

(b) -E- open

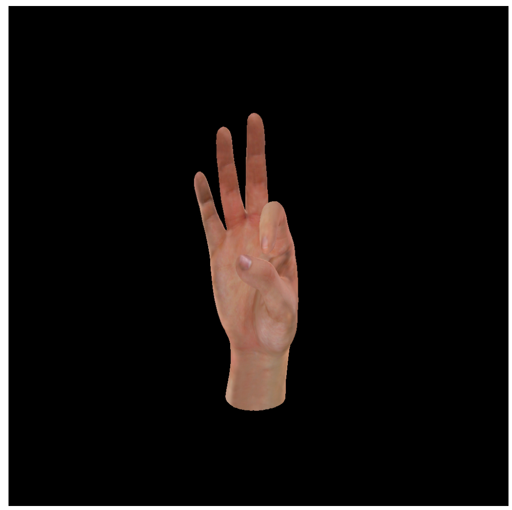
(c) -E- closed

(d) -E- closed alternate

Figure 2.17: Traditional versus new, open versus closed -E-

The traditional notation for -F- has the primary selected finger joint configuration flex-mid for base and non-base joints respectively. This is problematic because, as can be seen in figure 2.18, the fingers are too closed. For this reason, the new proposed -F- has the primary selected finger joint configuration mid-mid for base and non-base joints respectively). See table 2.9 for full specifications. Again, the index-thumb contact is not exactly perfect, although is quite good, additional refining of joint angles to allow for true contact is needed in future work.

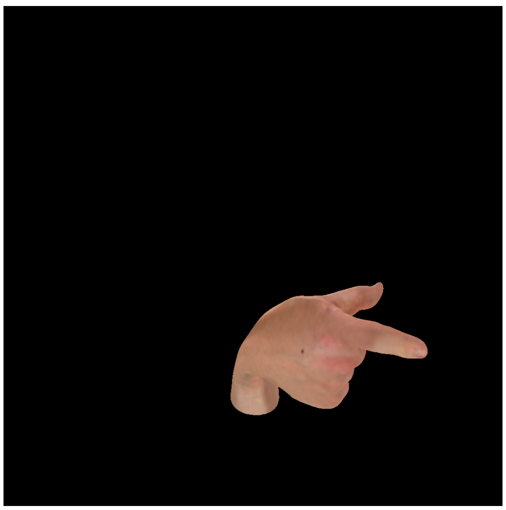
(a) -F- traditional



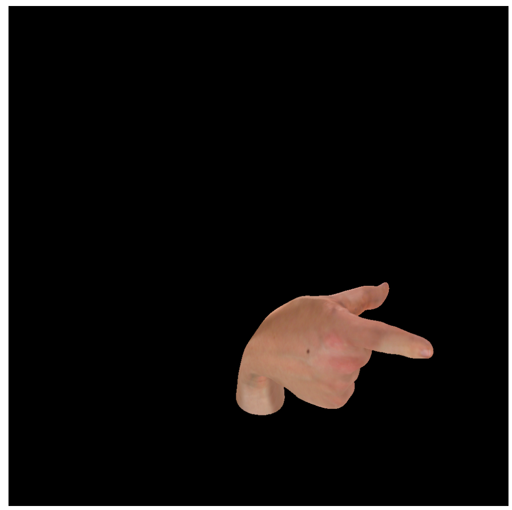
(b) -F-

Figure 2.18: Traditional versus new -F-

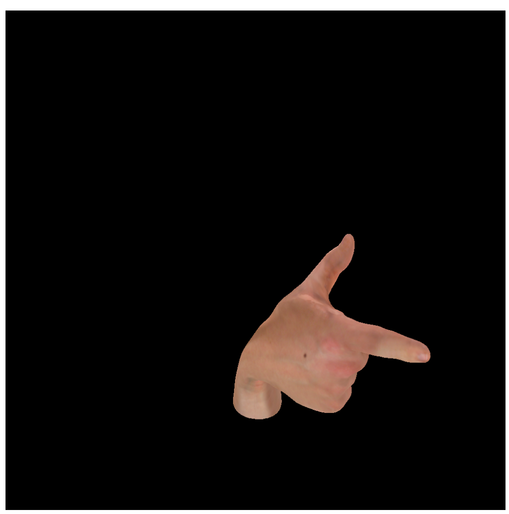
The traditional notation for -G- has the thumb in a secondary selected position. This is nearly identical with the thumb in the primary selected fingers group (see figure 2.19). Additionally, a variant with the thumb (radially) abducted has been observed, this is labeled -G- thumb. More work is needed to determine the distributional properties of these variants. Although this variation has been observed since the early days of sign language research, Stokoe's notation even used G to label all handshapes that had the index and only the index extended. See table 2.9 for full specifications.



(a) -G- traditional



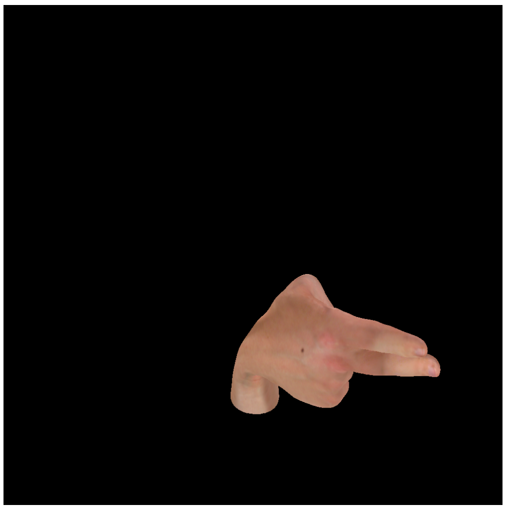
(b) -G-



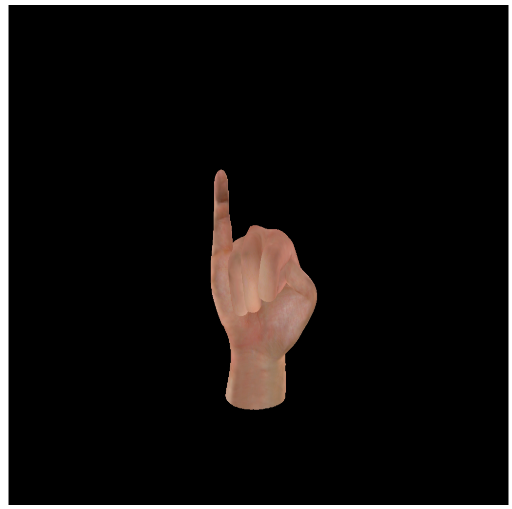
(c) -G- with thumb

Figure 2.19: Traditional versus new -G-

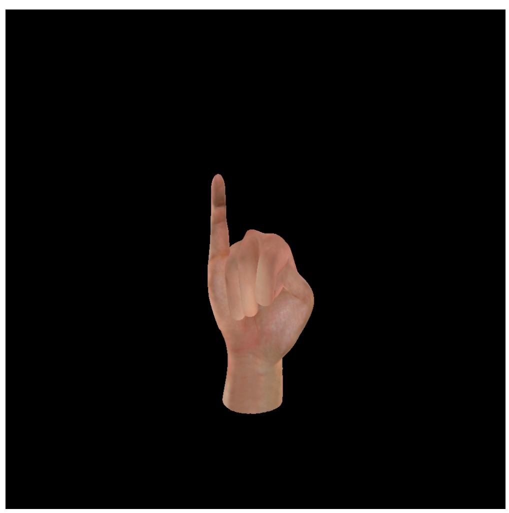
The traditional notations for -H-, -I-, and -J- did not need adjustment (see figure 2.20). Although, the thumb position for -I- and -J- could use more refining (flexing over the fingers), with methods for determining articulator contact like those discussed above. Additionally, variants with the thumb in the positions below are found. See table 2.9 for full specifications.



(a) -H-



(b) -I-



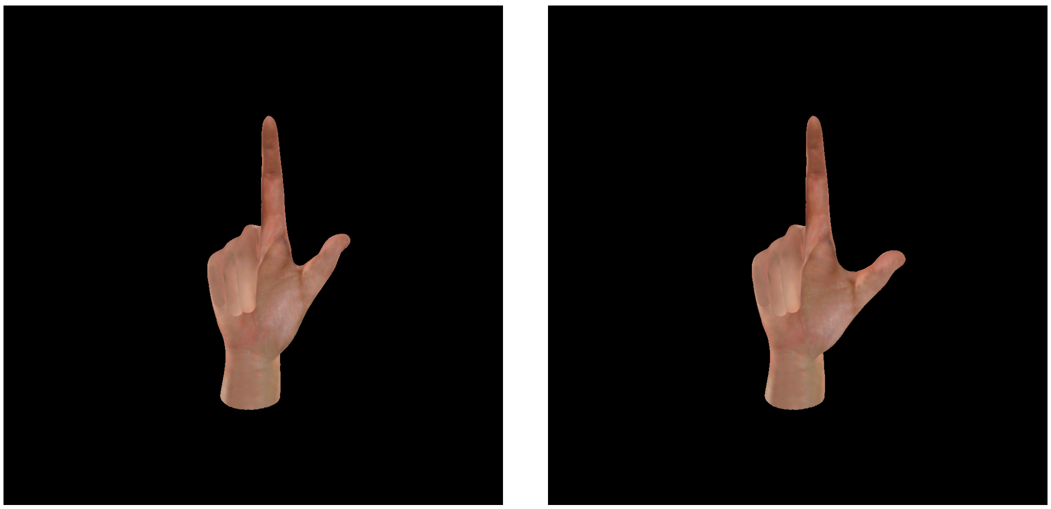
(c) -J- with thumb

Figure 2.20: Renderings of -H-, -I-, and -J-

The traditional notation for -K- uses the stacked feature to get the index and middle fingers in the appropriate position (where the index finger is ext-ext, and the middle finger is flex-ext for base and non-base joints respectively). The stacked feature does not have a straightforward implementation in the articulatory model. The current proposal (which accounts for any stacked handshape with only two fingers) is that one of the fingers is in the primary selected group, and the other is in the secondary selected group. This gets the visualization correct (see figure 2.21). Further work on this is needed, but as an initial proposal to account for more than two finger stacked handshapes using the articulatory model is that the end points of the stacked fingers (i.e. the index and the pinky for a handshape where all fingers are stacked) are in separate selected-finger groups, and when that happens, the other fingers interpolate flexion between these two end points. See table 2.9 for full specifications.

Figure 2.21: Rendering of -K-

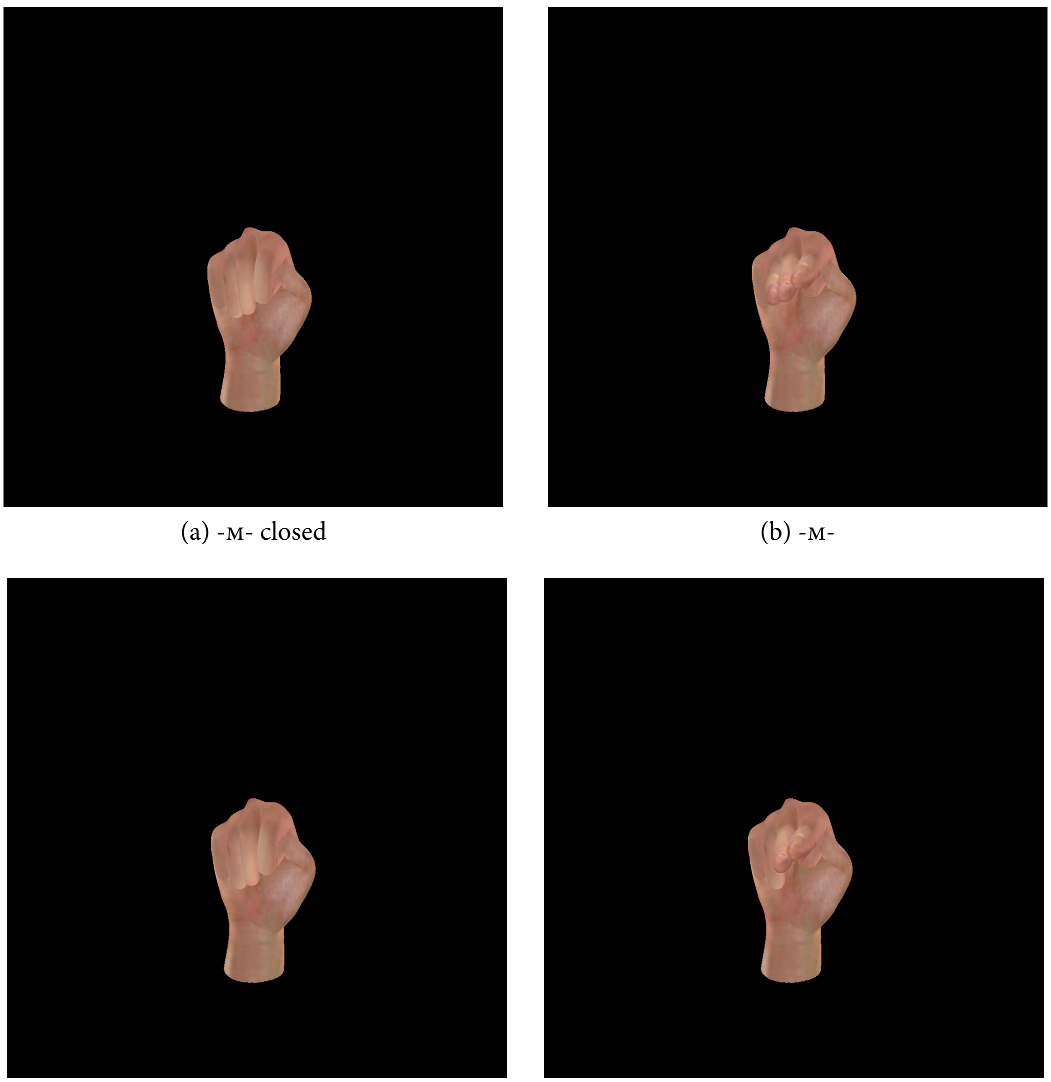
The traditional notation for -L- has the thumb in the same selected group as the index. Additionally, this group is not abducted. This is problematic because, as can be seen in figure 2.22, the thumb does not have enough radial abduction (although the contrast is quite subtle). For this reason, the new proposed -L- has the thumb in the secondary selected finger group, and is abducted. See table 2.9 for full specifications.



(a) -L- traditional (b) -L-

Figure 2.22: Traditional versus new -L-

The handshapes for -M- and -N- are controversial. Although the citation form is frequently shown as the index, middle (and for -M- ring) fingers fully flexed over the thumb, more frequently there is not full flexion of those digits, but rather the non-base joints are extended. Forms for both variants are proposed (see figure 2.23). For the closed variants, more work is needed in order for the fingers to be prevented from achieving full flexion when the thumb is in the way. See table 2.9 for full specifications.



(a) -M- closed

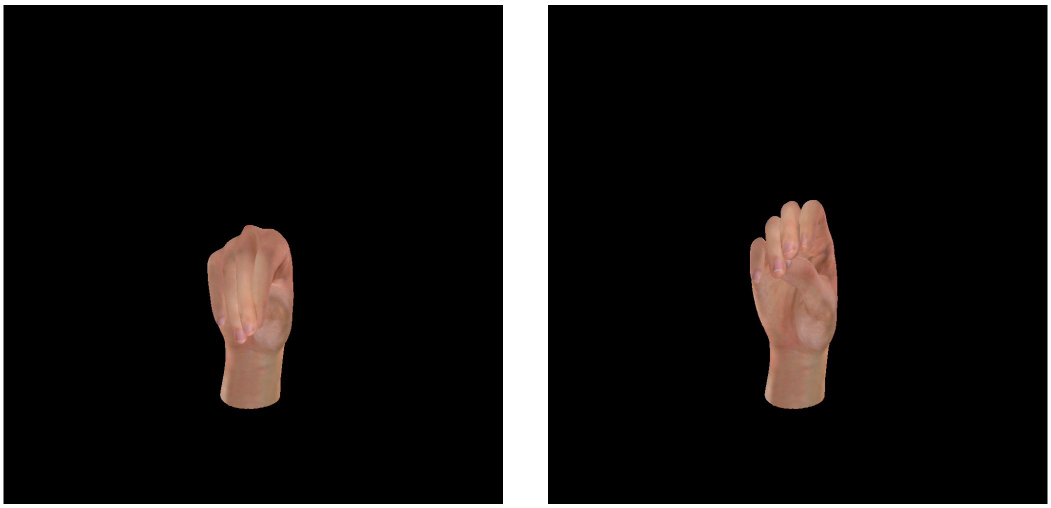
(b) -M-

(c) -N- closed

(d) -N-

Figure 2.23: Open versus closed -M- and -N-

The traditional notation for -o- has the joint configuration flex-mid for base and non-base joints respectively. This is problematic because, as can be seen in figure 2.24, the fingers are too closed. For this reason, the new proposed -o- has the joint configuration mid-mid for base and non-base joints respectively). See table 2.9 for full specifications.

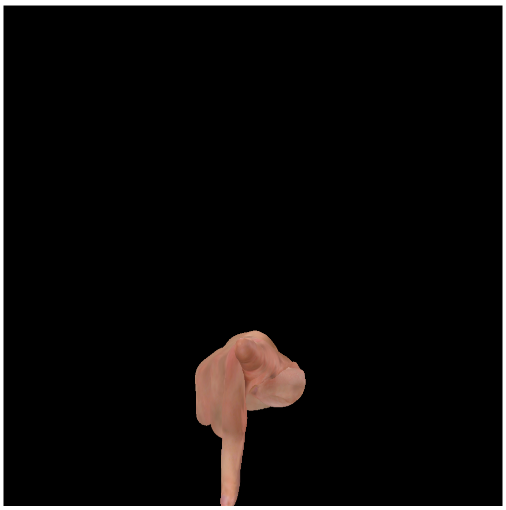


(a) -o- traditional

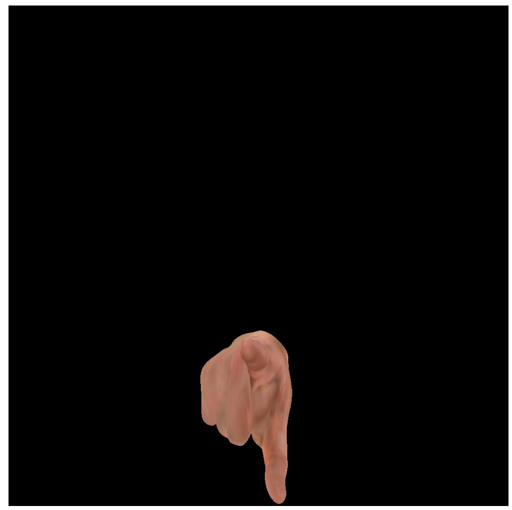
(b) -o-

Figure 2.24: Traditional versus new -o-

The traditional notations for -P-, -Q-, -R-, and -S- did not need adjustment (see figure 2.25). Both -R- and -S- need additional refinements to make the articulators not collide. Currently for -R-, the index and middle finger intersect. For -S-, the thumb intersects with the fingers rather than closing over them. Additional logic in the rendered can be added that evaluates if the hand mesh would intersect, and if it does adjust the joint angles.[25] See table 2.9 for full specifications.



(a) -P-



(b) -Q-

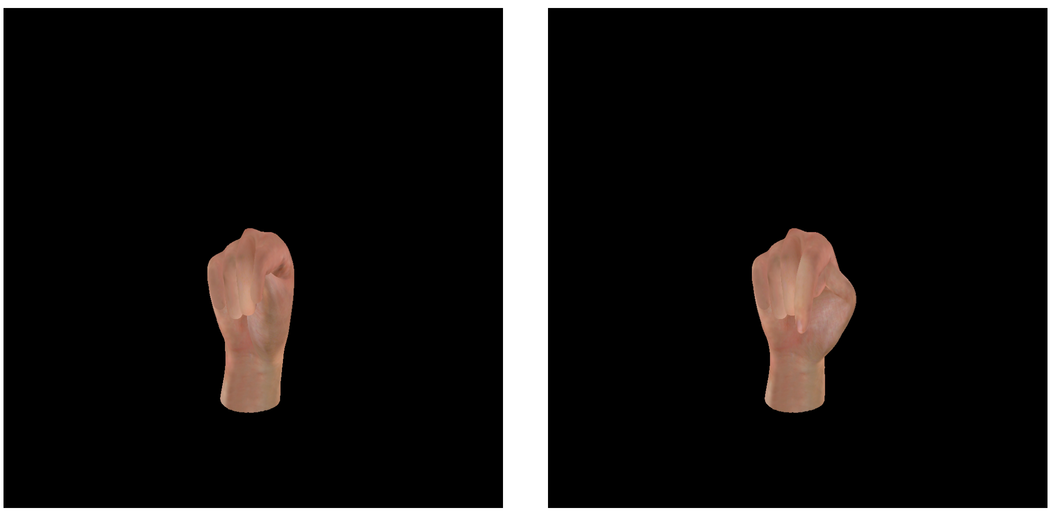
(c) -R-

(d) -S-

Figure 2.25: Renderings of -P-, -Q-, -R-, and -S-

25. This is similar to what happens in the physical world: for example, in an -S- handshape: the muscles contract for full flexion on the thumb, but if there is a finger in the way the thumb cannot fully flex, but rather is prevented from doing so by the fingers.

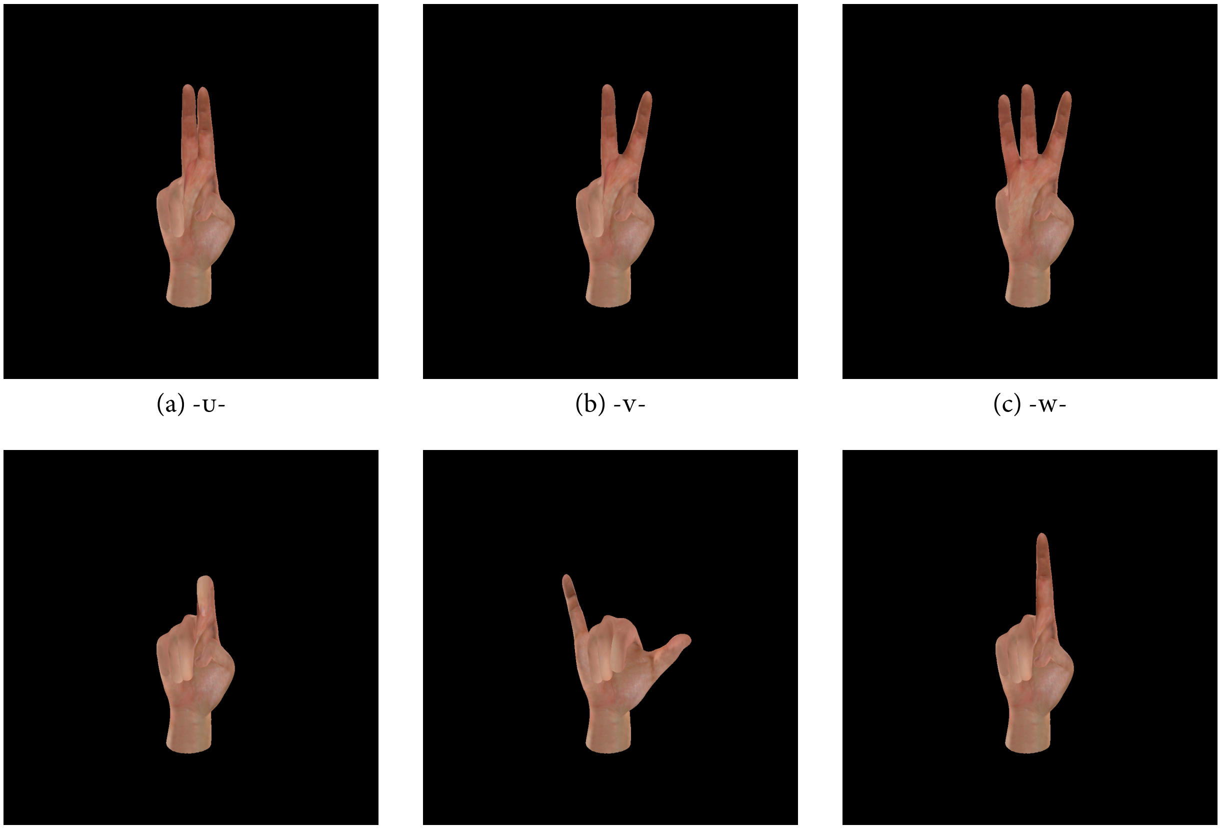
The traditional notation for -T- has the thumb in the selected finger group, and there is a crossing feature. This is problematic because, as can be seen in figure 2.26, the thumb is too far (radially) abducted. For this reason, the new proposed -T- has the thumb in the nonselected finger group. Although, another solution to this problem would be to alter the amohs model to have a more goal-oriented way of realizing the crossing feature (also described here as negative abduction). This could be achieved by coding specific abduction-adduction angles for each pair of digits to cross, or be implemented as a physiologically constrained attempt of the digits to cross over each other. See table 2.9 for full specifications.



(a) -T- traditional　　(b) -T-

Figure 2.26: Traditional versus new -T-

The traditional notations for -U-, -V-, -W-, -X-, -Y-, and -Z- did not need adjustment (see figure 2.27). See table 2.9 for full specifications.



(a) -U- (b) -V- (c) -W-

(d) -X- (e) -Y- (f) -Z-

Figure 2.27: Renderings of -U-, -V-, -W-, -X-, -Y-, and -Z-

2.6 Looking forward

This implementation will work for any handshape that has been (and can be) described using either the Prosodic Model notation system, or the Articulatory Model of Handshape. Although this chapter was limited to the handshapes used in ASL fingerspelling, the model itself is not limited to that. For example, the handshape that is used for the airplane classifier (also called the I love you or ILY handshape), can be generated (figure 2.28) from the prosodic notation for the handshape, `H;T-^;#`, directly using the code in figure 2.29. The hand alone could also be used for stimuli testing handshape or fingerspelling, although it is not clear how people would react to a disembodied hand producing handshapes or fingerspelling. Using the 3D rendering system in `ogre`, this hand could be attached to a full body model of a person signing for use in rendering a signing avatar. More work is needed to generate naturalistic movements of this full body avatar and join it to the handshapes produced here.

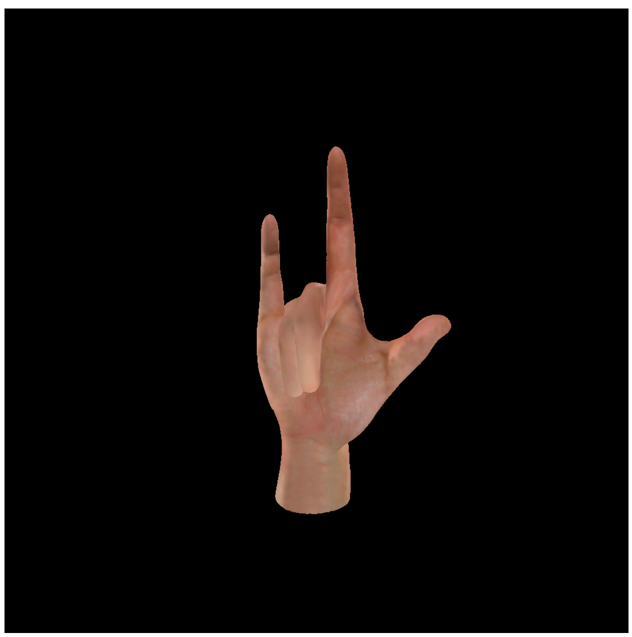


Figure 2.28: A rendering of the hand for the airplane classifier or ILY handshape.

```
import amohs
amohs.render.renderImage(amohs.hs.arm(handshape=amohs.pm.
   pmHandshape("H;T-^;#").toAMhandshape(), orientation="
   defaultFS").toArmTarget(), "./ily.png")
```

Figure 2.29: Code to render the airplane classifier or ILY handshape (in Prosodic Model notation: `H;T-^;#`).

2.7 Conclusions

The phonetics-phonology interface has not been explored extensively for sign languages. The Articulatory Model of Handshape adapts articulatory phonology to the limited task of modeling the phonetics-phonology interface for handshape in sign languages. The Articulatory Model of Handshape clearly links phonological specifications with predicted articulatory targets within an articulatory phonology framework. This implementation makes clear predictions about variation, namely that because speech generally (which includes signing), and fingerspelling specifically, are a set of dynamic gestures, the result of any given handshape is a function of the identity of the segment being articulated, as well as properties of the handshapes that precede and follow it. Additionally, this variation is not a simple averaging of all of the configurations, but rather is structured: the active (or selected) articulators will be less contextually influenced than articulators that are non-active (or nonselected). Additionally, because the articulatory model has been fully implemented computationally the entire influence of assumptions about phonological specifications, and how they are translated into articulatory targets can be seen. In developing the translation mechanisms it is clear that some small modifications to existing phonological specifications for a subset of the ASL fingerspelling handshape inventory had to be made. Finally, this computational implementation of the Articulatory Model of Handshape is a critical first step in developing a computational model that could produce the expected coarticulation. But, before numerical predictions (as well as 3D renderings of these predictions) of coarticulation can be implemented in the Articulatory Model, we need robust quantitative measures of factors that contribute to coarticulation phenomena, as well as robust quantitative details about the amounts of coarticulation that are associated with these factors.

Chapter 4 will be just this kind of quantitative analysis of pinky extension coarticulation. In addition, this analysis of coarticulation will confirm the general hypotheses that follow from articulatory phonology, and the Articulatory Model of Handshape.

Chapter 3

Timing and segmentation of ASL fingerspelling

3.1 Introduction

Segment duration is one of the most basic elements of phonetic description in any language. We know that segment duration is affected by numerous macro-factors (e.g. individual variation, utterance speed, and familiarity with the target item) as well as similarly numerous micro-factors (e.g. segment type, preceding and following segments, articulatory complexity, and stress; see Klatt (1976) for a review, and Peterson & Lehiste (1960); Lehiste (1972); Oller (1973); Port (1981) for specifics). As with all phonetic features, duration adds crucial information to the language signal. Voice onset time is a similar temporal phonetic feature that has a vast body of research on how it greatly influences the perception of segment identity. Segment duration is used by listeners to differentiate between segments, as discussed by (Klatt, 1976). Segment duration can also be used in speech recognition to facilitate processing. The macro-factors can be used to adjust algorithms, as well as to help a language model predict words likely to be spoken. For example, if a given word could be either a native word, or a foreign word, if the segment durations are longer than average it is more likely to be the foreign word, especially if speaker variation, and utterance speed have already been controlled for. The micro-factors are much more directly applicable to the speech processing itself, helping to predict on a segment by segment basis what the most likely one uttered was. Segment duration, by itself, is a very crude predictor of segment identity, but in conjunction with other details it becomes an important tool in automatic recognition of speech (Livescu & Glass, 2001; Chung & Seneff, 1999; Levinson, 1986).

Segment duration in fingerspelling provides a similarly crude — but important — tool in automated fingerspelling recognition. We expect segment duration in ASL fingerspelling to vary with many of the same macro-factors — they are almost exactly the same for fingerspelling as they are for spoken communication. Some micro-factors, on the other hand, will differ because of the change in modality. There will be similar articulatory factors, although these stem from the limitations of hands and arms rather than the mouth and vocal tract. Other micro-factors ought to remain the same. As we have seen in research on coarticulation (chapter 4, for example), what comes before and after a segment has an influence on the intermediate segment generally, including its duration. This is a result of producing language which is at a very low level the process of moving a set of articulators to targets in a sequence. These effects are those that the articulatory model of handshape described in chapter 2, based on articulatory phonology, predict as a result of modeling the systems involved with going from abstract mental representations of words, to the physical properties of the articulators responsible for executing the speech.

The structure of segments and their duration in fingerspelling has one large difference from segment duration in speech: there is no obvious dichotomy between segments that are short (consonants in spoken languages) and segments that are longer (vowels), which join together to form larger (syllabic or moraic) units. Rather, the segments of fingerspelling are a series of target handshapes that the articulators move through. These segments generally consist of brief holds of handshapes that correspond to the letters of the fingerspelled word[1], with transitions between these holds.

There have been a few descriptions of fingerspelling rate in the literature which all fall between 2.18 and 6.5 letters per second (154–459 msec/letter), with a mean of 5.36 letters per second (187 msec/letter). An overview of the previous studies can be found in table 3.1; the details of each study are described below.

Bornstein (1965) reports a rate of 5 letters per second (200 msec/letter). The fingerspelling was elicited here for inclusion in a video course to teach fingerspelling production and perception.

Zakia & Haber (1971) report rates of fingerspelling between 6.17 letter per second (162 msec/letter) and 5.26 letters/sec (190 msec/letter). There are rates as slow as 1.90 letters per second (527 msec/letter) for students not familiar with fingerspelling.

1. Although these points are generally where the handshape is closest to the canonical one for a given letter, as we discuss in chapter 4 they frequently deviate substantially from the canonical forms.

	rate – fastest		rate – slowest	
publication	letters/sec	msec/letter	letters/sec	msec/letter
Bornstein (1965)	5.00	200	5.00	200
Zakia & Haber (1971)	6.17	162	1.90	527
Hanson (1981)	6.15	162	5.65	177
Hanson (1982)	5.58	170	5.26	190
Wilcox (1992)	3.33	300	3.33	300
Jerde *et al.* (2003)	4.46	319	2.00	500
Quinto-Pozos (2010)	8.00	125	5.00	200

Table 3.1: **An overview of previous reports of fingerspelling rates** See the text for a more detailed discussion, as well as specifics about the signers and words for the fastest and lowest groups.

Hanson (1981, 1982) measured the timing properties of fingerspelled words, pseudowords, and nonce words to be used as stimuli in a fingerspelling perception study. In both, she measured the overall duration of the word, and divided it by the number of letters. (Hanson, 1981) reports rates of 6.15 letters per second (162 msec/letter) for English words, and 5.65 letters per second (177 msec/letter) for non-English words. Hanson (1982) reports average rates of 5.88 letters per second (170 msec/lettter) for (English) words, 5.56 letters per second (180 msec/letter) for (English) pseudowords, and 5.26 letters per second (190 msec/lettter) for nonce words.

Wilcox (1992) is the first known study to use kinematic data. His study was extremely limited and looked at only the fingerspelled B-U-T and the loan-fingerspelled T-O-O-B-A-D (he glosses as #TOO-BAD). Targets lasted for a mean of 91 milliseconds and transitions lasted a mean of 314 milliseconds across both types. For the fingerspelled B-U-T targets had a mean of 104 milliseconds, and and transitions lasted for a mean of 319 milliseconds. Adding together three holds and two transitions, and then dividing by the number of letters in B-U-T gives a rate of 3.33 letters/sec (300 milliseconds msec/letter).

Jerde *et al.* (2003) used a data glove to analyze coarticulation in fingerspelling. They had four signers total; three had ranges of 3.13–4.46 letters per second (224–319 msec/letter). One subject was considerably slower at a rate of 2–2.36 letters per second (434–500 msec/letter). None of the subjects in this study were deaf, all subjects were fluent hearing interpreters. It was not reported if these signers were native signers of ASL or not.

Quinto-Pozos (2010) found that there was an overall rate of 5–8 letters per second (125 – 200 msec/letter). There were significant differences between signers in less formal settings, although they went away in formal ones. Additionally, longer words had a faster rate than shorter words.

Geer (2010) looked at differences in native versus non-native signers. She did not report absolute times, but rather percentage of the fingerspelled word that was transition. She restricted her analysis to two-, three-, and four-letter abbreviations. Native signers' fingerspelling had between 67% and 73% transitions (across two conditions: in context and isolated).

Finally, Reich & Bick (1977) showed that word-medial letters are held for longer than initial or final letter, although this study looked at Visual English in an educational setting where English was spoken while words were fingerspelled. The fingerspelling system is the same between this form and the fingerspelling used in ASL discourse; however, because of the simultaneous language production the timing properties might differ substantially.

Building on these studies, we have collected and analyzed timing data from 4 ASL signers. We replicated many of the previous findings, and additionally found that there are:

- large differences between different letter types,
- different positions within a fingerspelled word,
- large individual differences,
- differences based on the type of word being fingerspelled, and
- finally, we found a heretofore undiscovered difference in the ratio of holds to transitions between signers.

. Section 3.2 describes the annotation methodology, section 3.3 describes a model of the rate of fingerspelling, which feeds into models of holds and transitions (sections 3.4 and 3.5 respectively), and finally motion capture data is explored as a second measure of rate in 3.6.

3.2 Methods

We recorded and analyzed timing information for 3 native ASL signers and 1 early learner, fingerspelling a total of 3,684 productions[2]. We annotated the video by identifying the hold (also known as posture or target) for each fingerspelled letter (which we call apogees). There were 21,453 apogees in total. Data from additional signers (1 new signer with 3 wordlists) and wordlists (1 additional wordlist for 2 signers, and 3 additional wordlists for the other 2) have been collected, but have not yet been annotated. The word lists used are in appendices B.1, B.2, and B.3. The following sections describe in detail the data collection and annotation process.

3.2.1 *Video recording*

The data was collected across different sessions that consisted of all of the words on one word list. During each session the signer was presented with a word on a computer screen. They were told to fingerspell the word, and then press a green button to advance if they felt that they fingerspelled it accurately, and a red button if they had made a mistake. If the green button was pressed the word would be repeated, the signer would fingerspell it again, and then they would move on to the next word. If the red button was pressed the sequence was not advanced, and the signer repeated the word. Most sessions were collected at a normal speed, which was supposed to be fluid and conversational, the signers were told to fingerspell naturally, as if they were talking to another native signer.[3] For a small number of sessions the signers were asked to fingerspell at a careful speed, which was supposed to be slow and deliberate.[4] For most sessions the signers sat in a chair with an armrest that they could use if they felt the desire to. In a small number of sessions the signers were asked

2. Each production is a single, specific fingerspelling of a word. These could also be called *word instances*, borrowing from computer science terminology.

3. The instructions, given in ASL were to: "proceed at normal speed and in your natural way of fingerspelling."

4. Again, in ASL "be very clear, and include the normal kind of transitional movements between letters." The signers were also specifically asked not to punch the letters with forward movements, as is often done for emphatic fingerspelling.

to stand rather than sit. Each session lasted between 25-40 minutes, there was a self-timed break in the middle of each session for the signer to stretch and rest.

Video was recorded using at least two cameras, both at 45 degrees angles from straight on. Each of these cameras recorded video that was 1920×1080 pixels, 60 fields per second, interlaced, and using the AVCHD format. These files were then processed using FFMPEG to deinterlace, crop, resize, and reencode the video files so that they were compatible with the ELAN annotation software. The command used to encode, and separate out each session was `cat [list of input mts files] | ffmpeg -i - -ss [start time] -t [duration] [options] -an -sameq -y [out file]` where options were either of the two: 1. deinterlaced, cropped, and scaled file for use with ELAN 2. full sized, deinterlaced only file for use with video recognition.

deinterlace+crop+scale `-vf "[in] yadif=1 [o1]; [o1] crop=1464:825:324:251 [o2]; [o2] scale=852:480 [out]"`

deinterlace `-vf "[in] yadif=1 [out]"`

3.2.2 *Annotation*

Our annotation method is separated into two main parts: 1. a simple task to identify approximate times of each apogee (*peak detection*) and 2. a verification task to determine precise timing for each apogee (*apogee verification*). The first is designed to be extremely quick, and allow multiple annotator judgements to be aggregated together. The second is much more exact, with the goal of providing precise data on the timing of handshape change during the fingerspelling[5].

Peak detection

Once this video was processed, 3–4 human annotators identified the peak of each apogee. Peaks were defined as the point where the articulators changed direction to proceed on to the next apogee

5. This system for annotation is applicable only for fingerspelling. It could be extended to other parts of ASL discourse, however it would be missing other critical parameters of the language: movement, location, and non-manuals.

(i.e. where the instantaneous velocity of the articulators approached zero). This point was also where the hand most closely resembled the canonical handshape, although in normal speed the handshape was often very different from the canonical handshape. Two FS-letters defied definition in this manner, namely -J- and -Z-, since they have movement. With these two FS-letters annotators were asked to just indicate a peak when they could determine that it was one of these two FS-letters. Peak detection is simple, and requires only minimal training; additionally, annotators found this task very intuitive.

In order to determine the most likely apogee locations the peaks from each annotator were averaged using an algorithm that minimized the mean absolute distance between the individual annotators' peaks. This algorithm allowed for misidentified peaks by penalizing missing or extra peaks from individual annotators. Using logs from the recording session, a best guess at the FS-letter of each peak was added using forced alignment (starting at the left edge of the word, matching each apogee with a letter in the word).

Apogee verification

Finally, a more experienced, second language learner of ASL or someone specifically trained in fingerspelling annotation went through each file and verified the location and identity of each apogee from the combined peaks from the peak detection stage. We defined apogee as the point when the handshape reached a configuration that was closest to the canonical handshape for a given FS-letter. If a handshape remained stable for more than one frame, each stable frame was marked. Details for the distribution of holds will come in the following sections.

A naive description of fingerspelling might be a series of handshapes (apogees), one for each letter in a word, with each being held briefly.[6] Although this is on some level accurate for some apogees, it is not the case for every one. We observed there were three seemingly distinct realizations for apogees.

6. Of course this ignores FS-letters that involve movement (-J- and -Z-).

Multiframe hold – a handshape that is held statically for more than one frame

Single frame hold – a handshape that appears static for a single frame

Instantaneous – a handshape that is most canonical in what appears to be a transition, but does not appear in a stable state

Impressionistically, the first and last apogees in words are frequently, although not always, multiframe holds. Handshapes that are neither fully flexed nor fully extended (-E- especially) were those most frequently in the instantaneous group. Other apogees populated the other groups with varying frequencies. The effect of position and FS-letter identity on hold duration will be explored in section 3.4 below.

Because this is a task of annotating handshape, a select set of bigrams may result in no change in handshape, although there should be two letters (e.g. H-U or U-H). For these the handshape stability is annotated and marked with the sequence of letters as in the word. The orientation is marked by aligning instantaneous markers when the hand is oriented in the most canonical position. For an example of the sequence of holds and transitions based on handshape holds for one token of the word -C-, -O-, -S-, and -T- see figure 3.1.

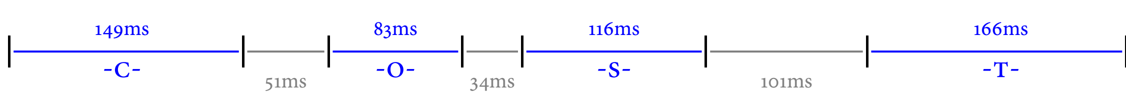


Figure 3.1: **A visualization of the holds and transitions for one token of the word -C-, -O-, -S-, and -T-** Blue lines represent the holds, and the grey lines represent the transitions.

Handshape – We defined a handshape as stable if all of digits assumed a position and maintained it with only minor fluctuations. As soon as any digit moved the handshape is considered to not be stable anymore. We were conservative with respect to holds, in that if a digit moves a small amount, but that movement is part of a larger movement that preceded or followed, that was not considered stable.

Orientation – Most FS-letters are produced with the palm facing away from the signer's body. The few exceptions to this are -G-, -H-, -P-, and -Q-[7] where the palm faces the signer (-G- and -H-; labeled *side* in the analysis described below), or faces down (-P- and -Q-; labeled *down* in the analysis described below). Because handshape and orientation changes are not always synchronized, we have annotated handshape stability as a hold, even if the hand is continuing to undergo an orientation change. Future annotation is necessary for orientation changes in detail and determine the pattern of stability and motion that exists there.

Movement – Two FS-letters are described as having movement: -J- and -Z-. -J- involves an orientation change, and -Z- traces the path of the letter[8]. For both of these FS-letters, again we have annotated a hold to be where the handshape is stable, regardless of orientation change, or path movement.

Handshape detail – A detailed (although not exhaustive) description of handshapes are given in table 3.2. This is meant to be guidance to annotators, and is intended to catch the core features for each handshape, allowing for the systematic variation known to exist in handshape. If some of these features match, but the handshape is significantly different than expected, annotators added a diacritic (+) to note a large amount of deviance. This is not intended to exhaustively mark all of the deviant handshapes, but only those that should be looked into further. There are some instances where an apogee is found, but no peak was detected. Although we have not analyzed this systematically these instances are frequently apogees that are instantaneous, or apogees that occur extremely close to each other. These apogees are noted with a different diacritic (*). Finally if two handshapes have compressed to form a single apogee a digraph is used to annotate the combined apogee. Examples that we have seen so far are -GH-, - IT-, - IN-, - IO-, - IL-, and - CI-. Here the digraph is simply two letters that seem to make up the single apogee; for consistency they should

7. These are the FS-letters traditionally described as having different orientation, there are other possibilities that we have found as well: -X- and -Y-.

8. This is frequently abbreviated to just a horizontal line, representing the top bar of the z.

be written in alphabetical order regardless of the orthographic order of the letters in the word being fingerspelled. See table 3.3 for a description of those found so far.

3.2.3 Data for analysis

In order to analyze the data that has been annotated so far, productions that were problematic in a variety of ways were excluded. Only productions where the signer felt they had fingerspelled them correctly (those where the signer pressed the green button), were included (89 productions were excluded). Only productions where each hold corresponded to a single letter were included (126 productions excluded). One place where this happens is where adjacent apogees only differed in orientation: an example of a production that fit this category was a production of H-U-S-B-A-N-D, where the handshape was not change between the -H- and the -U-, even though the orientation of the hand does change (which is exactly what distinguishes these two FS-letters). This allows us to set aside the issue of orientation change during a single handshape hold for these analyses, and considering that these represent only 3% of the total data, holding them out will not have a huge impact on the overall outcome of the analyses. Only productions where there was one and only one hold for each letter in the word were kept (600 productions were excluded). These excluded productions are likely occurrences of reduction and epenthesis, although an analysis of these phenomena would be merited, it is beyond the scope of this work. Productions where holds or transitions were excessively long were removed: any hold was longer than 500 milliseconds (half a second), or any transition was longer than 1000 milliseconds (one second) (94 productions excluded). Finally, when there was an obvious error in the trial (e.g. the signer repeated the word, but still pressed the green button) were excluded (2 productions excluded). This resulted in a data set of 2,918 productions (of words) and 16,959 apogees across two word lists, and 4 signers.

9. There are some instances where the index finger flexes as the thumb is moving away from the base between the middle and index finger. In these cases the apogee for -T- should be marked when the index finger has started coming down, and the thumb starts moving. Frequently there looks to be a slight brush of the tip of the thumb across the proximal phalange.

FS-Letter	description of the most canonical shape
-A-	all fingers are flexed, with thumb touching the radial side of the hand, or extended
-B-	all fingers are extended. The thumb is hyper flexed across the palm
-C-	all the fingers are curved.
-D-	the index finger is fully extended. At least the middle finger is making contact with the thumb: the ring and pinky may be either flexed, or making contact with the thumb
-E-	the thumb is bent and hyper flexed across the palm, the index finger is bent and may be touching the thumb. The other fingers may be bent, like the index finger, or flexed completely.
-F-	contact with the index and thumb. The middle, ring, and pinky are all extended
-G-	the index finger is fully extended. All other fingers are flexed. The thumb is either extended fully, or unextended, against the middle finger.
-H-	index and middle fingers are fully extended. All others are flexed. The thumb is unextended or extended
-I-	the pinky is fully extended, all other fingers are flexed. The thumb is either hyperflexed, or unexetended, against the radial side
-J-	the pinky is fully extended, all other fingers are flexed. The thumb is either hyperflexed, or unexetended, against the radial side
-K-	the index finger is fully extended, the middle finger is extended, but bent 90° at the joint closest to the hand
-L-	the index finger is fully extended, and the thumb is extended away from the hand. All other fingers are flexed
-M-	the index, middle, and ring fingers are closed, or flat-closed over the thumb, which is hyper flexed across the palm, possibly touching the base of the pinky and ring fingers
-N-	the index and middle fingers are closed, or flat-closed over the thumb, which is hyper flexed across the palm, possibly touching the base of the ring and middle fingers
-O-	the thumb and the index finger are touching in a curved, closed configuration. The other fingers are either in the same configuration, touching the thumb, or completely flexed.
-P-	the index finger is fully extended, the middle finger is extended, but bent 90° at the joint closest to the hand
-Q-	the index finger is fully extended. All other fingers are flexed. The thumb is either extended fully, or unextended, against the middle finger.
-R-	the index and middle fingers are extended and crossed over each other. All other fingers are flexed
-S-	all fingers are completely flexed, with thumb hyperflexed across the fist
-T-	the index finger is closed, or flat-closed over the thumb, which is hyper flexed across the palm, possibly touching the base of the middle and index fingers[9]
-U-	the index and middle fingers are completely extended, all other fingers are flexed, the thumb is hyperflexed across the palm
-V-	the index and middle fingers are completely extended and are spread apart, all other fingers are flexed, the thumb is hyperflexed across the palm
-W-	the index, middle, and ring fingers are completely extended and are spread apart, all other fingers are flexed, the thumb is hyperflexed across the palm
-X-	the index finger is bent similar to -E-, all other fingers are completely flexed
-Y-	the pinky is fully extended, the thumb is hyper extended away from the hand. All other fingers are flexed
-Z-	the index finger is fully extended, and all other fingers are flexed

Table 3.2: Description of handshapes

FS-Letter	description of the most canonical shape
-I- and -T-	the index finger is closed, or flat-closed over the thumb, which is hyper flexed across the palm, possibly touching the base of the middle and index fingers, and the pinky is extended. This should only be used if the hand reaches this configuration at a single frame.
-G- and -H-	index and middle fingers as well as the thumb are extended. Similar to the CL 3 handshape.
-I- and -N-	index and middle fingers are (partially) flexed over the thumb, and the pinky is fully extended.
-I- and -O-	index, middle, and ring fingers are looped and touching the thumb, and the pinky is fully extended.
-I- and -L-	the index and thumb are extended (the thumb is abducted), and the pinky is fully extended.
-C- and -I-	the index, middle, and ring fingers as well as the thumb are partially extended (also described as curved open), and the pinky is fully extended.

Table 3.3: Description of digraphs

3.3 Rate

Although we have time-annotated individual apogees, the first thing to quantify is the rate of fingerspelling. The reasons for this are two fold: first, most of the literature is framed in terms of overall fingerspelling rate, so this allows for direct comparison with previous results. Additionally, overall speech rate is always a large driver of segment and syllable duration in spoken languages (see the seminal (Crystal & House, 1982, 1988), among many others). We anticipate this will be the case for ASL fingerspelling as well, so rate needs to be calculated to be used as a predictor in models for segment duration that will be explored in sections 3.4 and 3.5.

Rate was calculated by measuring the duration (in seconds[10]) of the fingerspelled word, from the beginning of the first hold, to the end of the last hold, and then dividing it by the number of letters (with is also equal to the number of holds for this data). This data was then modeled using a hierarchical linear regression model (as a reminder, these are also known as linear mixed effects regressions). Some of the discussion that follows was included in chapter 1, but it is repeated and expanded upon here to serve as a reminder of the statistical reasoning that is integral to this chapter (as well as the following chapter). These regression models are similar to the models used in chapter 4, although instead of predicting the probability of a binary outcome, the prediction (also known as

10. Although milliseconds are used as measures of durations of apogee holds, for words durations are converted to, and reported in, seconds. The reasons for this are twofold: 1. word durations are typically on the order of seconds, where as apogee hold durations are on the order of tens of milliseconds. 2. Reporting word durations in seconds allows for easy calculation of rates in letters per second, which match how rates have been reported in previous literature.

the outcome, or dependent variable) is a continuous distribution, assumed to respond linearly to the predictors (also known as inputs, or independent variables). There are a number of advantages to using hierarchical regression models. First, hierarchical models are more robust against unbalanced designs (for example, here, 2 signers fingerspelled words from 2 word lists and 2 signers fingerspelled words from only 1 yielding double the amount of data for 2 of the signers compared with the other 2). Even more importantly, hierarchical models were chosen because they account for the structure among the properties of the data that is being analyzed. In order to illustrate this, consider the structure of the data being analyzed here. Each production in our data has a number of properties about it that we want to include in our analysis:

- it has a word identity (which it shares with a few other productions),
- it was fingerspelled in a given trial (pair of word repetitions),
- within this trial it was either the first or second repetition,
- these trials are ordered within each wordlist,
- each word has additional properties:
  - its length,
  - its type (i.e. name, noun, non-English word),
  - which wordlist it is a member of,
- finally, each production was fingerspelled by a specific signer.

Some of these properties are related, in that they are nested within each other: words are associated with a single wordlist. All of these properties may influence the rate of fingerspelling, and so need to be included in the model. There is a distinction between properties like these that are used as predictors (often called fixed effects), and properties that are used as grouping variables (often called random effects), that is, those that define groups, and the structure of those groups, within the data. The choice between what is used as a predictor and what is used as a grouping variable is not uncontroversial (Gelman & Hill, 2007; Barr *et al.*, 2013; r-sig-mixed-models listserv, 2010; glmm wiki, 2014).

> Treating factors with small numbers of levels as random will in the best case lead to very small and/or imprecise estimates of random effects; in the worst case it will lead to various numerical difficulties such as lack of convergence, zero variance estimates, etc..[sic] (A small simulation exercise shows that at least the estimates of the standard deviation are downwardly biased in this case; it is not clear whether/how this bias would affect the point estimates of fixed effects or their estimated confidence intervals.) In the classical method-of-moments approach these problems may not arise (because the sums of squares are always well defined as long as there are at least two units), but the underlying problems of lack of power are there nevertheless.
>
> (glmm wiki, 2014)

> Advice is sometimes given that multilevel models can only be used if the number of groups is higher than some threshold, or if there is some minimum number of observations per groups. Such advice is misguided. Multilevel modeling includes classical regression as a limiting case (complete pooling when group-level variances are zero, no pooling when group-level variances are large). When sample sizes are small, the key concern with multilevel modeling is the estimation of variance parameters, but it should still work at least as well as classical regression.
>
> (Gelman & Hill, 2007, 275)

Groups are always recommended for variables that have a large number of levels (word identities in this model, or frequently, subjects in psycholinguistic studies with large numbers of subjects). But they can also be used if one is interested in modeling the variation between the levels of the groups, and are required if the intent is to generalize beyond the population of groups that were sampled. Additionally, as discussed in the quote above, small numbers of groups will either lead to estimates of less variation between the groups than actually exists, or models not fitting. For these reasons, the signer group (with four levels), will be treated as a grouping variable, rather than a predictor variable[11].

The model gives an intercept for the outcome (the interpretation of which varies depending on the scales and types of predictors) and then for each predictor (and interactions specified between predictors) the model generates a coefficient which is the magnitude and direction of the effect that the predictor has on the outcome. Grouping variables make adjustments to the intercept (also called random intercepts) or predictor coefficients (also called random slopes) based on group membership of a given data point. Calculating p-values for hierarchical linear regressions is not as straightforward as it is for simple regressions because it is not clear how to calculate the degrees of freedom. There are a number of methods and approximations that have been proposed, they all have drawbacks. (Bates, 2010; Barr *et al.*, 2013; glmm wiki, 2014)

1. using t-statistic as if it were a z-statistic; this method is approximately true for large sample sizes, although can be anticonservative.

2. using likelihood ratio tests on models with the same structure leaving one predictor out at a time; this can be overly conservative because it relies on the χ^2 distribution.

3. Markov Chain Monte Carlo (MCMC) sampling with flat priors; this method has not been implemented for newer versions of `lme4`, and cannot handle complex grouping structures.

4. Parametric bootstrap by fitting a reduced model, simulating data with the reduced model, and comparing test statistics based on simulated data; can be prohibitively computationally intensive, especially with complex models.

Additionally, there are many (Bates (2010); Gelman & Tuerlinckx (2000); Gelman & Hill (2007); Gelman (2013) among others) who argue that calculating and using p-values as a cutoff for statis-

11. Additionally, models fit with signer as a predictor yield similar results. This shows that the choice between signer as predictor and signer as grouping variable is not the source of the effects that we observe here. Additionally, using signer as a grouping variable, allows for the resulting estimates to be for a generic other signer (allowing for generalization to signers outside of our sample).

tical significance is not the appropriate approach. Gelman & Tuerlinckx (2000) especially argues that one should rely on confidence intervals (95%, 99%, etc.) to determine the direction (sign) and magnitude of the effect. Because of this, all model predictors will be visualized using both 95% and 99% confidence intervals, grouping variables will be visualized with 95% confidence intervals (abbreviated CIs). If those intervals do not overlap zero, it is safe to conclude that the effect is real (statistically significant), and the sign indicates the direction of the effect. Additionally, in tables of model outputs stars have been associated with effects that are significant based on the z-statistic approximation. Where these differ, the confidence intervals will be the more conservative of the measures.

For the analysis of rate, the outcome is the rate of fingerspelling (in letters per second). The predictors are word type with levels noun (reference), name, and non-English, repetition with levels first (reference) or second, and their interaction. We expect that each of these will have a systematic effect on the rate of fingerspelling.

Predictors:

- word type with levels noun (reference), name, and non-English,
- repetition with levels first (reference) or second,
- *interaction* word type × repetition

The grouping factors are as follows: We include intercept adjustments for signer (1, 2, 3, or 4), as well as slope adjustments for word type, repetition, and their interaction. We expect there to be large amounts of intersigner variation, and this variation may even include variation in how the signers react to the various predictors. These grouping factors will allow for the effects of the predictors to be separated from signer variation, as well as provide an estimate of the amount of signer variation that is observed. We include intercept adjustments for word length. Word lengths were included as a grouping factor and not as a predictor because we expect that although there will be variation in the rate based on the length of the word being fingerspelled, we do not except that this variation will be systematic or linear. As will be discussed later, and as can be seen in the visualization of the intercept adjustments in figure 3.4, this turns out to be the case. We included intercept adjustments for trial and intercept adjustments for words, which are nested within wordlists. Both of these are included as grouping factors, because we expect that there will be some variation based on their levels, but this variation will not be large or systematic. Although there is debate about using model selection to fit the most parsimonious model that is justified by the data (Bates, 2010) versus fitting full models of all the possible (and measured) predictors and groups Gelman & Hill (2007); Barr *et al.* (2013), consensus seems to be forming around the latter. For that reason, full models (or, as full as will still fit, given the large number of groups and predictors) will be used here.

Grouping factors:

- intercept adjustments for signer (1, 2, 3, or 4), as well as slope adjustments for
  - word type,
  - repetition,
  - their interaction
- intercept adjustments for word length
- intercept adjustments for trial
- intercept adjustments for words, which are nested within wordlists

Overall (for reference levels, nouns on the first repetition) the rate is 5.84 letters per second. For the predictors in the model, there are significant effects for word type, as well as the interaction of repetition and word type. For word type the reference level is noun. Names are slightly (although statistically significantly) slower than nouns (0.40 letters per second slower, or 5.44 letters per second); non-English words are quite a bit slower than nouns (1.27 letters per second slower, or 4.57 letters per second). The second repetition trends towards slower (outside of the 95% CI, but not the

99% CI). The interaction of (second) repetition and names trends towards faster Finally, the interaction of (second) repetition and non-English words is significantly faster (0.40 letters per second faster). The model is visualized in figure 3.2 and full model output is in table 3.4.

Results:

- overall rate: 5.84 letters per second
- significant effects of:
  - word type
  - interaction of repetition and word type

An adjustment to the overall intercept (along with a variance) is estimated for each level in each grouping variable (frequently called random intercepts). Additionally, for some grouping variables the effects of the predictors are allowed to vary (frequently called random slopes). Plots follow showing the magnitude and direction (as well as 95% confidence intervals) for each (figures 3.3–3.6). Starting with signer (figure 3.3), we can see that overall signers 2 and 4 are slower than 1 and 3. Additionally they vary slightly with non-English and names: for signer 3 the effects of names and non-English words are dampened, for signer signer 1 the effect of non-English words is dampened, and for signers 2 and 4 the effect of non-English words is magnified. Moving on to repetition, for signer 4 the effect of the second repetition is dampened, for signer 3 it is magnified, and for both of the others it is unchanged (close to zero). The interaction of name and repetition is magnified for signer 4, dampened for signer 2, and is unchanged for signers 1 and 3. Finally, The interaction of non-English words and repetition is dampened for signers 1 and 3, magnified for signer 4, and unchanged for signer 2. For all of the interactions the effect adjustments are fairly small in magnitude, especially when compared with the intercept adjustments by signer.

Word length does not appear to vary a large amount (see figure 3.4[12], the word lengths are not ordered or grouped in any way). Only 3 and 7 letter words do not overlap zero, and even for those, the magnitude is small. Additionally, the word lengths are not ordered or grouped in any way. If they were ordered, from, say the shortest to the longest, that would indicate that word length could be a linear predictor of rate. If they were grouped in some way, say with all word lengths < 7 in a distinct group and all word lengths > 7 in a separate group, that would indicate that word length was a categorical predictor. Neither of these being true, we are confident that the length of the word does not introduce systematic variation, and should be a grouping predictor like word and trial (although it does still introduce variation, which the hierarchical model accounts for).

Trial and word do not show systematic variation (see figure 3.5 and 3.6 respectively). For both, although there are some instances where the confidence interval does not overlap zero, given the number of groups this is expected: for a normal distribution of 576 levels (how many different word levels there are) we expect to see almost 29 fall outside of a 95% confidence interval.

12. The plots of intercept adjustments for word length, trials, and words that are given for each model are not strictly necessary: In each, there is little or no systematic variation. This is exactly what we expect for a model like this, and grouping factors like word length, trials, and words. For example, any individual word will vary in the rate of fingerspelling, but this variation is typically not large or particularly systematic (specific groups of words pattern together and others differently). These plots are included for each model to show statistical due diligence, but can be skipped without missing anything substantive, if the reader is so inclined.

	coefficient (standard error)
(Intercept)	$5.84(0.38)^{***}$
wordtypename	$-0.39(0.16)^{*}$
wordtypenonEnglish	$-1.28(0.20)^{***}$
repetition2	$-0.47(0.21)^{*}$
wordtypename:repetition2	$0.17(0.08)^{*}$
wordtypenonEnglish:repetition2	$0.39(0.12)^{***}$
AIC	7161.87
BIC	7347.23
Log Likelihood	-3549.94
Deviance	7099.87
Num. obs.	2920
Num. groups: wordList:word	577
Num. groups: trialWR	549
Num. groups: lengthFact	11
Num. groups: signer	4
Variance: wordList:word.(Intercept)	0.22
Variance: trialWR.(Intercept)	0.09
Variance: lengthFact.(Intercept)	0.02
Variance: signer.(Intercept)	0.55
Variance: signer.wordtypename	0.07
Variance: signer.wordtypenonEnglish	0.14
Variance: signer.repetition2	0.18
Variance: signer.wordtypename:repetition2	0.00
Variance: signer.wordtypenonEnglish:repetition2	0.04
Variance: Residual	0.47

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3.4: Coefficient estimates and standard errors of the hierarchical linear model for fingerspelling rate

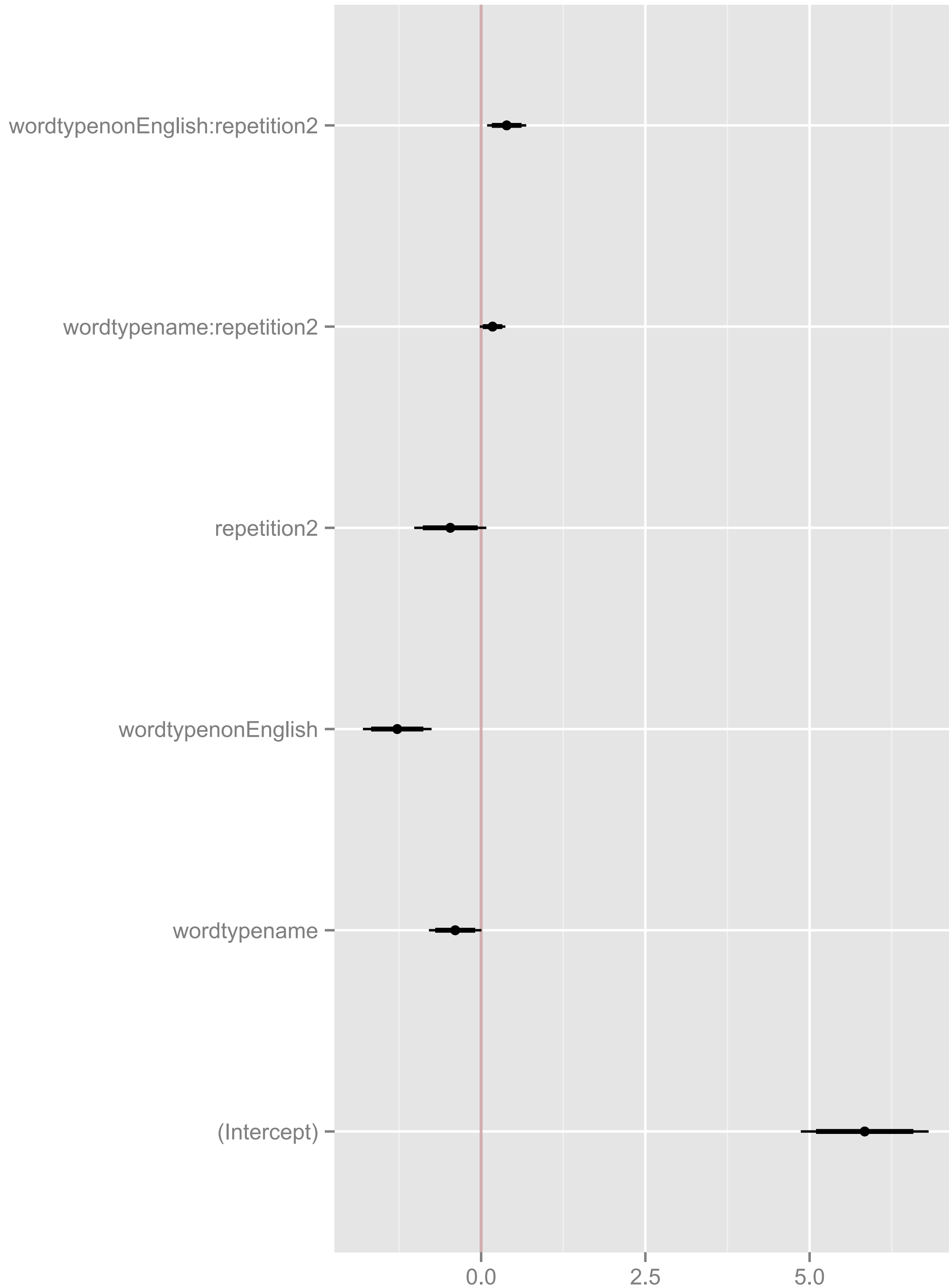


Figure 3.2: **Coefficient plot for the predictors of the hierarchical linear model for fingerspelling rate** Thick lines represent 95% confidence, thin lines 99% confidence, and dots are the estimates of the coefficients (or intercept).

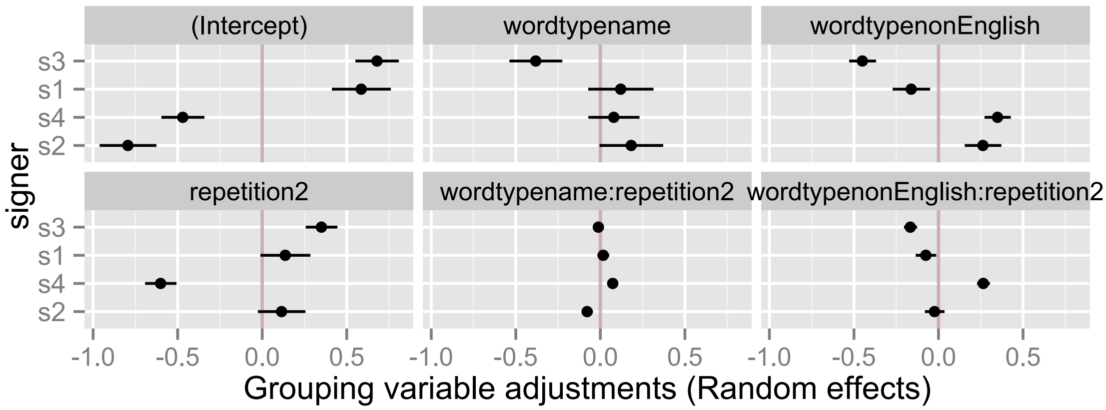


Figure 3.3: **Plot of intercept adjustments (random intercepts) for signer, as well as slope adjustments (random slopes) for word type, repetition, and their interaction of the hierarchical linear model for fingerspelling rate** As discussed in detail above, there is a large amount of intersigner variation (seen in the intercept facet), additionally, there is some variation among signers with respect to the effects of word type, repetition, and their interaction. The levels on the y-axis are signers, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

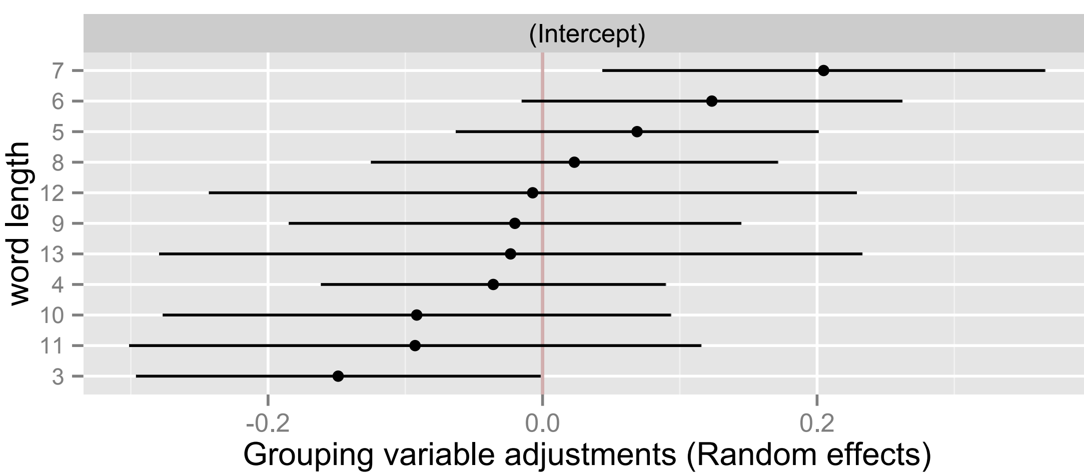


Figure 3.4: **Plot of intercept adjustments (random intercepts) for length of word of the hierarchical linear model for fingerspelling rate** As discussed in detail above, there is not much systematic variation of rate between word lengths. The levels on the y-axis are the word lengths, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

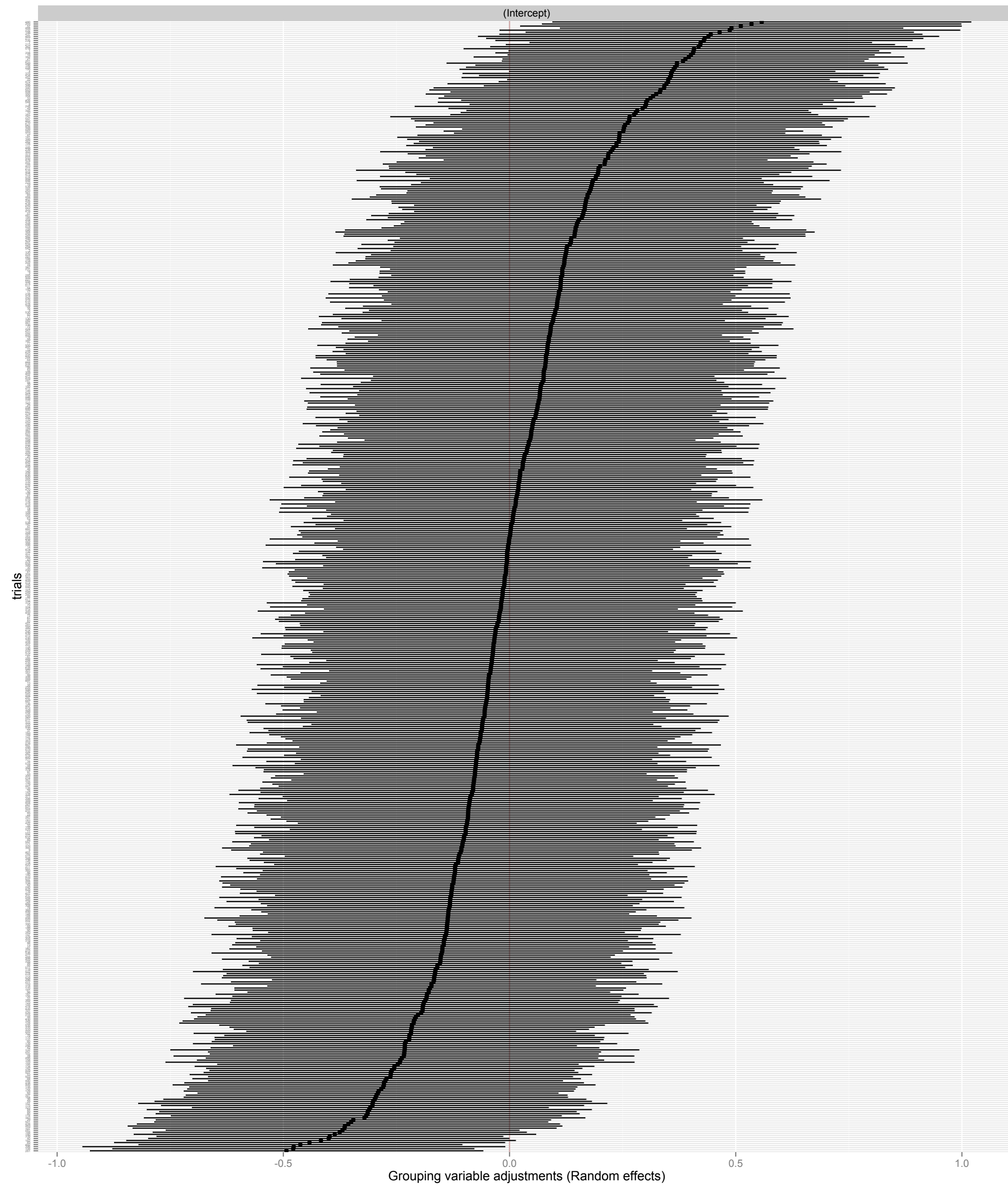


Figure 3.5: **Plot of intercept adjustments (random intercepts) for trials of the hierarchical linear model for fingerspelling rate** Because there are a large number of trials, there are many levels on the y-axis. Although it is difficult to read individual trials, as discussed in detail above, there is not much systematic variation of rate between trials. The sigmoidal shape is due to the fact that the intercept adjustments are modeled on a normal distribution. The levels on the y-axis are trial (numbers), and they are ordered by the magnitude of the intercept adjustment from smallest on the bottom to largest on the top.

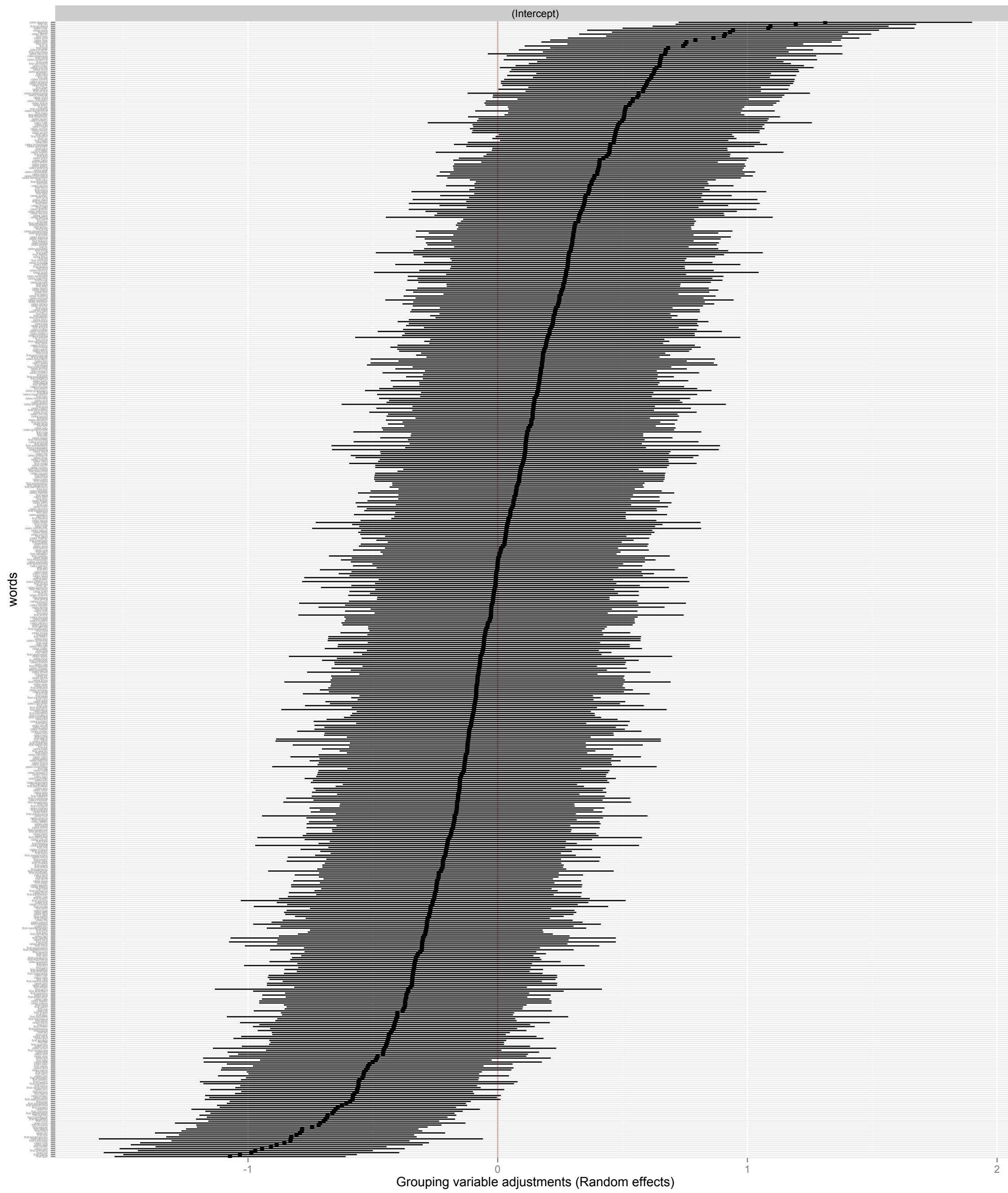


Figure 3.6: **Plot of intercept adjustments (random intercepts) for words nested in word lists of the hierarchical linear model for fingerspelling rate** Because there are a large number of words, there are many levels on the y-axis. Although it is difficult to read individual words, as discussed in detail above, there is not much systematic variation of rate between words. The sigmoidal shape is due to the fact that the intercept adjustments are modeled on a normal distribution. The levels on the y-axis are words (with the word list prefixed to them, to show the nested structure), and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

In conclusion, the rate findings here align with those that have been found previously: the overall rate for English words is 5.84 letters per second (171 msec/letter), right in the middle of the rates reported in previous studies. non-English words are fingerspelled at a slower rate 5.44 letters per second, replicating the findings of (Hanson, 1982), although this difference is reduced in second repetitions. There is a large amount of intersigner variation: with a range from 6.52 (signer 3) to 5.50 (signer 2) letters per second. There was not systematic variation in rates based on the length of the word. The only previous study that found a significant effect of word length was (Quinto-Pozos, 2010), which defined short words as 3 letters or less, and long words as more than 3 letters. In our data 3 letter words might be slightly slower (see figure 3.4), but the difference is not big enough to have confidence in, and this effect does not hold as words get longer. This says that word length is not the main (or even a systematic) driver of rate, but rather, other factors like word type (English vs. non-English) or intersigner variation.

3.4 Hold duration

We now move on to measuring individual apogees within each word. First, we will look at the duration of holds (duration of transitions will be discussed in section 3.5). As some have noted (including Reich & Bick (1977), for Visual English) the first and last apogees in fingerspelled words are frequently held for longer. Beyond that, many (including Wilcox (1992)) remark that the rest of the apogees are impressionistically rhythmic, with each taking about the same amount of time to execute. Because of this pattern, in addition to the fact that there are transitions between each hold, it is difficult to generalize from rate calculations to individual hold durations in a reliable manner. For example, depending on how much longer the first and last apogees are they will have a disproportionate impact on the rate by lowering it more for smaller words (where the beginning and end represent a larger percentage of holds) than longer words. However, there have been two studies (Wilcox, 1992; Jerde *et al.*, 2003) that look at sub-word units, although neither explicitly look at the differences between holds at the edges of words and those in the middle. Of these, only Wilcox

(1992) measured holds versus transitions (again, this was with only two fingerspelled words, one was already well on its way to being a loan sign). He found that holds of 91 milliseconds and transitions lasted a mean of 314 milliseconds across both types, and for the fingerspelled B-U-T (as opposed to the loanword #TOO-BAD) holds had a mean of 104 milliseconds, and and transitions lasted for a mean of 319 milliseconds.

Using the annotation scheme described in section 3.2.2, we have precise timing annotations for each apogee within each fingerspelled production. The video was shot at 60 FPS, (technically, 59.94 FPS, to account for drop frames.) which means that each frame is 16.68 msec from the adjacent frames[13]. This means that it is not possible to detect differences that are shorter than 16.68 msec, as they cannot be recorded by our cameras. The annotation software that was used, ELAN, allows for annotations that are as small as 1 msec. Because the signal that was annotated has this lower bound of sensitivity, all annotations were aligned to correspond to frames rather than raw milliseconds as they were exported from ELAN. Because frames are closest to the measure, all of the hold and transitions models are fit with the outcome being frames rather than milliseconds. Translating from frames to milliseconds is simple, just multiply by 16.68. Additionally, this technological limit means that holds that are shorter than two frames cannot be measured accurately. The video frames could look exactly the same if a hold is 30 msec, 20 msec, or shorter, that is, it will appear to be a single frame hold (or an instantaneous apogee, that has its canonical shape only in a transitionary state). For this reason, this data will have an artificially large number of holds that are 1 frame or 16.68 msec (see figure 3.7). This artificial cutoff violates one of the assumptions of linear models: that the outcome be linear. In order to compensate for this, and ensure that these specific holds are not the ones responsible for driving the effects, all models have been fit with all of the data, as well as only the data including holds of 2 frames and longer. The model with all of the holds is reported here,

13. This assumes that the shutter on the camera is instantaneous, which is not quite accurate, but for the purposes of this explanation can be ignored. The shutter rate used was typically high, (~1 msec) in order to stop motion blur as much as possible.

and the additional models are reported in appendix C; the effects are by and large the same across all models.

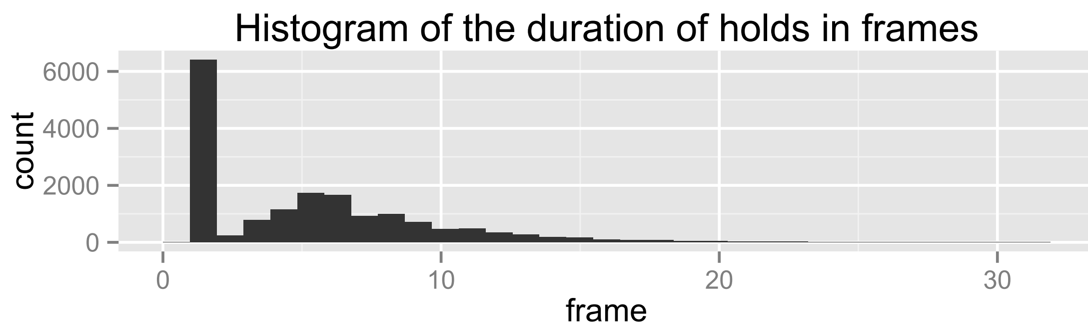


Figure 3.7: **Plot showing the distribution of hold durations in frames** Although the distribution looks normal centered around 5 frames, there is a large spike of holds (nearly 1/3 of all holds) at 1 frame.

3.4.1 Full words, including single-frame holds

For hold durations, the outcome was the number of frames of each hold in the word. Predictors are: the fingerspelling rate for the word, scaled and centered at zero (this was the outcome variable in the model in section 3.3); word type (using the same levels as before: noun (reference), name, and non-English); repetition with levels first (reference) and second; current apogee orientation or movement phonological group (abbreviated currGroup[14]) with levels default (reference), down (for FS-letters -P- and -Q-), movement (FS-letters -J- and -Z-), and side (FS-letters -G- and -H-); previous apogee orientation or movement phonological group (abbreviated prevGroup) with the same levels as currGroup; following apogee orientation or movement phonological group (abbreviated follGroup) with the same levels as currGroup; position of the apogee (abbreviated position), with levels 2–12, first, and last (where the first apogee is first, the second is 2, the third is 3, etc.). The final

14. The abbreviation currGroup is from current apogee's phonological orientation or movement group. This is to contrast with the previous apogee's phonological orientation or movement group (prevGroup) and following apogee's phonological orientation or movement group (follGroup).

apogee is always labelled as last, no matter what position it is in (for example, in the word T-A-X-I, -T- is 1, -A- is 2, , -X- is 3, and , -I- is last), and the first apogee is labelled first. As well as the interactions of rate × word type; word type × repetition; and the three-way interaction of rate × word type × repetition.

Predictors:

- rate
- word type
- repetition
- current apogee orientation or movement phonological group
- previous apogee's phonological orientation or movement group
- following apogee's phonological orientation or movement group
- position in the word
- *interaction* rate × word type
- *interaction* word type × repetition
- *interaction* interaction of rate × word type × repetition

The grouping factors for hold durations: intercept adjustments for signer (1, 2, 3, or 4), as well as slope adjustments for rate, word type, and repetition; intercept adjustments for word length; intercept adjustments for the FS-letter of the current apogee; intercept adjustments for the FS-letter of the previous apogee; intercept adjustments for the FS-letter of the following apogee; intercept adjustments for trial; and intercept adjustments for words, which are nested within wordlists.

Grouping factors:

- intercept adjustments for signer (1, 2, 3, or 4), as well as slope adjustments for

    - rate
    - word type,
    - repetition,

- intercept adjustments for word length
- intercept adjustments for current apogee FS-letter
- intercept adjustments for previous apogee FS-letter
- intercept adjustments for following apogee FS-letter
- intercept adjustments for trial
- intercept adjustments for words, which are nested within wordlists

Overall (for reference levels: mean rate, nouns, first repetition, with default orientation, in the first position) the hold duration is 3.60 frames (or 60 msec). There is a significant effect of rate, as the rate increases as the hold duration decreases. For every standard deviation slower the rate is, the hold is 5.58 frames longer (almost double the overall hold duration). There is a significant effect for word type: non-English words have shorter holds than nouns (1.28 frames shorter), additionally there is a trend that is not outside of the confidence intervals for names. There is a significant effect for phonological type of the current apogee: apogees that have movement are significantly longer than those with default orientations (7.39 frames longer); the effect is smaller, but also significant for the down and side orientations (1.40 and 1.41 frames longer respectively). There are no significant effects for the phonological group of the previous or following apogee, with the exception that there is a trend for a shorter current apogee if the following apogee has movement (again, outside of the 95% CI, but not the 99% CI). There is a significant effect of position: there are no strong relationships in the medial apogees, however the first apogee is significantly longer (1.99 frames) and the last apogee is significantly longer (6.04 frames longer). The interaction of rate and word type is only significant for the non-English words, where non-English words are held for even shorter when the rate increases, in other words, at high rates, the holds are shorter than predicted by the effect of word type alone. The interaction of rate and repetition is significant where apogees in second repetition have shorter holds than predicted by the effect of rate alone. Finally, neither the interaction between word type and repetition, or the three-way interaction of rate, word type, and repetition are significant. The model is visualized in figure 3.8 and full model output is in table 3.5.

Results:

- overall hold duration: 3.60 frames (or 60 msec)
- significant effects of:
  - rate
  - word type: non-English words differ from English
  - phonological type of the current apogee:
    * apogees that have movement are significantly longer than those with default orientations
    * the effect is smaller, but also significant for the down and side orientations
  - position: all medial positions are shorter than the first position, and the last apogee is significantly longer
  - the interaction of rate and word type (for the non-English words)
  - the interaction of rate and repetition

Grouping variable adjustments to intercepts and slopes are visualized in figures 3.9–3.15. Starting with signer (figure 3.9), we can see that there is a large amount of individual variation in the intercept adjustment: signer 3 has much longer holds, and signer 2 has much shorter holds than either of the other two signers, who are closer to the middle. For signer 3 and 4 the effect of rate is dampened, and for signers 1 and 2 it is magnified. For signer 4 the effect of names and non-English words is dampened, for signer 3 it is magnified, and for the other two it is unchanged. There is not a large amount of variation for the effect of repetition among the signers.

Word length does not appear to vary a large amount (see figure 3.10). Only 5-letter words do not overlap zero, and even for those, the magnitude is small. There is a lot of variation based on the FS-letter of the current apogee (see figure 3.11) the intercept for holds is adjusted considerably longer for FS-letters -X-, -C-, and -K- and adjusted considerably shorter for FS-letters -R-, -O-, and -E-. The level of FS-letter identity is, of course, nested within the phonological orientation/movement group. For that reason we have to look at the intercept adjustments for each phonological group separately: for the movement FS-letters, -Z- apogees have an intercept adjustment up, making them longer than -J- apogees. -G- apogees are adjusted up compared to -H- as well. For the down orientation, there does not seem to be a large difference between -P- and -Q-. The previous apogee FS-letter does not show large amounts of variation, with the exception of apogees with an -I- before them are shorter (see figure 3.12). The following apogee FS-letter does not show large amounts of variation, with the exception of apogees with an -S- after them are longer, and apogees with a -T- or -N- after them are shorter (see figure 3.13). Finally, trial and word do not show systematic variation (see figure 3.14 and 3.15 respectively).

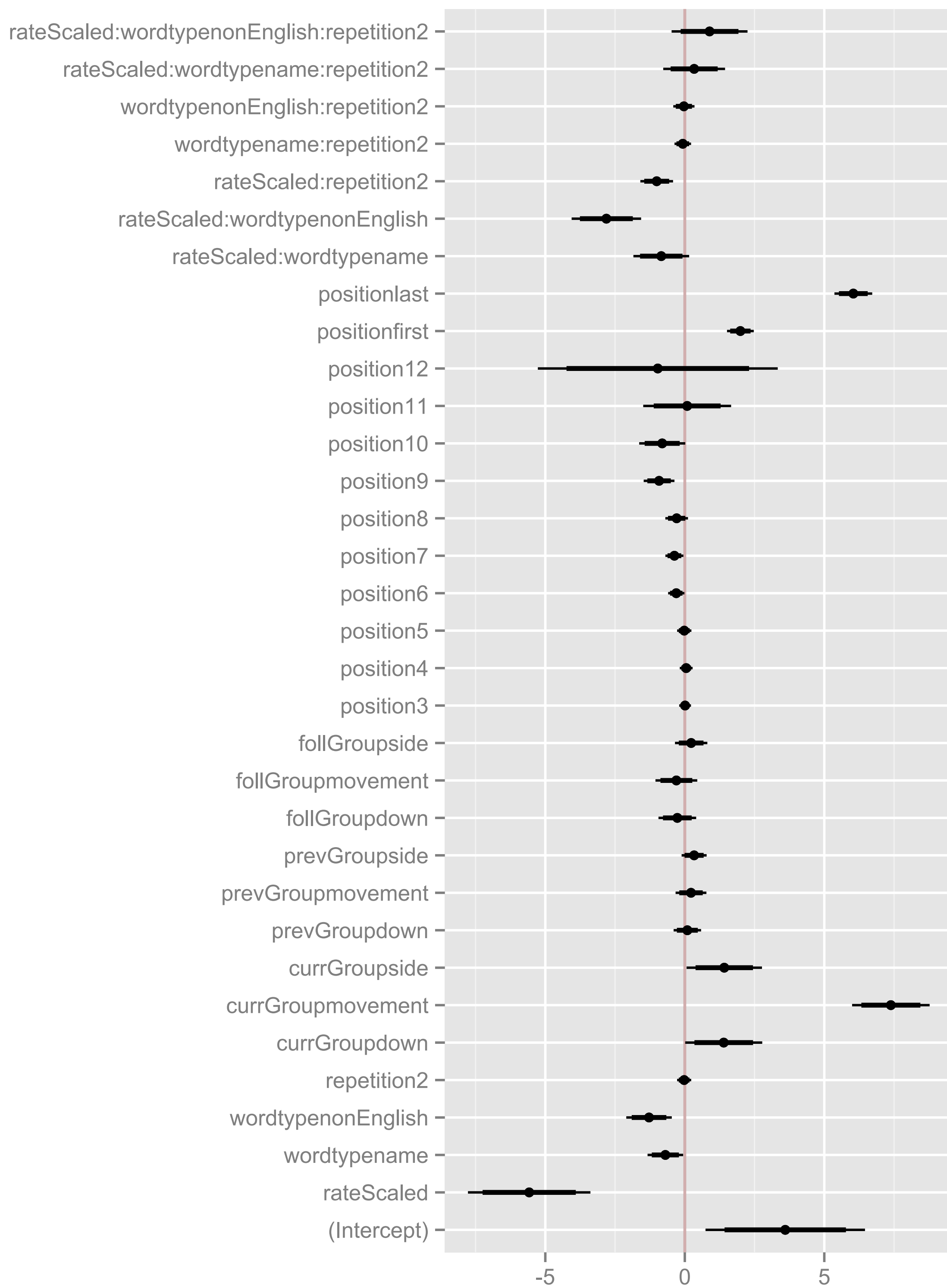


Figure 3.8: **Coefficient plot for the predictors of the hierarchical linear model for hold durations in full words, including single frame holds** Thick lines represent 95% confidence, thin lines 99% confidence, and dots are the estimates of the coefficients (or intercept).

	coefficient (standard error)
(Intercept)	$3.60(1.11)^{**}$
rateScaled	$-5.58(0.85)^{***}$
wordtypename	$-0.70(0.25)^{**}$
wordtypenonEnglish	$-1.28(0.32)^{***}$
repetition2	$-0.02(0.10)$
currGroupdown	$1.40(0.54)^{**}$
currGroupmovement	$7.39(0.54)^{***}$
currGroupside	$1.41(0.52)^{**}$
prevGroupdown	$0.09(0.19)$
prevGroupmovement	$0.22(0.21)$
prevGroupside	$0.33(0.17)$
follGroupdown	$-0.27(0.26)$
follGroupmovement	$-0.30(0.29)$
follGroupside	$0.23(0.23)$
position3	$0.01(0.08)$
position4	$0.05(0.09)$
position5	$-0.02(0.10)$
position6	$-0.31(0.11)^{**}$
position7	$-0.37(0.13)^{**}$
position8	$-0.29(0.16)$
position9	$-0.92(0.21)^{***}$
position10	$-0.81(0.32)^{*}$
position11	$0.08(0.61)$
position12	$-0.97(1.67)$
positionfirst	$1.99(0.19)^{***}$
positionlast	$6.04(0.26)^{***}$
rateScaled:wordtypename	$-0.84(0.39)^{*}$
rateScaled:wordtypenonEnglish	$-2.81(0.48)^{***}$
rateScaled:repetition2	$-1.01(0.23)^{***}$
wordtypename:repetition2	$-0.07(0.12)$
wordtypenonEnglish:repetition2	$-0.03(0.15)$
rateScaled:wordtypename:repetition2	$0.34(0.43)$
rateScaled:wordtypenonEnglish:repetition2	$0.89(0.53)$
AIC	84399.75
BIC	84825.40
Log Likelihood	-42144.87
Deviance	84289.75
Num. obs.	16967
Num. groups: wordList:word	577
Num. groups: trialWR	549
Num. groups: follLetter	27
Num. groups: prevLetter	27
Num. groups: apogeeLetter	26
Num. groups: lengthFact	11
Num. groups: signer	4
Variance: wordList:word.(Intercept)	0.82
Variance: trialWR.(Intercept)	0.43
Variance: follLetter.(Intercept)	0.06
Variance: prevLetter.(Intercept)	0.03
Variance: apogeeLetter.(Intercept)	0.48
Variance: lengthFact.(Intercept)	0.03
Variance: signer.(Intercept)	4.75
Variance: signer.rateScaled	2.71
Variance: signer.wordtypename	0.16
Variance: signer.wordtypenonEnglish	0.29
Variance: signer.repetition2	0.02
Variance: Residual	7.76

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3.5: Coefficient estimates and standard errors of the hierarchical linear model for full words, including single frame holds

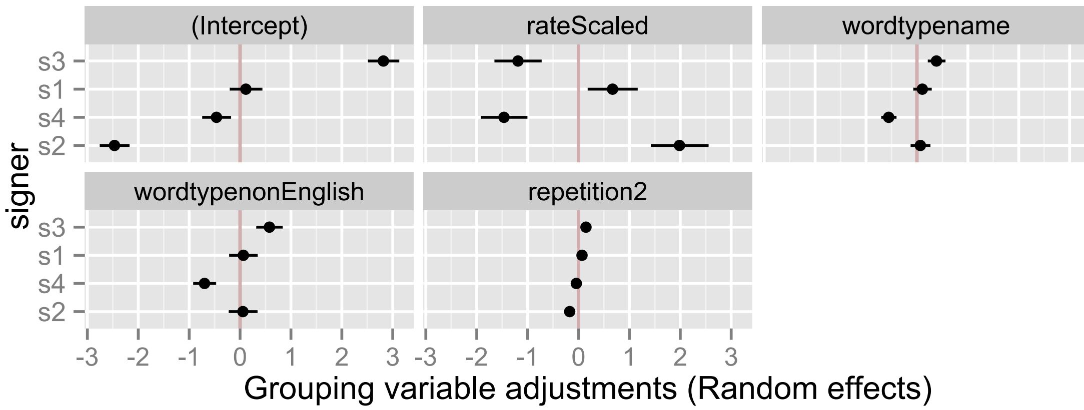


Figure 3.9: **Plot of intercept adjustments (random intercepts) for signer, as well as slope adjustments (random slopes) for rate, word type, and repetition of the hierarchical linear model for hold durations in full words, including single frame holds** As discussed in detail above, there is a large amount of intersigner variation (seen in the intercept facet), additionally, there is some variation among signers with respect to the effects of word type and repetition. The levels on the y-axis are signers, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

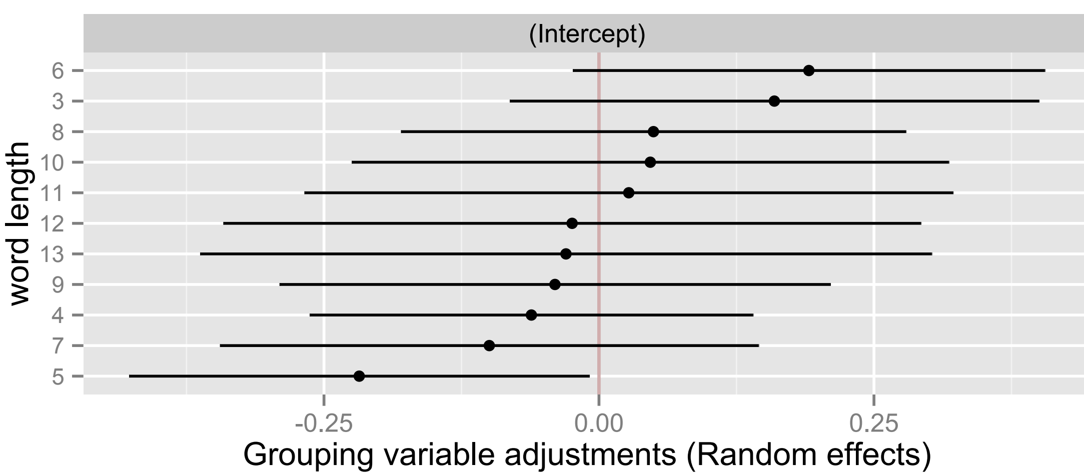


Figure 3.10: **Plot of intercept adjustments (random intercepts) for length of word of the hierarchical linear model for hold durations in full words, including single frame holds** As discussed in detail above, there is not much systematic variation of hold durations between word lengths. The levels on the y-axis are the word lengths, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

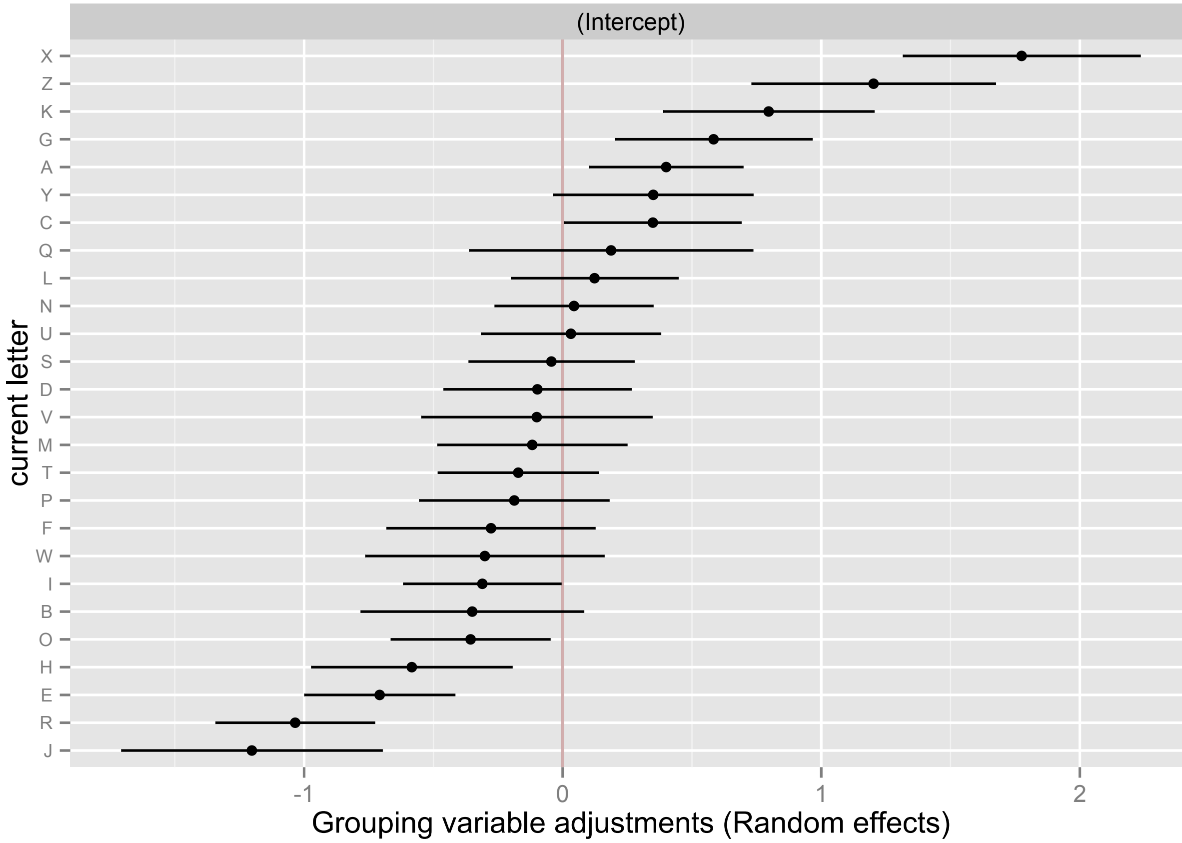


Figure 3.11: **Plot of intercept adjustments (random intercepts) for current FS-letter of the hierarchical linear model for hold durations in full words, including single frame holds** As discussed in detail above, some FS-letters are considerably shorter (-R-, -O-, and -E-) and some FS-letters are considerably longer (-X-, -C-, and -K-) than most other FS-letters. The levels on the y-axis are current FS-letters, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

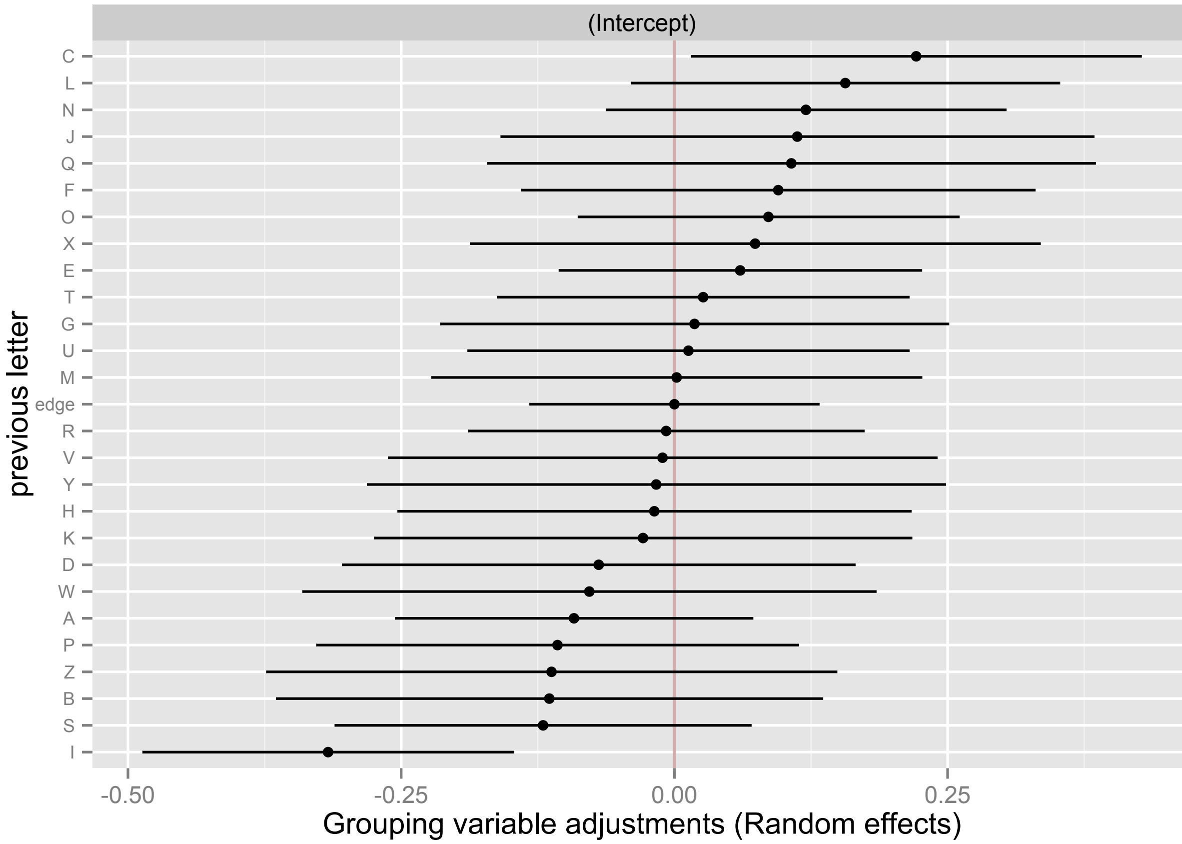


Figure 3.12: **Plot of intercept adjustments (random intercepts) for previous FS-letter of the hierarchical linear model for hold durations in full words, including single frame holds** As discussed in detail above, there is not much systematic variation of hold durations between previous FS-letters. The levels on the y-axis are previous FS-letters, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

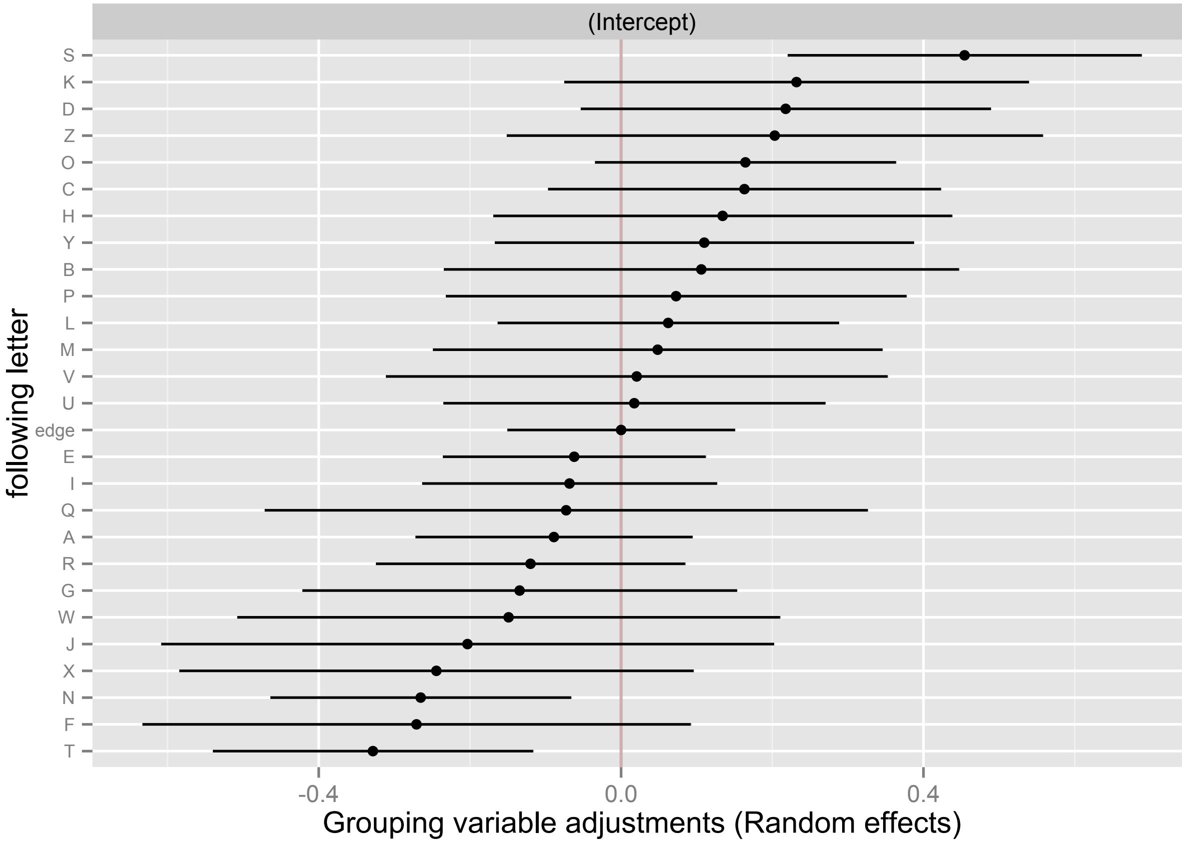


Figure 3.13: **Plot of intercept adjustments (random intercepts) for following FS-letter of the hierarchical linear model for hold durations in full words, including single frame holds** As discussed in detail above, some following FS-letters have considerably shorter current holds (-S-) and some following FS-letters have considerably longer current holds (-T- and -N-) than most other following FS-letters. The levels on the y-axis are following FS-letters, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

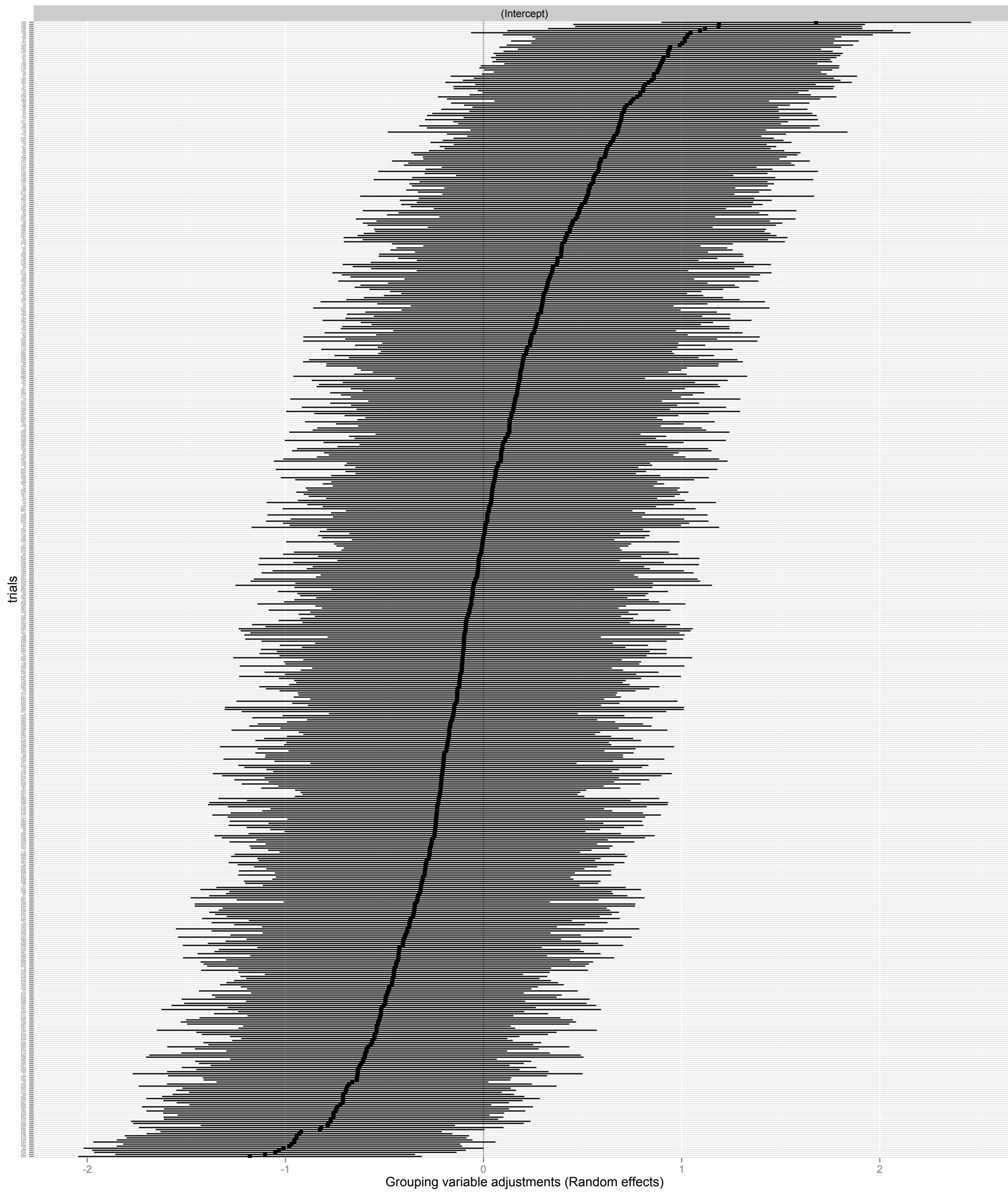


Figure 3.14: **Plot of intercept adjustments (random intercepts) for trials of the hierarchical linear model for hold durations in full words, including single frame holds** Although it is difficult to read individual words, as discussed in detail above, there is not much systematic variation of hold durations between trials. The sigmoidal shape is due to the fact that the intercept adjustments are modeled on a normal distribution. The levels on the y-axis are trial (numbers), and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

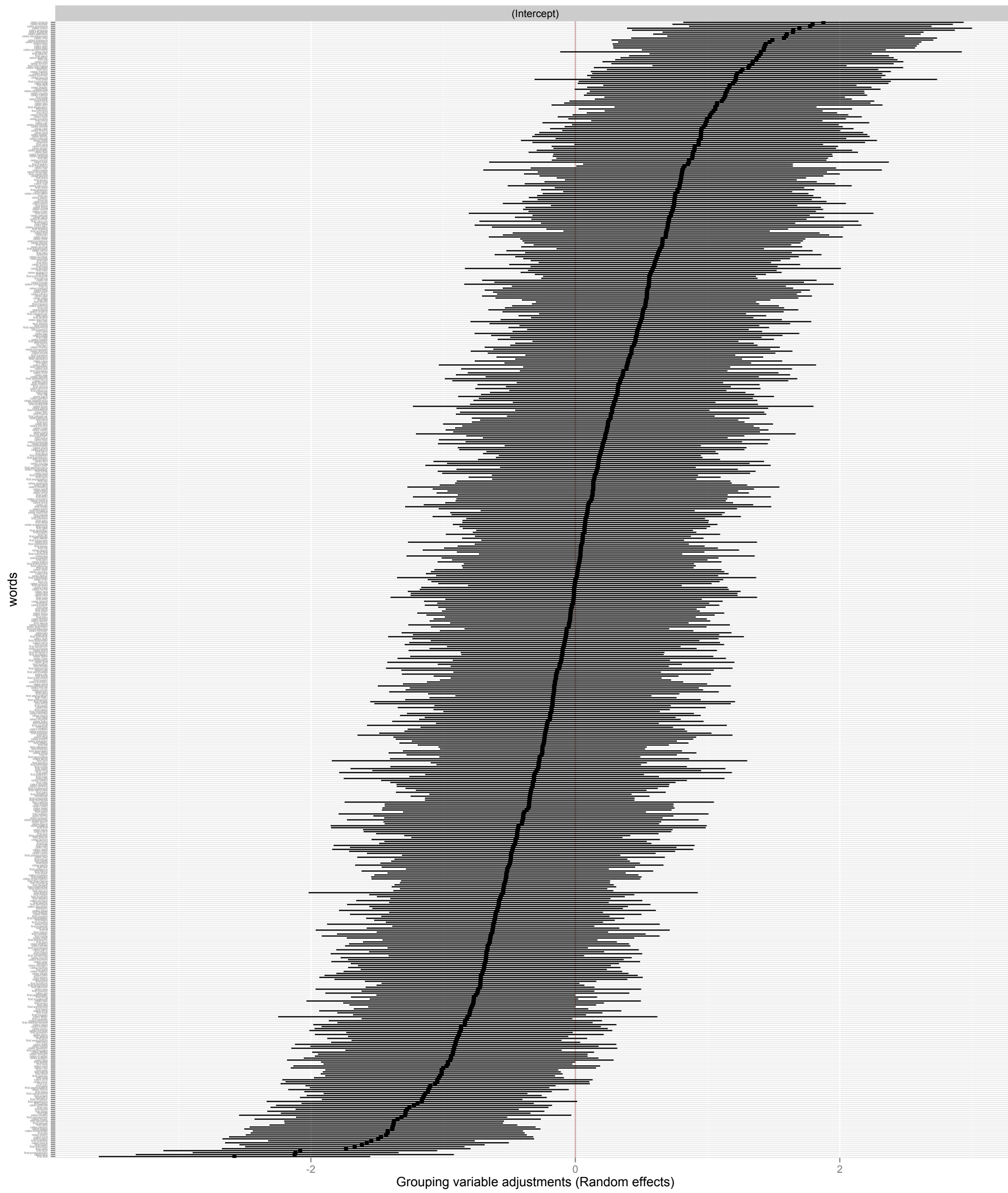


Figure 3.15: **Plot of intercept adjustments (random intercepts) for words nested in word lists of the hierarchical linear model for hold durations in full words, including single frame holds** Because there are a large number of words, there are many levels on the y-axis. Although it is difficult to read individual words, as discussed in detail above, there is not much systematic variation of hold durations between words. The sigmoidal shape is due to the fact that the intercept adjustments are modeled on a normal distribution. The levels on the y-axis are words (with the word list prefixed to them, to show the nested structure), and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

3.4.2 Conclusions concerning hold durations

Summarizing the holds model, the overall duration of a medial hold is 3.60 frames (or 60 msec). By far, the largest effect is the rate of fingerspelling: as the rate slows down, the holds get longer. Just a one standard deviation slower rate will produce holds that are double the duration of mean rate holds. There is an effect of position: initial or final position is longer: 5.59 frames (or 93 msec) for initials, and 9.64 frames (or 160 msec) for finals. There might be a slight trend for the medial holds to get shorter in longer words, but the effect is not robust across all models. Apogees with movement are held for longer than those without (by more than doubling the duration of the hold), and apogees with non-default orientations (down or side) tend to be held for longer, although the effect is smaller and is not significant in every model. The interactions of these effects are significant in only some of the models, and their magnitude is not particularly large when they do.

Additionally, there is large amount of signer variation: in each model signer 3 has the longest holds, and in all but one model signer 2 has the shortest. The difference between these two signers is between 4 and 6 frames on average. There is also some variation between individual FS-letters: for default orientations: -X- and -K- are typically longer than other FS-letters, and -R-, -E-, and -O- are shorter; for FS-letters with movement -Z- is held longer than -J-; for side orientation FS-letters -G- is sometimes longer than -H-; there is not much variation between the down orientation FS-letters -P- and -Q-.

All of these findings are in line with what has been found before: the overall durations are a bit shorter than those found by Wilcox (1992) (91 milliseconds), but that is not surprising given that the only word he measured was one that had three letters (B-U-T). If we add up estimates (from the models fit here) for first, medial, and last positions (93+61+160 msec) and divide by 3 apogees, we get an average of 103 msec, within 12 msec of what Wilcox found. Additionally, we replicate the finding from Reich & Bick (1977) that first and last positions are held for longer than medial positions. In addition to replicating these findings, we have found a huge effect of rate, as well as larger intersigner variation, and inter-FS-letter variation that is not accounted for by orientation or movement categories.

3.5 Transition durations

Although rate does have a large effect on hold durations, they are not perfectly correlated. This is because in addition to the holds, the fingerspelling sequence has transitions between each of the holds. We fit a model to transition durations in order to see if there were similar effects as hold on durations, as well as similar variation among signers. The model structure is similar to those above. The outcome was the number of frames of each transition in the word. Predictors are: the fingerspelling rate for the word, scaled and centered at zero (this was the outcome variable in the model in section 3.3); word type (using the same levels as before: noun (reference), name, and non-English); repetition with levels first (reference), and second; previous apogee orientation or movement phonological group (abbreviated prevGroup) with levels default (reference), down (for FS-letters -P- and -Q-), movement (FS-letters -J- and -Z-), and side (FS-letters -G- and -H-); following apogee orientation or movement phonological group (abbreviated follGroup) with the same levels as prevGroup; position of the transition (abbreviated position), with levels 1–12. As well as the interactions of rate × word type; word type × repetition; and the three-way interaction of rate × word type × repetition.

Predictors:

- rate
- word type
- repetition
- previous apogee's phonological orientation or movement group
- following apogee's phonological orientation or movement group

- position in the word
- *interaction* rate × word type
- *interaction* word type × repetition
- *interaction* interaction of rate × word type × repetition

The grouping factors for transition durations: intercept adjustments signer (1, 2, 3, or 4), as well as slope adjustments for rate, word type, and repetition; intercept adjustments word length; intercept adjustments for the FS-letter of the previous apogee; intercept adjustments for the FS-letter of the following apogee; intercept adjustments for trial; and intercept adjustments for words, which are nested within wordlists.

Grouping factors:

- intercept adjustments for signer (1, 2, 3, or 4), as well as slope adjustments for
  - rate
  - word type,
  - repetition,
- intercept adjustments for word length
- intercept adjustments for previous apogee FS-letter
- intercept adjustments for following apogee FS-letter
- intercept adjustments for trial
- intercept adjustments for words, which are nested within wordlists

Overall (for reference levels: mean rate, nouns, first repetition, with default orientation, in the first position) the transition duration is 8.33 frames (or 139 msec). There is no effect of rate. There is a significant effect of word type, both names and non-English words have longer transitions than nouns (longer by ~1 frame). There is a significant effect of the phonological group of the previous group: with down orientations, or movements in the previous apogee, the current transition will be longer. There are no significant effects for the phonological group of the following apogee. There is a significant effect of position: the immediately adjacent positions are not significantly different from each other, but positions that are 2 or 3 or more separated are (e.g. position 2 and 3 are not significantly different, but position 2 and 5 are) where later positions in the word have shorter transitions. The interaction of rate and word type is only significant for the non-English words, where non-English words have even shorter transitions when the rate increases, in other words, at high rates, the transitions are shorter than predicted by the effect of word type alone. The interaction of rate and repetition is not significant. Finally, neither the interaction between word type and repetition, or the three-way interaction of rate, word type, and repetition are significant. The model is visualized in figure 3.16 and full model output is in table 3.6.

Results:

- overall transition duration: 8.33 frames (or 139 msec)
- significant effects of:
  - word type: non-English words differ from English
  - phonological type of the previous apogee:
    * previous apogees that have movement are significantly longer than those with default orientations
    * the effect is smaller, but also significant for the down and side orientations in previous apogees
  - position: transitions shorten in later positions of a word
  - the interaction of rate and word type (for the non-English words)

Grouping variable adjustments to intercepts and slopes are visualized in figures 3.17–3.22. Starting with signer (figure 3.17), we can see that there is a large amount of individual variation in the intercept adjustment: signer 2 has much longer transitions, and signer 3 has much shorter transitions than either of the other two signers, who are closer to the middle. There is not a lot of variation among the signers on the effect of names. For signers 2 and 4 the effect of rate is dampened, and for signers 1 and 3 it is magnified. For signer 2 the effect of non-English words is dampened, for signer 1 it is magnified, and for the other two it is unchanged. There is not a large amount of variation for the effect of repetition among the signers.

Word length does not appear to vary a large amount (see figure 3.18). The previous apogee FS-letter shows some variation, with transitions following -Y- being longer[15], and transitions following -U-, -R-, -C-, -E-, -F-, and -O- being shorter (see figure 3.19). The following apogee FS-letter shows some variation, with transitions preceding -T-, -S-, and -U- being longer, and transitions preceding -X- and -W- being shorter (see figure 3.20). Finally, trial and word do not show systematic variation (see figure 3.21 and 3.22 respectively).

The overall transition duration, at 139 msecs, is shorter than those found by Wilcox (1992), who found transition durations of 314 milliseconds. Surprisingly, there is no effect of rate, which suggests that signers alter their rate by altering their hold durations rather than their transitions. The effect of previous movements and orientation changes suggest that the alignment of handshape and orientation changes or movement execution are timed to the beginning of the handshape holds rather than the the end of handshape holds.

Finally, the variation among signers' transition durations is surprising: signer 2 has much longer transitions, and signer 3 has much shorter transitions, which is the exact opposite pattern that is observed in the hold durations, where signer 3 has much longer holds, and signer 2 has much shorter holds. This shows that individual signers vary not only in overall rate, hold duration, and transition duration, but also in the ratios of holds to transitions. Study of many more signers is needed to see

15. This is consistent with the finding that -Y- frequently has a movement associated with it (Keane, 2010; Keane *et al.*, 2011).

how these differences relate to individual style differences, or possibly socio-cultural background, language exposure, language use, etc.

	coefficient (standard error)
(Intercept)	$7.93(1.53)^{***}$
rateScaled	$-2.94(2.10)$
wordtypename	$0.93(0.22)^{***}$
wordtypenonEnglish	$1.01(0.29)^{***}$
repetition2	$0.22(0.07)^{**}$
prevGroupdown	$1.16(0.38)^{**}$
prevGroupmovement	$1.56(0.38)^{***}$
prevGroupside	$0.37(0.37)$
follGroupdown	$0.49(0.42)$
follGroupmovement	$0.54(0.43)$
follGroupside	$-0.71(0.40)$
position3	$-0.27(0.07)^{***}$
position4	$-0.41(0.08)^{***}$
position5	$-0.72(0.09)^{***}$
position6	$-0.87(0.10)^{***}$
position7	$-1.27(0.11)^{***}$
position8	$-1.55(0.14)^{***}$
position9	$-2.02(0.19)^{***}$
position10	$-1.74(0.29)^{***}$
position11	$-2.21(0.56)^{***}$
position12	$-3.57(1.51)^{*}$
positionfirst	$0.40(0.07)^{***}$
rateScaled:wordtypename	$-0.49(0.42)$
rateScaled:wordtypenonEnglish	$-1.09(0.47)^{*}$
rateScaled:repetition2	$-0.23(0.23)$
wordtypename:repetition2	$-0.07(0.12)$
wordtypenonEnglish:repetition2	$-0.20(0.15)$
rateScaled:wordtypename:repetition2	$0.01(0.43)$
rateScaled:wordtypenonEnglish:repetition2	$0.06(0.53)$
AIC	67519.75
BIC	67897.26
Log Likelihood	-33709.88
Deviance	67419.75
Num. obs.	14047
Num. groups: wordList:word	577
Num. groups: trialWR	549
Num. groups: follLetter	26
Num. groups: prevLetter	26
Num. groups: lengthFact	11
Num. groups: signer	4
Variance: wordList:word.(Intercept)	1.19
Variance: trialWR.(Intercept)	0.75
Variance: follLetter.(Intercept)	0.26
Variance: prevLetter.(Intercept)	0.22
Variance: lengthFact.(Intercept)	0.09
Variance: signer.(Intercept)	9.13
Variance: signer.rateScaled	17.43
Variance: signer.wordtypename	0.08
Variance: signer.wordtypenonEnglish	0.20
Variance: signer.repetition2	0.00
Variance: Residual	6.31

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3.6: Coefficient estimates and standard errors of the hierarchical linear model for all transitions

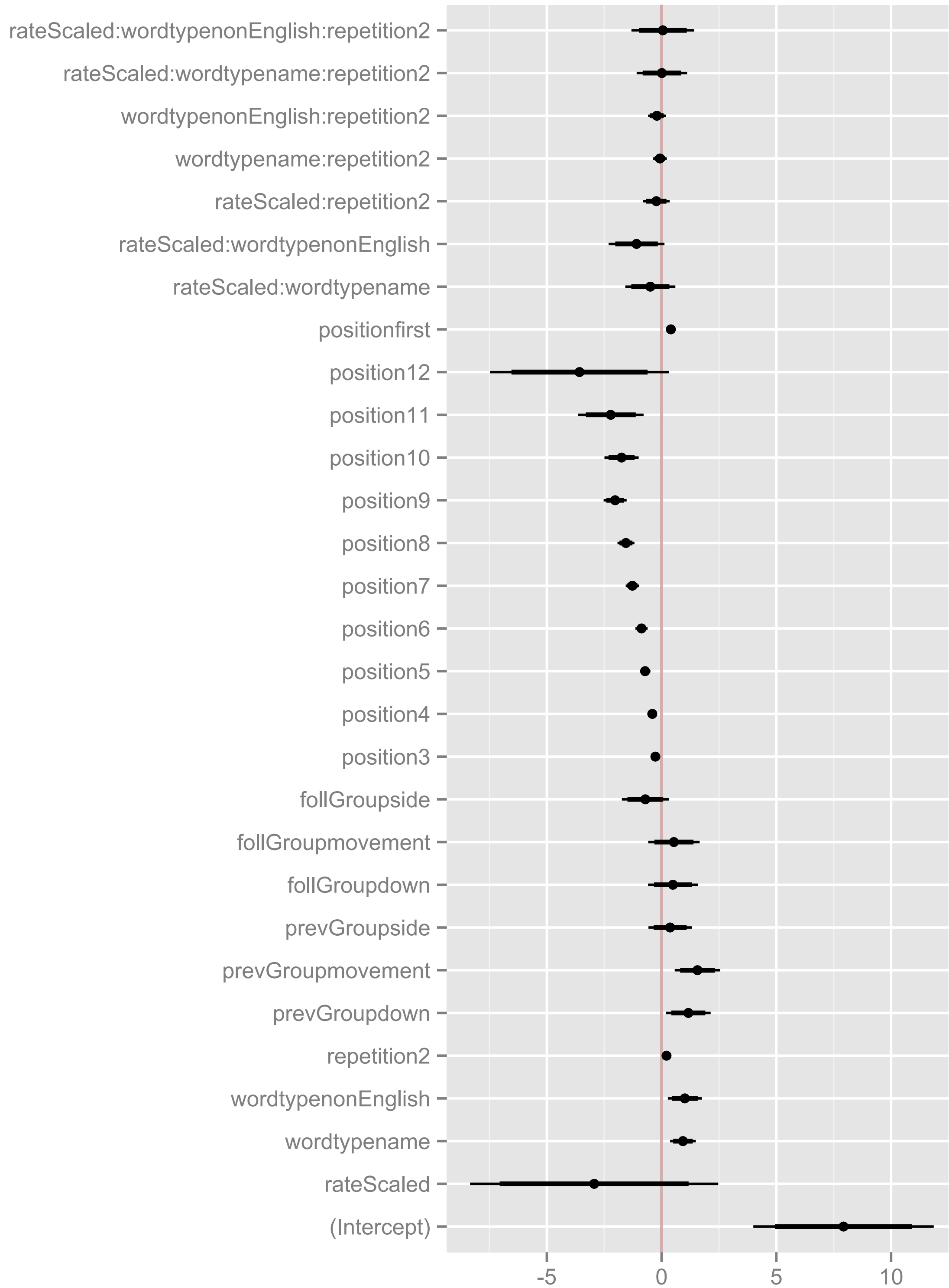


Figure 3.16: **Coefficient plot for the predictors of the hierarchical linear model for all transitions** Thick lines represent 95% confidence, thin lines 99% confidence, and dots are the estimates of the coefficients (or intercept).

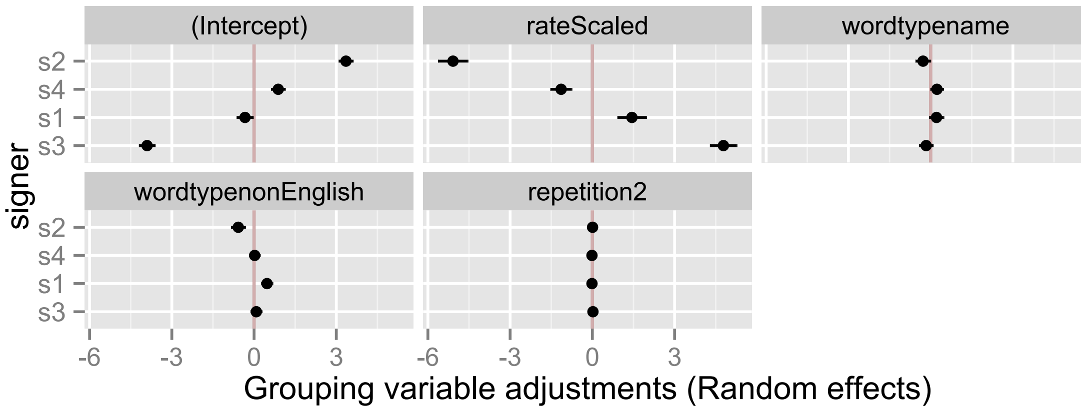


Figure 3.17: **Plot of intercept adjustments (random intercepts) for signer, as well as slope adjustments (random slopes) for rate, word type, and repetition of the hierarchical linear model for all transitions** As discussed in detail above, there is a large amount of intersigner variation (seen in the intercept facet), additionally, there is some variation among signers with respect to the effects of word type and repetition. The levels on the y-axis are signers, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

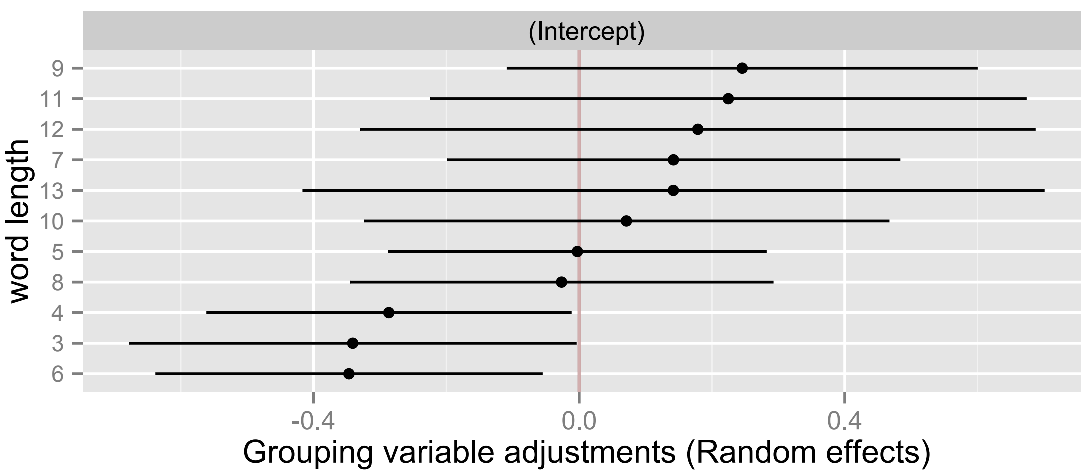


Figure 3.18: **Plot of intercept adjustments (random intercepts) for length of word of the hierarchical linear model for all transitions** As discussed in detail above, there is not much systematic variation of transition durations between word lengths. The levels on the y-axis are the word lengths, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

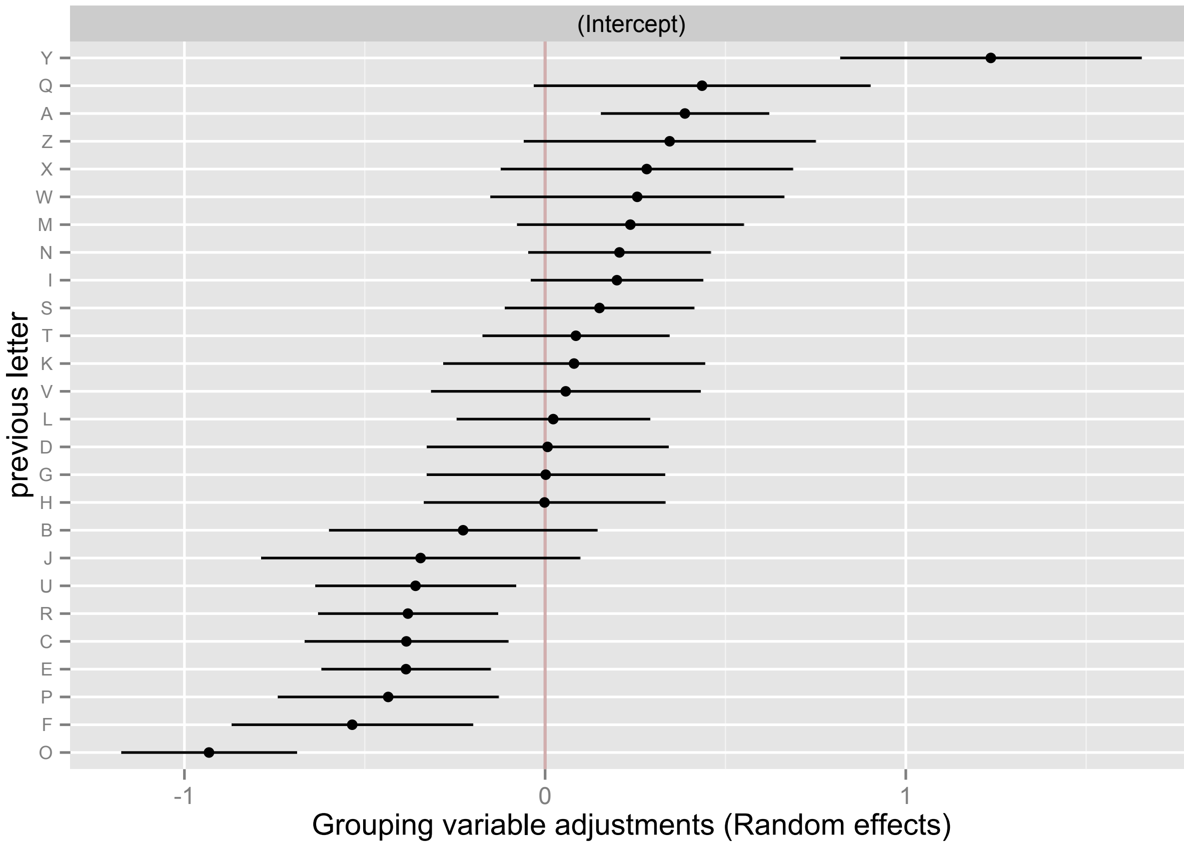


Figure 3.19: **Plot of intercept adjustments (random intercepts) for previous FS-letter of the hierarchical linear model for all transitions** As discussed in detail above, some previous FS-letters have considerably shorter transitions (-U-, -R-, -C-, -E-, -F-, and -O-) and some previous FS-letters have considerably longer transitions (-Y-) than most other previous FS-letters. The levels on the y-axis are previous FS-letters, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

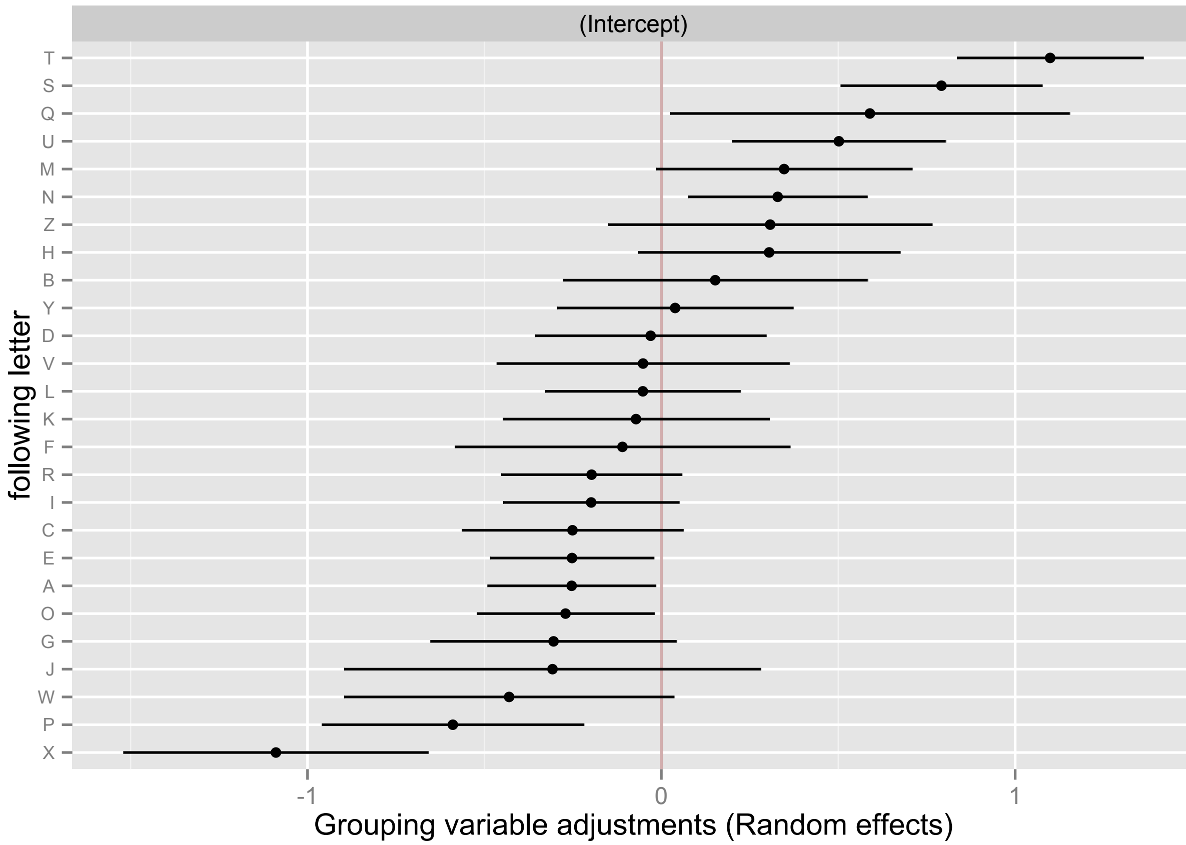


Figure 3.20: **Plot of intercept adjustments (random intercepts) for following FS-letter of the hierarchical linear model for all transitions** As discussed in detail above, some following FS-letters have considerably shorter transitions (-X- and -W-) and some following FS-letters have considerably longer transitions (-T-, -S-, and -U-) than most other following FS-letters. The levels on the y-axis are following FS-letters, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

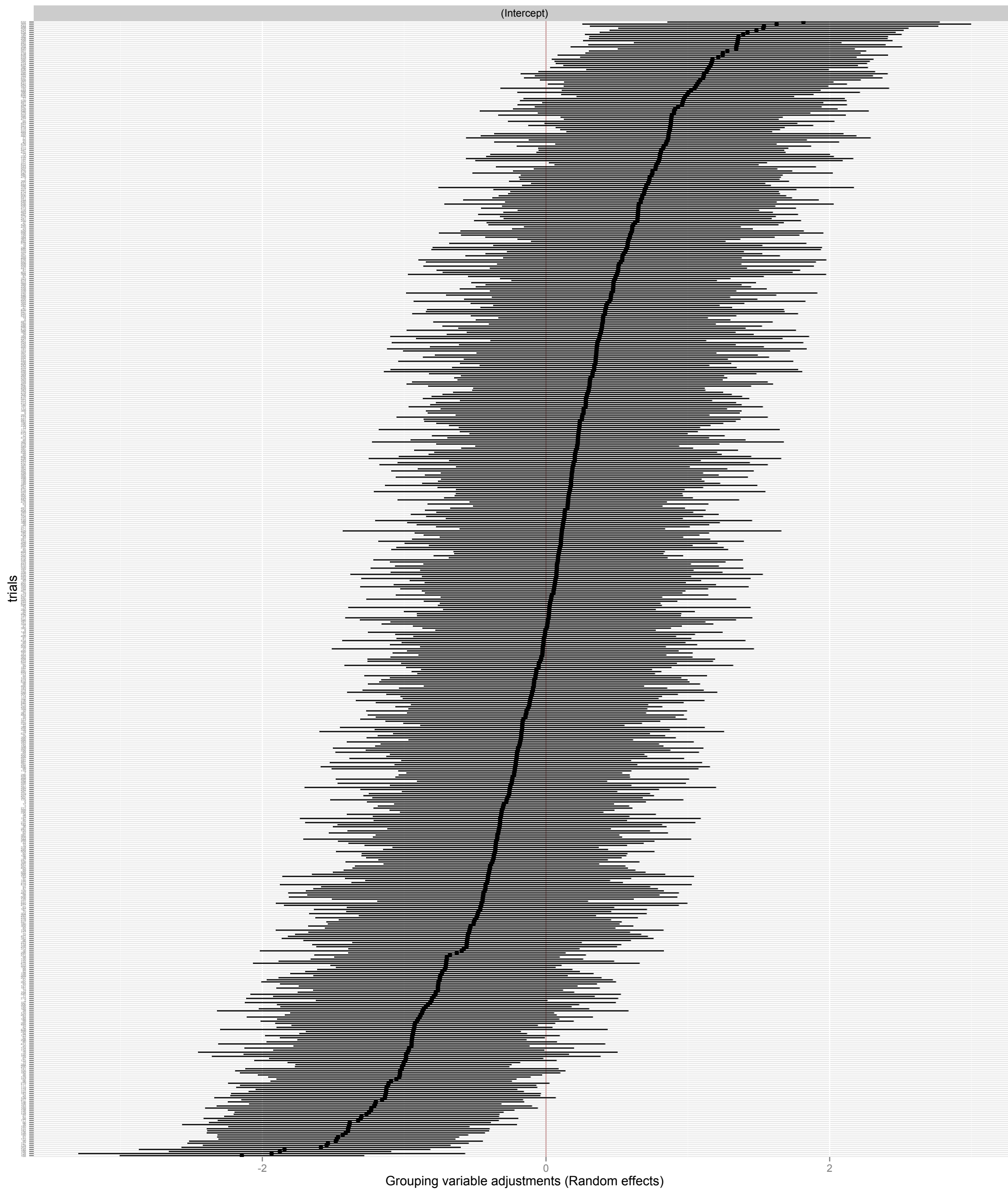


Figure 3.21: **Plot of intercept adjustments (random intercepts) for trials of the hierarchical linear model for all transitions** Because there are a large number of trials, there are many levels on the y-axis. Although it is difficult to read individual words, as discussed in detail above, there is not much systematic variation of transition durations between trials. The sigmoidal shape is due to the fact that the intercept adjustments are modeled on a normal distribution. The levels on the y-axis are trial (numbers), and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

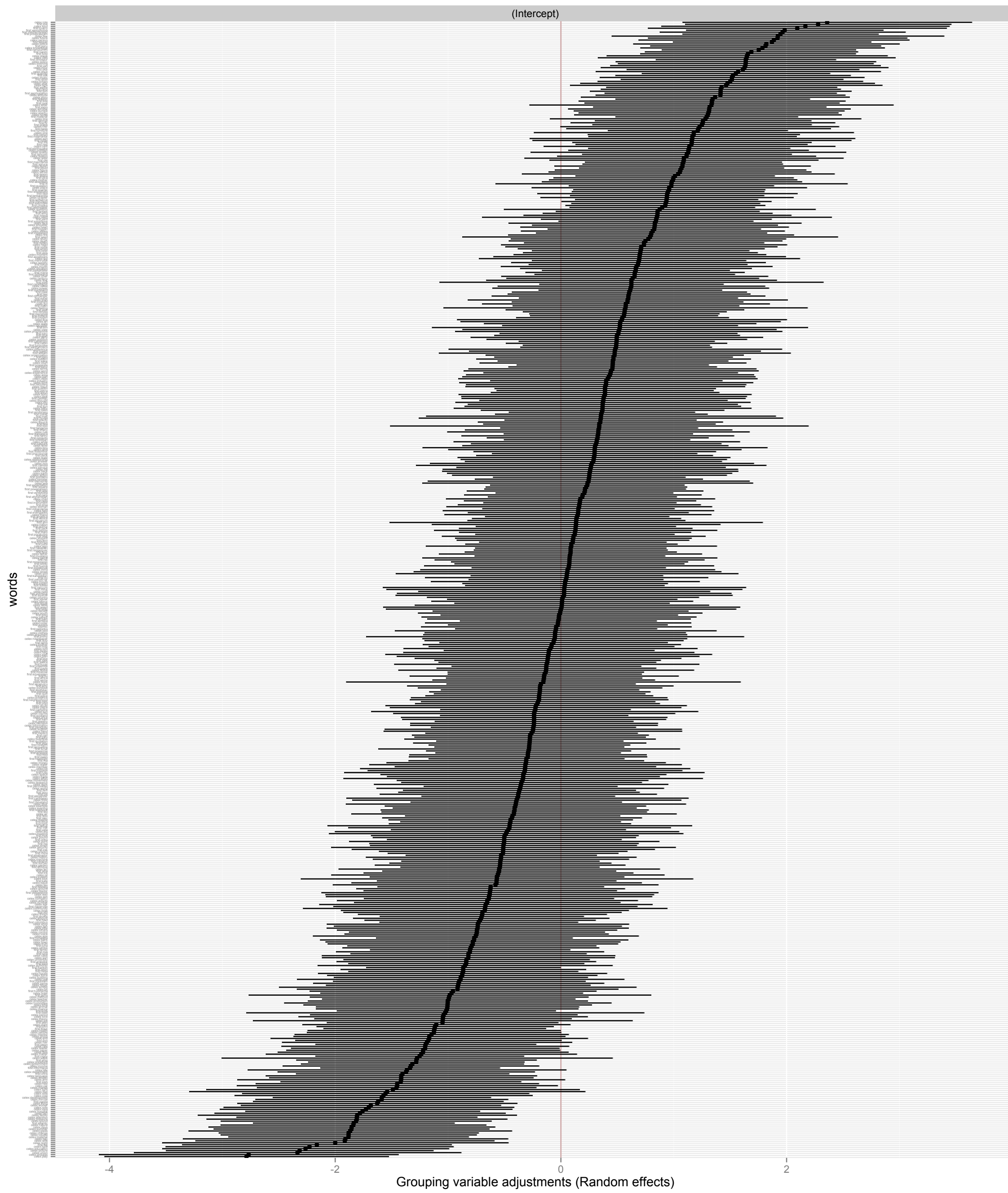


Figure 3.22: **Plot of intercept adjustments (random intercepts) for words nested in word lists of the hierarchical linear model for all transitions** Because there are a large number of words, there are many levels on the y-axis. Although it is difficult to read individual words, as discussed in detail above, there is not much systematic variation of transition durations between words. The sigmoidal shape is due to the fact that the intercept adjustments are modeled on a normal distribution. The levels on the y-axis are words (with the word list prefixed to them, to show the nested structure), and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

3.6 Motion capture, rate

3.6.1 Motion capture setup

Instrumented capture of articulators producing speech has been used for speech for a relatively long time. The technology (and, critically, application of that technology) to signed languages is relatively newer. A few researchers have used a variety of motion capture (and other) technologies to look at phonetic variation as well as other topics in ASL and other sign languages (Cheek (2001); Tyrone (2002); Tyrone & Mauk (2010); Tyrone *et al.* (2010); Mauk (2003); Mauk & Tyrone (2008); Jantunen (2013); Eccarius *et al.* (2012) among others). Typically motion capture has not been used for measuring handshape (with the notable exception of (Cheek, 2001)). One reason for this is that passive optical motion capture technologies are plagued by the problems of marker identification and swapping. For this reason the system used here is an active-marker optical system (the PhaseSpace motion capture system) that uses small LEDs that emit light at specific frequencies. This allows the system to positively identify each marker, without swapping. Using cameras positioned around the signer, the system can be used to track multiple articulators in rapidly changing handshapes, where passive optical systems simply could not.

Using this system, we have developed a protocol to collect data on the handshape of signers as they are signing (or fingerspelling.) The marker setup and protocol were designed to allow for the calculation of joint angles using a joint constrained inverse kinematic model. Although this work is ongoing, we have interim results for fingerspelling rate, which will be reported in the remainder of this chapter.

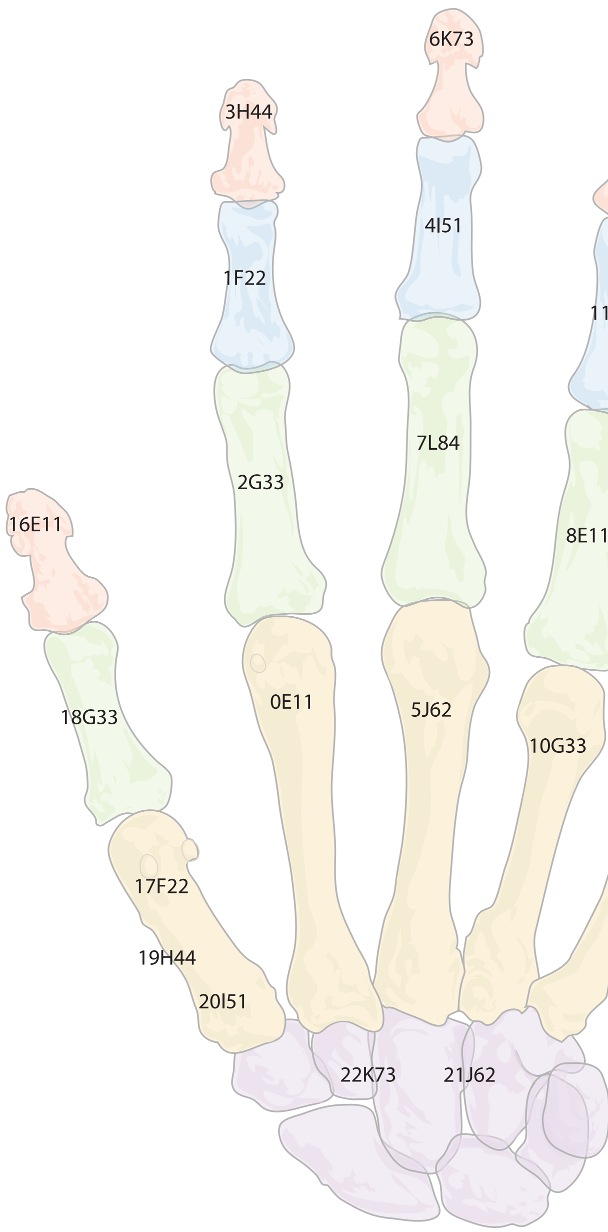


Figure 3.23: **Marker placement for motion capture data collection** Each marker is represented by a string of characters that represent its (one or two digit) ID number, one letter representing its position on the individual string, one number indicating its group for an 8 group setup, and one number indicating its group for a 4 group setup. For example, the marker that will be used most throughout the rest of the chapter (15L84) has a marker ID of 15, is in the L position on its string (of 8 markers), and is in group 8 for an 8 group setup, and in group 4 for a 4 group setup. Marker 23L84 was placed halfway up the forearm.

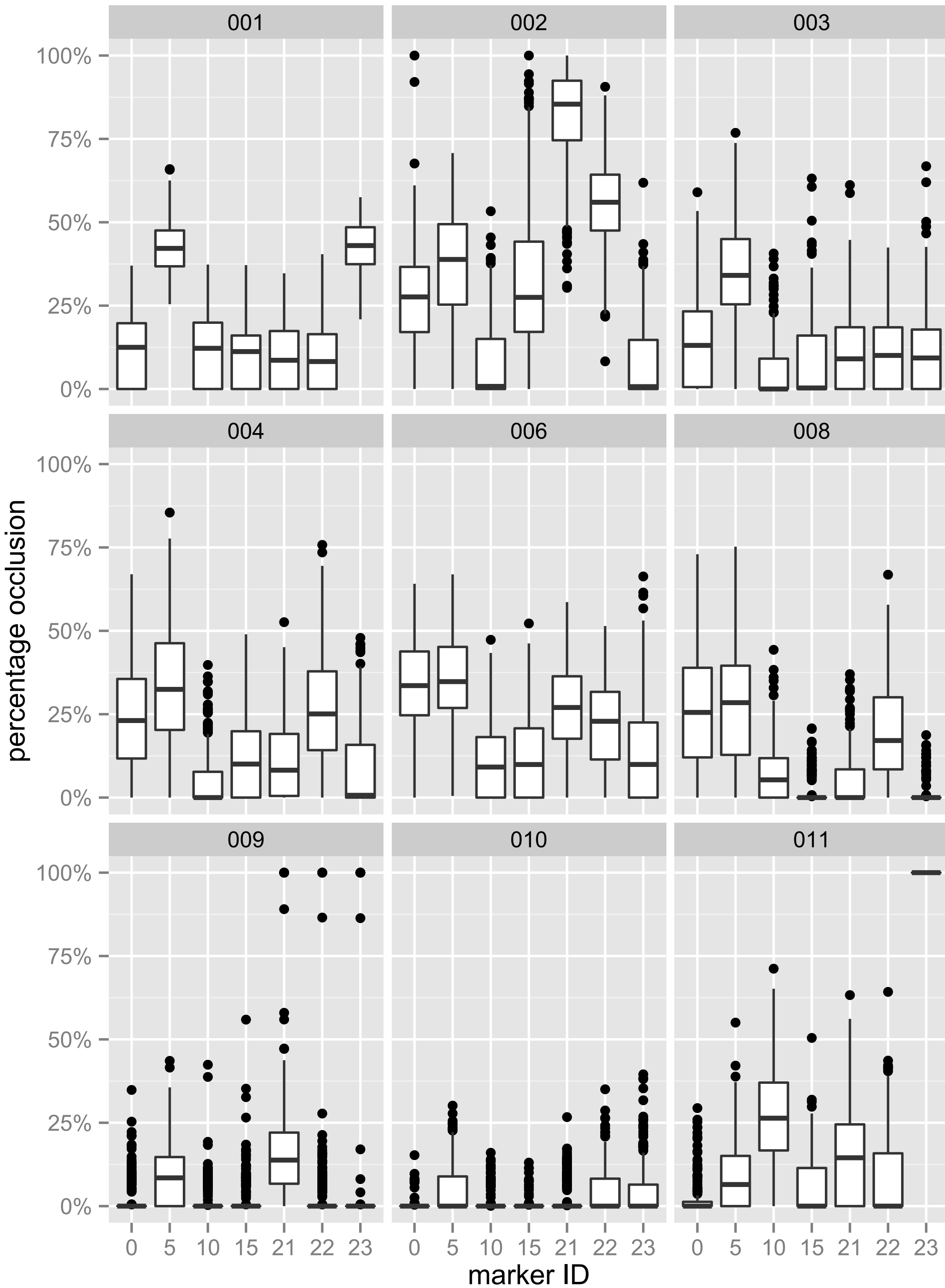


Figure 3.24: A series of boxplots showing the amount of occlusion for each marker on the back of the palm, separated by signers. The y-axis is the percentage of occluded frames where higher is more occlusion.

We have collected data from 11 signers, two were excluded from this analysis because they misunderstood the instructions, and fingerspelled at an exceptionally slow and deliberate rate throughout the data collection. This leaves us with 9 signers, all of who are native signers or earlier learners, and use ASL as their main mode of communication. Each signer was set up with 24 markers on their right (which was also their dominant) hand (see figure 3.23 for the marker setup). Each marker blinks at a unique frequency, which allows the system to positively identify it. Additionally there are groups that each marker is associated with. For any given frame, only one group is illuminated at a time. The more groups there are, the slower the effective capture rate is for each individual marker. The system operates at 480Hz, but for example, with 4 groups, each marker is sampled at 120Hz. Overall, occlusion was low for most signers (see figure 3.24). The camera setup was modified slightly starting with signer 007 (although that signer was excluded for other reasons), and maintained for the rest of the signers. The modifications were two: one camera was moved from the left side of the signer to the right (this meant there were fewer occlusions for all markers except for those on the thumb), the number of groups used during data collection was reduced from 8 to 4, allowing for an effective doubling of the rate of data acquisition (from 60Hz to 120Hz). These two things combined account for lower rates of occlusion for subjects 008, 009, 010, and 011. Stimuli were presented as printed words on a computer screen in front of the signers using PsychoPy (Peirce, 2007). The wordlist used was 186 words, which was a subset of the CELEX word list, as well as additional items designed to test the capabilities of the system for measuring orientation changes[16] (see appendix B.4 for the full wordlist). Otherwise, the data collection followed the same procedures as the video data collected, which was discussed above: each word was fingerspelled two times in a row, and the signer had the opportunity to self-correct by pressing a red button if they felt they had made a mistake.

16. Because it was a combination of two different word lists, there are some duplicate words in this wordlist, all duplicates were presented separately by the system, as if they were any other new word.

3.6.2 *Methods of analysis*

The first step in analyzing handshape data is to temporally segment the motion capture data into apogees, as was done with the video data as laid out in section 3.2.2 above. Because we have kinematic data, most of the (time consuming) human annotation can be replaced with methods used in speech recognition and signal detection to determine periods of holds based directly on the motion capture data. But, even before holds can be identified, we need to be able to identify the periods of the trials where the signer is fingerspelling. In the remaining part of this chapter, a few methods of detection of the period of fingerspelling for motion capture data will be proposed, compared, and then rate measures from the two most successful methods will be compared to the video data.

Because of the setup of the data collection (the buttons to advance the trials are placed at about desk height, and the signers are sitting down), the signer's hands are relatively low in the motion capture space during the periods of the trial when the signer is not fingerspelling. For this reason, a simple first approach to finding the periods of fingerspelling is to use the height of the marker (along the y-axis in this case) in the lab coordinate space[17]. Although this is not the only-axis where there is movement (there is also movement along the z-axis away from the button which is in front of the signer), the motion in the y-axis is the largest and most robust. Given this measure, there are six possible methods for determining when the hand is in the fingerspelling position (given in 1–6 below). These methods can be separated into two groups based on their approach; within each approach one method uses all of the data at once, one groups the data by signer, and the third groups the data by signer and trial. The first group (methods 1–3 below) use a simple threshold value based on some subset of the data: if the marker is above the threshold the signer is (considered to be) fingerspelling, if the marker is below they are (considered to be) not. The second group (methods 4–6 below) uses Hidden Markov Models (HMM) to learn and predict that difference between fin-

17. The lab coordinate space is defined with its origin on the floor, approximately below where the signer is sitting. The x-axis is (from the perspective of the signer, sitting at the origin) left (+) and right (-), the y-axis is up (+) and down(-), and the z-axis is front(+) and back(-) in the room. The stimuli display screen is near the front of the room, slightly offset from center to the right, so from the origin, positive z slightly negative x.

gerspelling and not fingerspelling. Hidden Markov Models are used extensively in the automatic speech recognition literature (see (Rabiner, 1989) for an introduction, (Gales & Young, 2008) for discussion of applications). At their core, they are models that allow us to predict something that is unobserved by something that we can observe. For our purposes here, what we want to predict is if the signer is or is not fingerspelling. What we can observe is (among other things) the height of the markers on the signers' hand. So, using an HMM, we would use the height of the hand as the observed variable, and the states that we want this to predict are fingerspelling or non-fingerspelling position.

1. **one threshold** Set a threshold value (defined here as in the highest 10% of values observed) for all trials across all signers, if the marker is above the threshold the signer is fingerspelling, otherwise they are not.

2. **threshold per-signer** Set a threshold value (defined here as in the highest 10% of values observed) for all trials but for each individual signer, if the marker is above the signer-specific threshold the signer is fingerspelling, otherwise they are not.

3. **threshold per-trial**[18] Set a threshold value (defined here as in the highest 10% of values observed) for each trial and each signer, if the marker is above the trial and signer specific threshold the signer is fingerspelling, otherwise they are not.

4. **one HMM** Use a two state Hidden Markov Model that has been trained across all trials and all signers. The state that is higher is the fingerspelling state.

5. **HMM per-signer** Use a two state Hidden Markov Model that has been trained across all trials for each individual signer. The state that is higher is the fingerspelling state.

18. The grouping here is technically by signer and by trial, because each trial is associated with one and only one signer.

6. **HMM per-trial**[19] Use a two state Hidden Markov Model that has been trained across for each trial and each signer. The state that is higher is the fingerspelling state.

The first five methods were used with this data. The last one was attempted, but was not successful because of technical limitations. The `RHmm` package that was used to fit the HMMs could not accept training data that included frames with occlusion (and thus had NA values for the marker). Because of this, when training the HMM, trials where a given marker was occluded at all during the trial were excluded. For the one HMM method, and the HMM per-signer methods, this reduction in training data is not a concern, because there are many trials in both of these groups with no occlusion for a given marker. The fit HMM can then be used to predict states based on data for all trials, even those with occlusion. For the HMM per-trial method, however, we would only have HMMs for the trials that had no occlusion, and so could only get predictions for those trials, which would vastly limit the number of trials we have results for.

Each of these methods produce data where (for each marker) each frame is categorized as either in the fingerspelling state or in the non-fingerspelling state. From this data: a quick measure of success is how many trials does each method correctly predict that the trial starts with a non-fingerspelling state, stay in that state for a number of frames, and then there is a single fingerspelling state (that also has a duration of multiple frames), and then there is a non-fingerspelling that lasts for the rest of the trial. This measure is understandably coarse. For example: the signer could, in a given trial, move their hand to the fingerspelling state without actually starting fingerspelling, move their hand down to their lap, and then once again bring their hand up to the fingerspelling state to fingerspell the word. This trial would be counted as an error using this metric, even though the fingerspelling position finding methods are actually working like they should. Further work manually verifying the locations and durations of fingerspelling states with video of each session in the future will help refine this metric, although it is outside of the scope of this dissertation.

19. Again, the grouping here is technically by signer and by trial, because each trial is associated with one and only one signer.

We used this metric to evaluate the success of each of the methods: figure 3.25 shows, for each method and each marker the percentage of trials for each signer that it identifies as having the sequence non-fingerspelling state, fingerspelling state, non-fingerspelling state. The single threshold method (labeled [markerID]Thresh) has little success. For most signers and markers it is near 0% identification. The threshold per-signer (labeled [markerID]subjThresh) does not do much better. The threshold per-trial (labeled [markerID]trialThresh) is more successful, with around 25% identification (with considerable variation between signers, and higher rates for signers 008–011, who had the optimized camera setup). Both the one HMM model (labeled [markerID]HMM) and the HMM per-signer model (labeled [markerID]subjHMM) were even more successful, with around 32% and 35% identification respectively. Again, there is considerable variation between signers, and higher rates for signers 008–011, who had the optimized camera setup. Some signers and markers have rates of identification as high as 92%. Of the different markers that were tested, the one that had the least occlusion was marker 15, which is the marker on the back of the hand, on the ulnar side, just below the metacarpophalangeal (MCP) joint on the pinky finger.

Based on these results, we have concluded that the one large HMM or by subject HMMs, based on the height of the marker on the ulnar side of the back of the hand, are fairly successful at identifying the fingerspelling position. This determination was based on the average percentage of correctly identified trials across all signers for a given method. As can be seen in figure 3.25, there is considerable variation across signers. For example, the one HMM model for signer 11 preforms considerably worse than other signers. This could be because signer 11 has a distinct signing style, signer physiology, or the signer was even sitting in a distinct position in the room that set them apart from the other signers. Further work is needed to determine what factors contribute to large differences like this. Additionally, combining identification methods by using more complex HMMs, or other methods will almost certainly produce more accurate identifications, but those are outside the scope of this dissertation.

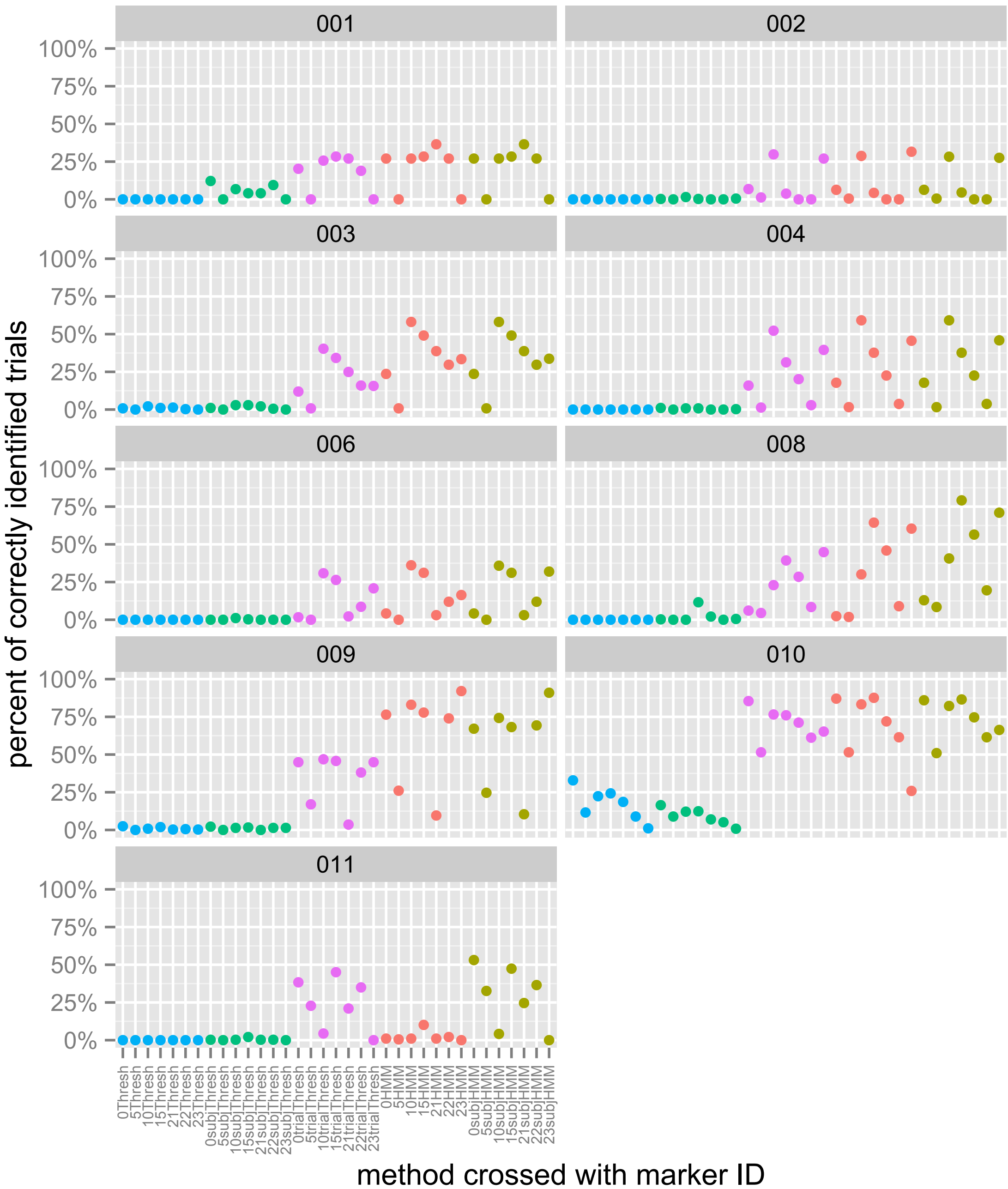


Figure 3.25: **A plot showing the percentage of trials with the pattern non-fingerspelling state, fingerspelling state, non-fingerspelling state for each marker and method, separated by signer** Each method is grouped together, and colored the same. The method groups are, from left to right: one threshold (blue), threshold per-signer (green), threshold per-trial (purple), one HMM (red), HMM per-signer (olive). The y-axis is the percentage of trials correctly identified out of total trials, where higher is more correctly identified trials. As discussed in the text above, the two methods that standout as the best are one HMM and HMM per-signer (for most markers, but especially marker 15 which has the least occlusion of all of the markers).

3.6.3 *Fingerspelling rate, as measured with motion capture data*

Rates from the one HMM for all signers model

We will now use the durations of the fingerspelling state for each successfully identified trial to measure rate of the 9 signers we have motion capture data for. The entire duration of the fingerspelling state was assumed to be the duration of the fingerspelled word. Rate was then calculated by dividing this duration by the number of letters in the word. We then fit a hierarchical linear regression (similar to those in section 3.3).

For the analysis of rate the outcome is the rate of fingerspelling (in letters per second). The predictor is only repetition with levels: first (reference) or second. The grouping factors are: intercept adjustments for signer (001–011), as well as slope adjustments for repetition; intercept adjustments word length; intercept adjustments for trial; and intercept adjustments for words.

Overall (for reference levels: the first repetition of a word) the rate is 5.39 letters per second. For the predictor in the model: there is no effect of repetition. The model is visualized in figure 3.26 and full model output is in table 3.7.

Grouping variable adjustments to intercepts and slopes are visualized in figures D.1–D.4 in the appendix. Starting with signer (figure D.1), we can see that there is a large amount of individual variation in the intercept adjustment: signers 002, 004, 001, and 010 have higher rates and signers 007, 009, 005, and 008 have slower rates with the other signers in the middle. There is not much difference for each signer with respect to the effect of rate. Word length does not appear to vary a large amount (see figure D.2). Only 7-letter words do not overlap zero, and even those, the magnitude is small. Finally, trial and word do not show systematic variation (see figure D.3 and D.4 respectively), although there are a handful of trials that seem to have very high rates.

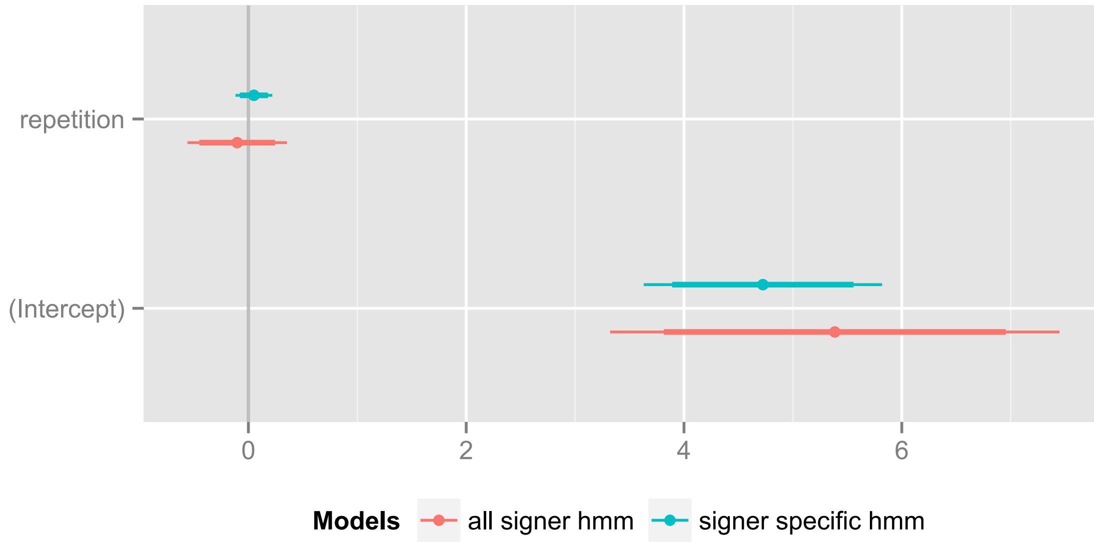


Figure 3.26: **Coefficient plots for the predictors of the hierarchical linear models for rates, using the all signer HMM model and using the signer-specific HMM model** Thick lines represent 95% confidence, thin lines 99% confidence, and dots are the estimates of the coefficients (or intercept).

	all signer hmm	signer specific hmm
(Intercept)	$5.39(0.80)^{***}$	$4.72(0.42)^{***}$
repetition	$-0.10(0.18)$	$0.05(0.07)$
AIC	7640.14	3975.90
BIC	7687.21	4024.01
Log Likelihood	-3811.07	-1978.95
Deviance	7622.14	3957.90
Num. obs.	1381	1550
Num. groups: trialWR	349	355
Num. groups: word	178	178
Num. groups: length	10	10
Num. groups: subj	9	9
Variance: trialWR.(Intercept)	1.96	0.17
Variance: word.(Intercept)	1.52	0.09
Variance: length.(Intercept)	0.35	0.17
Variance: subj.(Intercept)	4.51	1.41
Variance: subj.repetition	0.07	0.02
Variance: Residual	11.67	0.54

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3.7: Coefficient estimates and standard errors of the hierarchical linear model for sign rate, with motion capture data using both the all signer HMM model and the signer-specific HMM model

Rates from the signer-specific HMM model

This model is the exact same as the one before, but uses the predictions from the signer-specific HMM model.

Overall (for reference levels: the first repetition of a word) the rate is 4.72 letters per second. For the predictor in the model: there is no effect of repetition. The model is visualized in figure 3.26 and full model output is in table 3.7.

Grouping variable adjustments to intercepts and slopes are visualized in figures D.5–D.8 in the appendix. Starting with signer (figure D.5), we can see that there is a large amount of individual variation in the intercept adjustment: signers 002, 004, 001, and 010 have higher rates and signers 007, 009, 005, and 008 have slower rates with the other signers in the middle. There is not much difference for each signer with respect to the effect of rate. word length does not appear to vary a large amount (see figure D.6). Only 7 letter words do not overlap zero, and even those, the magnitude is small. Finally, trial and word do not show systematic variation (see figure D.7 and D.8 respectively), although there are a handful of trials that seem to have very quick rates.

Overall, the rate of fingerspelling in the motion capture setup is 4.72–5.39 letters per second. Although close to it, this is slightly slower than the rate found for video: 5.84 letters per second. There could be a number of reasons for this (e.g. the signers recruited for the motion capture experiment just happened to fingerspell slower, or the marker setup on the hand made the signers fingerspell slower). But, impressionistically, after inspecting a few videos together with the fingerspelling states from different methods, it was clear that signers will frequently pause with a neutral handshape in the fingerspelling position before starting to fingerspell the word. This is also observed in the regular video data analyzed at the beginning of this chapter. But, with the regular video data, the beginning of the word was defined as the beginning of the initial hold, not when the signer put their hand in the fingerspelling position. This extra bit of time at the beginning of each word for the motion capture data will generate an artificially lower fingerspelling rate when compared with the manually annotated fingerspelling durations.

We would like to see if the difference in fingerspelling rates found above is due to a brief pause before the word being included in the motion capture word durations, but not in the regular video word durations. To do this we fit hierarchical linear models to the regular video data, aand the two methods for the motion capture data to predict the duration of the fingerspelled word (in seconds). For these models the outcome is duration of the word. The predictors are the repetition, the number of letters in the word, and their interaction (for the regular video data, we also included a predictor for word type). The grouping factors are: intercept adjustments signer (001–011), as well as slope adjustments for repetition and length of the word; intercept adjustments for trial; and intercept adjustments for words (for the regular video data this is nested within wordlist).

Word duration from video

For the regular video data, we get an overall word duration of -0.20 seconds. Although the negative intercept seems problematic, remember that this is the duration for words at reference levels for categorical predictors, and at zero for continuous predictors: for this model that means this is for nouns, in the first repetition, with a length of zero letters (which is of course, not actually an interpretable point because there is no such thing as a zero letter word). There is a significant effect of word length (0.21 seconds longer per letter). So, for a 4 letter noun in the first repetition the model predicts that the duration would be 0.64 seconds long. None of the other predictors or their interactions are significant. There is signer variation in the intercept, as well as the effect of word length, and there is little systematic variation by word or trial (see intercept adjustments in section D.2).

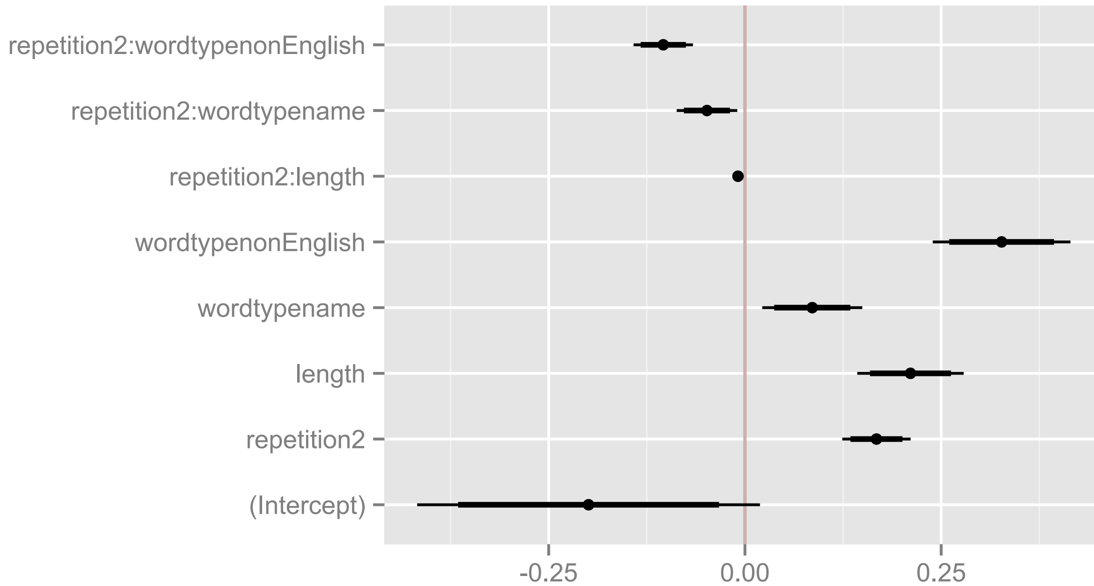


Figure 3.27: **Coefficient plot for the predictors of the hierarchical linear model for all word durations** Thick lines represent 95% confidence, thin lines 99% confidence, and dots are the estimates of the coefficients (or intercept).

	coefficient (standard error)
(Intercept)	$-0.20(0.08)^{*}$
repetition2	$0.17(0.02)^{***}$
length	$0.21(0.03)^{***}$
wordtypename	$0.09(0.02)^{***}$
wordtypenonEnglish	$0.33(0.03)^{***}$
repetition2:length	$-0.01(0.00)^{***}$
repetition2:wordtypename	$-0.05(0.01)^{**}$
repetition2:wordtypenonEnglish	$-0.10(0.01)^{***}$
AIC	-1515.16
BIC	-1389.60
Log Likelihood	778.58
Deviance	-1557.16
Num. obs.	2920
Num. groups: wordList:word	577
Num. groups: trialWR	549
Num. groups: signer	4
Variance: wordList:word.(Intercept)	0.01
Variance: trialWR.(Intercept)	0.00
Variance: signer.(Intercept)	0.03
Variance: signer.length	0.00
Variance: signer.wordtypename	0.00
Variance: signer.wordtypenonEnglish	0.00
Variance: Residual	0.02

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3.8: Coefficient estimates and standard errors of the hierarchical linear model for all durations

Word duration from motion capture

For the motion capture data with the one HMM for all, we get an overall word duration of 0.25 seconds. There is a significant effect of word length (0.20 seconds longer per letter). None of the other predictors or their interactions are significant. See figure 3.28 and table 3.9 for full details of both models. There is signer variation in the intercept, as well as the effect of word length, and there is little systematic variation by word or trial (see intercept adjustment visualizations in section D.3).

For the motion capture data with the signer-specific HMMs, we get an overall word duration of 0.23 seconds. There is a significant effect of word length (0.20 seconds longer per letter). None of the other predictors or their interactions are significant. See figure 3.28 and table 3.9 for full details of both models. There is signer variation in the intercept, as well as the effect of word length, and there is little systematic variation by word or trial (see intercept adjustment visualizations in section D.4).

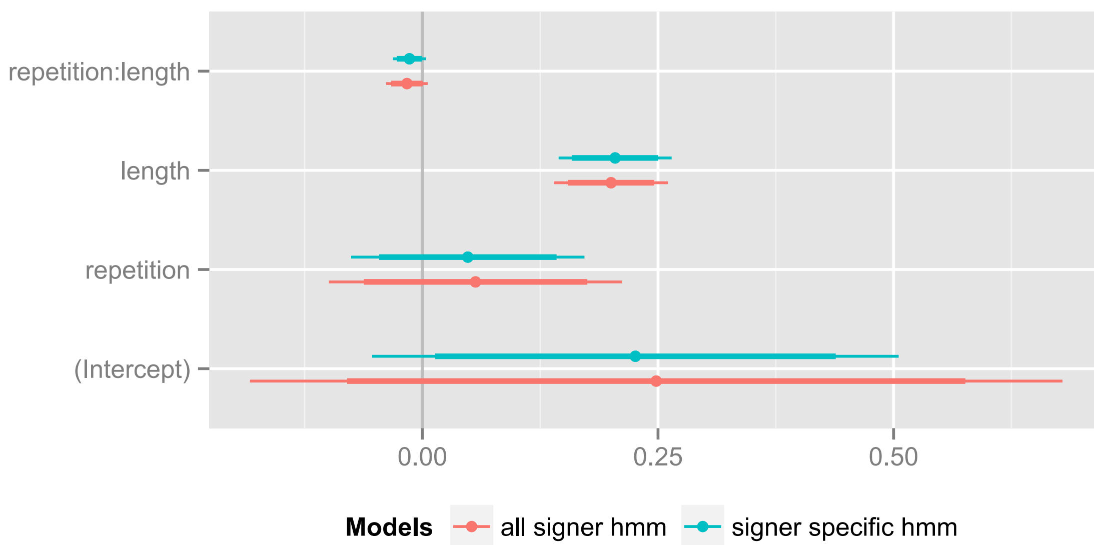


Figure 3.28: **Coefficient plots for the predictors of the hierarchical linear models for all word durations using motion capture data, using the all signer HMM model and using the signer-specific HMM model** Thick lines represent 95% confidence, thin lines 99% confidence, and dots are the estimates of the coefficients (or intercept).

	all signer hmm	signer specific hmm
(Intercept)	0.25(0.17)	0.23(0.11)*
repetition	0.06(0.06)	0.05(0.05)
length	0.20(0.02)***	0.20(0.02)***
repetition:length	−0.02(0.01)	−0.01(0.01)*
AIC	1467.84	1156.15
BIC	1535.84	1225.65
Log Likelihood	-720.92	-565.08
Deviance	1441.84	1130.15
Num. obs.	1381	1550
Num. groups: trialWR	349	355
Num. groups: word	178	178
Num. groups: subj	9	9
Variance: trialWR.(Intercept)	0.04	0.03
Variance: word.(Intercept)	0.02	0.02
Variance: subj.(Intercept)	0.16	0.05
Variance: subj.repetition	0.01	0.00
Variance: subj.length	0.00	0.00
Variance: Residual	0.12	0.08

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$

Table 3.9: Coefficient estimates and standard errors of the hierarchical linear model for word duration, with motion capture data using both the all signer HMM model and the signer-specific HMM model

Looking at these three models of word duration together we see something striking: The intercept for the two motion capture duration models are similar ~0.24 seconds, and the effect of letters is positive: for each additional letter the durations are 0.20 seconds longer. For the regular video data, the intercept is much lower (-0.16 seconds), but the effect of word length is almost exactly the same: for each additional letter the durations are 0.21 seconds longer. The similarity in the effect of letters suggests that each additional letter is contributing the same amount of time to the duration of fingerspelled words in both the regular video data as well as the motion capture data. The differences in intercept (with the motion capture intercepts being larger) can be explained by the brief pause with the hand in fingerspelling position before the signer starts the first hold that was observed impressionistically. That brief pause will add a little bit of duration to each word in the motion capture duration. This extra time is not related to how many letters there are in the word (or any of the other predictors). Although this is not a replacement for confirming that this is the case in all of the motion capture data either through manual annotation, or a robust model for finding the holds for fingerspelled words in the motion capture data, it is strong evidence that the difference in rate is not an underlying difference in the fingerspelling, but rather it is in artifact of the specific measures of duration that are used to calculate that rate.

3.7 Conclusions

Previous literature has a huge range of reports for the rate of fingerspelling production in ASL, anywhere from 2.18 to 6.5 letters per second. Most studies used fairly small data sets, and measured rate by measuring the duration of fingerspelling, and then dividing by the number of letters in the word. Using a large set of both regular video data as well as motion capture data, we have found an overall fingerspelling rate that is within this range, although it is a bit higher than the mean (5.84 letters per second). The rate from the motion capture data is slightly lower, although this appears to be an artifact of the specific measure of word duration including a brief pause before the fingerspelled word, as opposed to the strict beginning of the first hold to the end of the final hold measure used with the video data. This difference could be one possible explanation for some of the lower rates that have been reported in the literature. Detailed discussion of the coding methodologies are not available for each study, so this cannot be confirmed.

In addition, each individual hold associated with each apogee in each fingerspelled word was annotated for the regular video data, which allowed for sub-word duration analysis. Although this analysis is stretching current technological capabilities because fingerspelling is so quick, and current, non-specialized video cameras do not record faster than 60 FPS, we found a number of effects on hold duration. First, the rate of fingerspelling has a large effect on the duration of holds: the faster the rate, the shorter the holds. The first and last apogees are held for much longer than word-medial apogees. Of the word-medial apogees, holds tend to be the same duration with only slight differences between them. FS-letters with movement are held for longer. FS-letters with orientations that are down or to the side might be held for longer, although this is complicated by the alignment of handshape changes with orientation changes. Additionally there is a large amount of variation between signers for the overall duration of their holds. Transition durations vary more based on the orientation or movement of the apogee before them than after. This could be evidence that orientation and handshape changes are aligned to the beginning of the holds for apogees, as opposed to both the beginning and ends of holds. Transitions get considerably shorter in later positions in words. Finally, there is a large amount of variation between signers' transitions as well. Strikingly, the inter-signer variation for holds and transitions does not follow the same pattern: signers with long holds do not necessarily also have long transitions. Rather, there is considerable variation in the ratio of holds to transitions among different signers. This is counter to what has generally been assumed in the past, for example: "If the targets of fingerspelled words are only briefly achieved. then much of the time spent in fingerspelling is in transitional movements. If we ask which unit is likely to be more salient the targets or the transitions a reasonable answer would be that the temporally longer transitions may carry a substantial portion of the information in a fingerspelled word" (Wilcox, 1992, p59). We find that some signers do have fingerspelling that has relatively shorter holds and longer transitions, but other signers have fingerspelling with relatively longer holds and shorter transitions. This variation may be one of the features that people are (subconsciously) aware of when they describe different individual styles of fingerspelling. This variation might also explain the huge range of rates reported in the literature. Since most studies included only a few signers, it is not surprising, given the huge amount of variation, that there is a wide range of rates reported.

3.8 Looking forward

This work on the temporal properties of fingerspelling contributes to the field in a number of ways. First, it is part of a due diligence, basic description of a language phenomenon. No study before this has gone into as much detail, with this amount of data, to discover what the basic timing properties of fingerspelling are. This is intrinsically interesting and useful in work on developing automatic fingerspelling recognition tools, as well as automatic fingerspelling production (avatar) tools. Beyond these direct links, temporal information is important as a predictor in other linguistic analyses of fingerspelling (like that in chapter 4).

Concretely, findings from this timing data have been instrumental in additional studies that are ongoing (Keane & Geer, 2014; Geer & Keane, 2014). As discussed above, Wilcox (1992) proposed that transitions are more salient because they are temporally longer. As we have seen in this chapter, it turns out that this is not always true: some signers have a more balanced hold to transition ratio. This allowed for reevaluation of the idea that transitions are the most salient part of fingerspelling. In these (and other) ongoing studies, the temporal annotations described in this chapter were used to test which part of the signal allows for more successful fingerspelling perception (when the hold-transition ratio is approximately even): holds only, transitions only, or the full fingerspelling signal. When students of ASL are given stimuli from these groups, they are better at identifying fingerspelled words in the holds only and full signal conditions, than they are in the transitions only conditions (see Keane & Geer (2014); Geer & Keane (2014) for more detailed discussion of these results, as well experimental set up). These studies show that transitions are not more important than holds for fingerspelling perception, but rather the opposite. The temporal analysis discussed in this chapter was critical to both the formulation of the experiment, and construction of stimuli. This is just one extension of this timing work that would not be possible without understanding the basic temporal properties of fingerspelling. Finally, the timing analysis described here is a critical first step in the analysis of pinky extension coarticulation that will be explored in chapter 4.

The findings from the motion capture data are important. They show that motion capture technologies can be used to investigate (at least) temporal properties of fingerspelling. There are numerous benefits that motion capture technologies have: 1. In general they have higher temporal resolution than video data (the PhaseSpace system used here has a sample rate of 480 Hz compared to 60 Hz for the video data). This is important because it will help us tease apart differences in extremely short holds, which will allow us to further understand the large number of very short holds we found in the video data. 2. Automatic methods of hold identification will allow for much more data to be analyzed (because the time and monetary bottleneck of human annotation will be relieved). 3. Because motion capture data includes the position of markers on the hand, handshape variation like pinky extension (which will be discussed in detail in chapter 4) can be measured much more precisely.

Although fingerspelling is a distinct part of the ASL lexicon (see Brentari & Padden (2001); Keane *et al.* (2012b)), these findings have a few implications for understanding the temporal properties of the rest of ASL, and signed languages in general. As noted above, fingerspelling is distinct from other parts of ASL, because it uses only handshape (and for a very small number of FS-letters orientation and movement) for contrast. The rest of the ASL lexicon uses not only handshape, but also movement, location, orientation, and non-manuals to drive lexical contrasts. Because the other parameters all involve joints that are more proximal, and thus drive the movement of larger articulators (e.g. the elbow moves both the hand and the forearm, the shoulder moves the hand, forearm, and upper arm), segments that contrast across these parameters will likely be slower than segments in fingerspelling which contrast over basically just the joints of the hand. For this reason, the timing properties cannot be straightforwardly ported to lexical ASL signing, however some of the findings for fingerspelling could hold for signing more broadly. For example, we expect the signer variation found in fingerspelling will be present in comparable amounts for lexical signing. There is some work on prosodic patterns found in ASL and other sign languages, and the positional differences found in fingerspelling are similar to those found in signing: the last sign of an utterance is generally longer than utterance medial signs (Liddell (1978); Wilbur (1999), among many others). It is possible that the pattern found in fingerspelling with respect to holds and transitions in different positions of the word is similar for lexical signs: in fingerspelling, word-medial holds are all generally the same duration, with possibly only slightly shorter holds in later positions in words; however, the transitions show a significant reduction in duration in later positions in the word. In other words, as the word goes on, signers generally speed up the overall rate of fingerspelling by shortening the transitions but not the holds. This pattern should be tested at the utterance level for ASL and other sign languages: compare the durations of the lexical portions of the signs (since some signs involve movement, just using holds would not suffice for this definition) to the transitions between these signs. Finally, the methods used to determine fingerspelling location are the beginnings of methods to determine and distinguish the location of signs within the signing space. Of course there are more than two locations, but Hidden Markov Models with more than two states can be implemented and used to detect this distinct locations that the hand is in during lexical signing.

Chapter 4

Pinky extension coarticulation in ASL fingerspelling

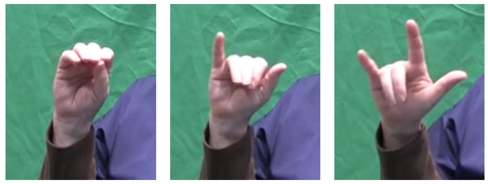
Coarticulation has been studied broadly for spoken language for a number of reasons, including the following: 1. As a phenomenon it is interesting in itself. One example of this is that understanding coarticulation can help in the automatic recognition of naturalistic speech (Deng & Sun (1994); Richardson *et al.* (2000); Livescu & Glass (2001) among others). 2. It can be used to test theories of phonological specification, as well as theories of the phonetics-phonology interface.

Coarticulation has seen a small amount of research in signed languages already (see section 2.1 for a more detailed review). For fingerspelling specifically, Jerde *et al.* (2003) found that there is both assimilatory as well as dissimilatory coarticulation for various parts of the hand. Hoopes (1998) notes the existence of pinky extension coarticulation in fingerspelling as well as signing, although he is interested in pinky extension as a sociolinguistic marker and sets aside the coarticulatory examples in his work. This chapter will look at pinky extension coarticulation in detail, using the same corpus of fingerspelling that was used in chapter 3. This data, and the analysis of it, will then be used to argue for the articulatory model of handshape that was proposed in chapter 2. The articulatory model of handshape makes clear predictions: the nonselected (nonactive) fingers will be the ones to undergo coarticulation (i.e. these will be the target of coarticulatory pressures), and these will be conditioned by the configurations of surrounding selected (active) fingers (i.e. these will be the sources of coarticulatory pressures). These will ultimately be supported through an analysis of pinky extension coarticulation in ASL fingerspelling.

4.1 Case studies

Three case studies have been conducted using visual estimation of extension to examine how the articulator positions change over time, and how well that aligns with any periods of stability. For each word below, the overall extension of every finger was estimated frame by frame for the entire period of time that the signer was fingerspelling the word. An extension value of zero was defined as when the finger was fully flexed; that is when all three of the joints of the finger (the metacarpophalangeal (MPC), proximal interphalangeal (PIP), and distal interphalangeal (DIP) joints) were completely flexed. An extension value of one was defined as when the finger was fully extended; that is when all three of the joints of the finger were extended completely. The thumb's measurement of extension is lateral across the palm (this is also described as radial-ulnar abduction), with zero being on the side of the hand, negative when the thumb is crossing over the palm, and positive when it is extended away from the thumb. Although these measurements of extension are coarser than other phonetic transcription systems (i.e. that of Johnson & Liddell (2011b); Liddell & Johnson (2011a,b)), they are sufficient for the purposes of these case studies.

Figures 4.1 and 4.2 show the extension of each finger over time for one signer, and one example of the word O-I-L. For each frame and each finger, a visual approximation of extension was made. Towards the bottom (a value of zero) is the most flexed that particular finger can be, and towards the top (a value of one) is the most extended. Lines are given for the observed values (thick, black) and the expected values (thin, red). Additionally gray boxes extend over periods of hand configuration stability, labeled with the associated FS-letter. For each period of handshape stability, the extension values for the selected fingers of a given FS-letter are overlaid (in darker, red boxes) as deviations form the dotted line at zero. This visualization is meant to function in a way similar to the gestural scores used by Browman & Goldstein (1986, 1992) among others (as shown in figure 2.9). The expected values line is generated by using the extension values of both the selected and nonselected fingers from the phonological specification of a canonical version of the handshape for a given FS-letter, with spline interpolation between apogees.



(a) -O- (b) -I- (c) -L-

Figure 4.1: Still images at apogees for O-I-L

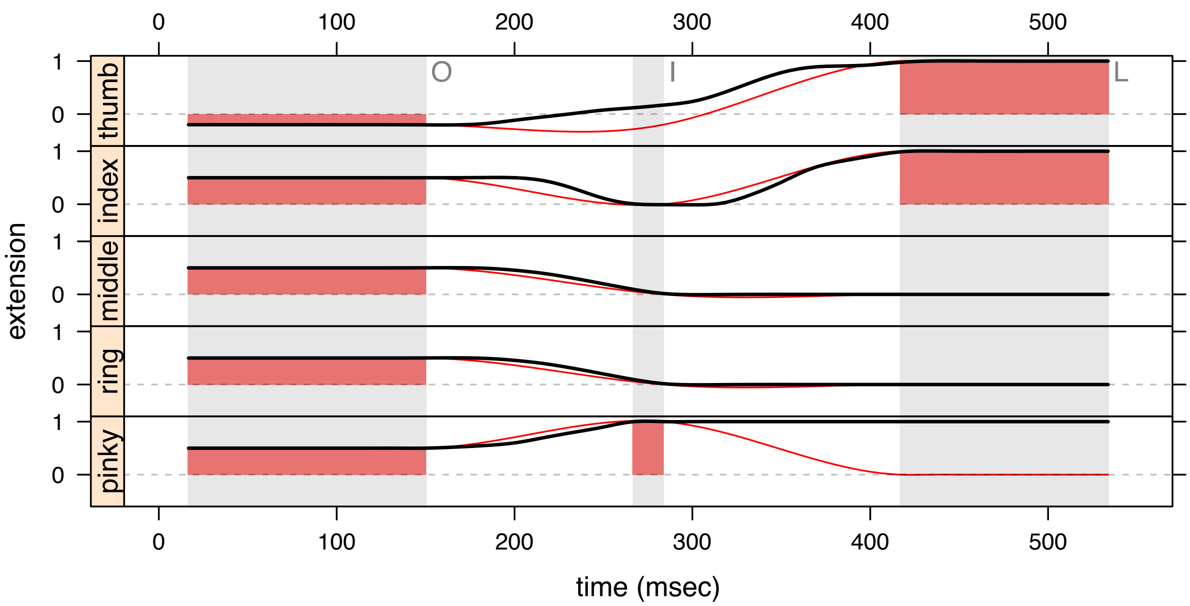
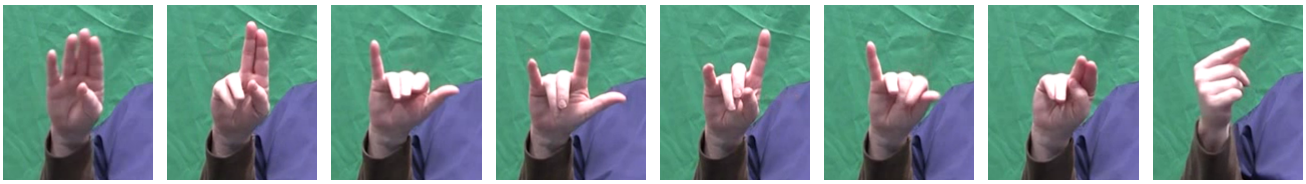


Figure 4.2: **Articulator trajectories for O-I-L** Gray boxes represent periods of hand configuration stability, thick, black lines represent observed extension (visually estimated), and the thin, red lines represent articulator trajectories if each apogee's hand configuration were canonical, with smooth transitions.

Starting with the first apogee, -O-, the observed and expected extension values match. For this FS-letter, all of the fingers are selected, for the fingers, the joints are phonologically specified so that they should have about 0.5 extension, and for the thumb there should be a little bit less than zero extension, because it is crossing over the palm. Moving on to the second apogee, the -I-, only the pinky finger is selected, which should be fully extended (ext = 1). All of the other fingers are nonselected, and should be fully flexed (ext = 0). For this apogee the observed extension for the fingers aligns with the phonological specification, the thumb, however, deviates slightly, which makes it more extended than expected. This deviation makes the thumb more like the configuration for the FS-letter that follows it: -L-. Finally, for the last apogee, the -L-, only the index finger and the thumb are selected, both being fully extended. The rest of the fingers are nonselected, and should be completely flexed. The thumb, as well as the index, middle, and ring fingers match the expected extension values. The pinky, however, stands out: although it should be flexed, it is almost completely extended. The pinky has the same extension as the apogee before it (the -I-), an example of the coarticulation that will be discussed in further detail in section 4.2. In this word, the only two deviations from expected values of extension occur with digits that are nonselected and should be flexed, but are realized as more extended, being more like the configurations of surrounding apogees (the following -L- in the case of the -I- and the preceding -I- in the case of -L-) .



(a) -B- (b) -U- (c) -I- (d) -L- (e) -D- (f) -I- (g) -N- (h) -G-

Figure 4.3: Still images at apogees for B-U-I-L-D-I-N-G

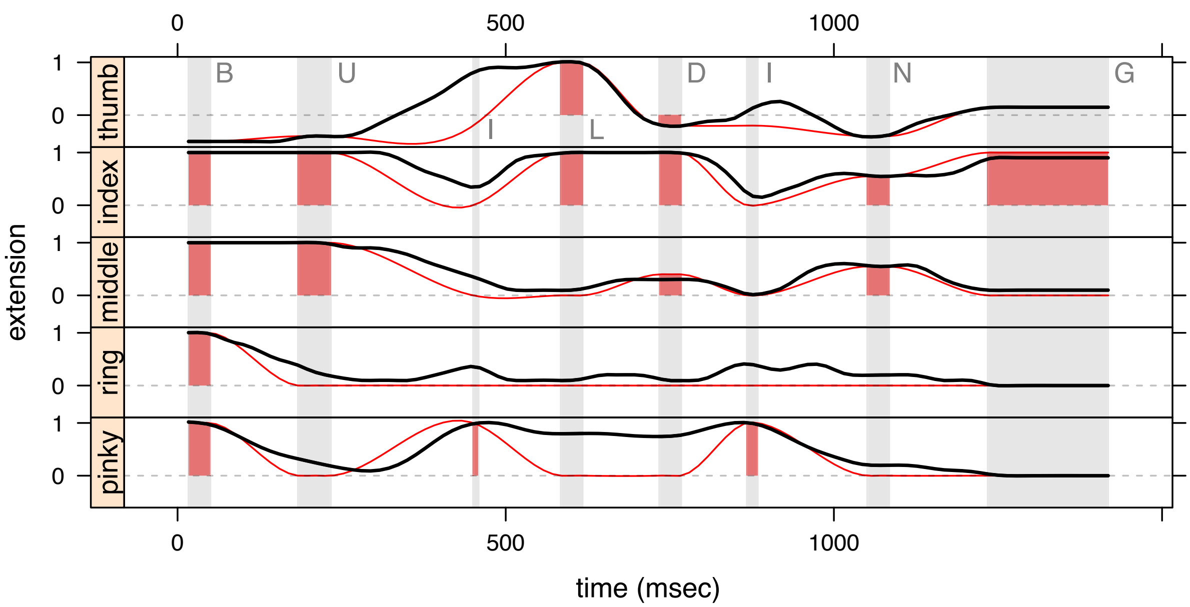
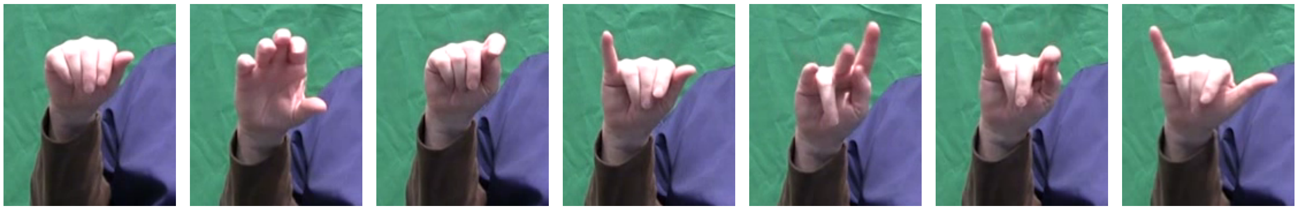


Figure 4.4: **Articulator trajectories for B-U-I-L-D-I-N-G** Similar to figure 4.2.

Figures 4.3 and 4.4 show the extension over time for the word B-U-I-L-D-I-N-G. The first apogee, -B- shows no deviation from the expected extension. The next apogee, -U-, shows no deviation for the thumb or the index or middle finger (the latter two, are selected), however the ring and pinky fingers, which are nonselected, are a little bit more extended than expected. The next apogee, the first -I-, shows a lot of deviation from expected extension values. The only digit that matches the expected extension value is the pinky, which is also the only selected finger. The ring, middle, and index fingers all are slightly more extended than expected, and the thumb is completely extended, matching the configuration of the following apogee. For the -L- apogee, the thumb and index finger are selected, and both match their expected extension values. The middle and the ring finger are slightly more extended than expected, and finally the pinky is nearly fully extended, which matches the -I- before it. In the next apogee, the -D-, the thumb as well as the index and ring finger are selected[1]; and they all match the expected extension values. The ring and pinky fingers are nonselected; the ring finger matches the expected extension, however the pinky is much more extended than expected. Across the last two apogees the pinky is more extended than expected given the phonological specification for each handshape, however there is a handshape with an extended pinky on either side of these two (both -I-s), which is conditioning coarticulation of pinky extension. Moving on to the second -I- apogee, the pinky is selected, and matches the expected extension value. The other digits approximate their expected values, with the exception of the thumb and ring finger. Following that, the -N- apogee, has the index and middle fingers selected, both of those, along with the other digits match the expected values. There are only slight deviations of the ring and pinky fingers, both of which are not selected. Finally, the last apogee, -G-, has the index finger selected, which matches the expected extension value. Additionally, all of the other digits similarly match their expected exten-

1. What fingers are selected for the FS-letter -D- is not actually a settled matter. In some models the thumb as well as the middle, ring, and pinky fingers are selected, the index finger is either nonselected and extended, or secondary-selected. However, Keane *et al.* (2012a) have shown shown that -D- is frequently realized as what is referred to as baby-D-, that is with the pinky and middle fingers completely flexed, the middle finger and the thumb forming a loop, and the index finger fully extended. The apogee here, shows this pattern with flexion in the ring finger, although the pinky is extended because of coarticulation from -I- apogees around it. With that configuration the middle finger and thumb would be selected, and the index finger secondary-selected, while the ring and pinky fingers are nonselected.

sion values. This case study shows again, that there is quite a bit of extension variation for fingers that are nonselected; especially on the pinky finger when it has apogees with pinky extension on either side. In contrast, the selected fingers of a given apogee always match the expected extension.



(a) -A- (b) -C- (c) -T- (d) -I- (e) -V- (f) -IT- (g) -Y-

Figure 4.5: Still images at apogees for A-C-T-I-V-I-T-Y

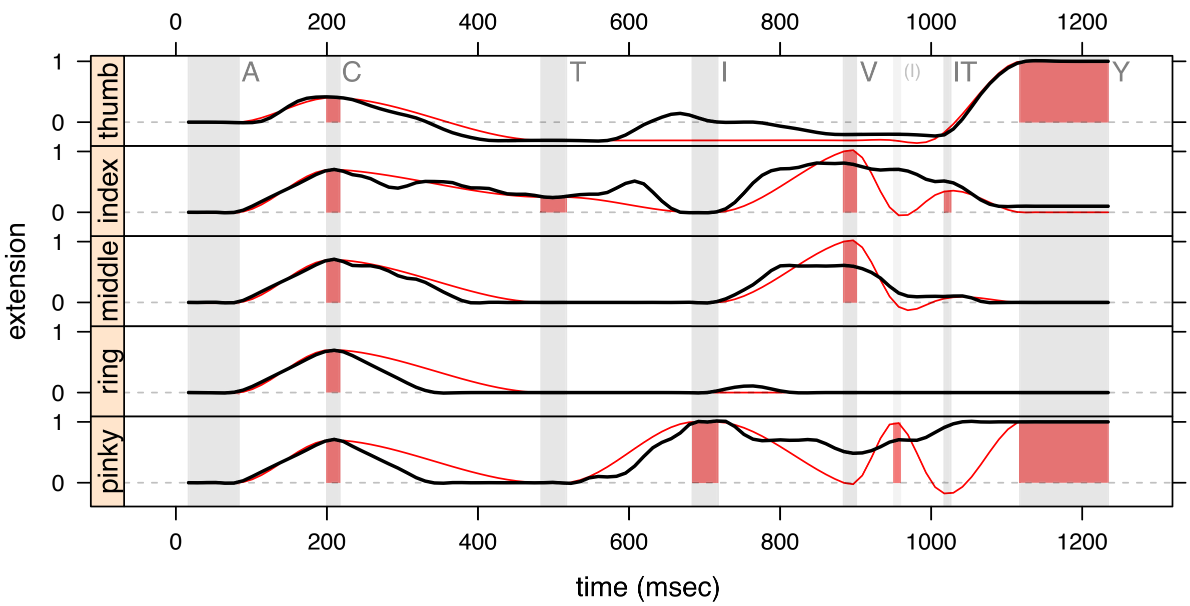


Figure 4.6: **Articulator trajectories for A-C-T-I-V-I-T-Y** Similar to figure 4.2, the only exception is that the light gray associated with the second -I-, is placed halfway between the -V- and -T- and -I- apogees in order to show the trajectories expected for canonical realization.

Moving on to a more complicated example, A-C-T-I-V-I-T-Y in figures 4.5 and 4.6, continue to show the relationship between selected and nonselected fingers. The first observed extension matches the expected extension for the first five apogees (-A-, -C-, -T-, and -I-) for both the selected and nonselected fingers. After that, however, there is quite a bit of deviation: the next apogee, -V-, has unexpected pinky extension, as well as some articulatory undershoot for the two selected fingers (the index and the middle finger). After that the next period of stability is actually two apogees (-I- and -T-) fused together to form -I- and -T-. The selected fingers for these two FS-letters do not clash: for the -I- the only selected finger is the pinky, whereas for the -T- only the index finger is selected. The two sets of selected fingers are separate, and thus do not conflict. The observed extension for the index and pinky fingers reach the extension targets for -I- and -T- at the same time, and thus the two apogees occupy the same period of time. In figure 4.6, a period of stability has been inserted halfway between the -V- and -I- and -T- to show what the articulators are expected to do if the fusion did not occur. The last apogee, -Y- matches the expected configuration. This case study shows two things: First, during the period of time between the two -I- apogees (including the fused -I- and -T- apogee), the pinky does not ever completely flex, but rather stays at least partially extended as a result of coarticulation, and it is not selected except in any of the intervening apogees. Second, in some extraordinary cases, apogees that do not have conflicting selected fingers can be fused temporally, where the articulators reach their phonologically specified targets at the same time.

Although rare, the apogee fusion seen here is not a solitary example. There are also examples of -TI-, -NI-, and -OI-; the last one is even documented as one strategy that is used in rapid lexicalization (Brentari, 1998). Two out of three of these share the property that the selected fingers of the two FS-letters are distinct, and thus there is no conflict. The -O- and -I-, however seems to present a problem because a canonical -O- should have all fingers selected. There is some work (Keane *et al.*, 2012a) that shows that there are instances of -O- where the ulnar digits (typically the pinky and ring fingers) are completely flexed rather than having the same configuration as the radial digits (typically the index and the middle fingers). This happens in approximately 25% of -O-s in this corpus. The analysis of these variants are that these handshapes have different selected fingers than the canonical forms, that is, only the index and middle fingers are selected, while the pinky and ring fingers are nonselected. Additionally the one example of -O- and -I- in our corpus shows increased flexion of the ring finger, just like with the -D- in the B-U-I-L-D-I-N-G case study, suggesting that this case of -O- and -I- fusion might involve an -O- variant that does not have the pinky finger selected. More work, and more data, are needed to fully understand and model how these two different types of variation interact, work which, in part, is included in other chapters of this thesis.

With a model of handshape that treats handshape as a whole, these fusions would have to represent examples of new kinds of segments in the inventory of FS-letters. However, if finger selection is taken into account, these fused apogees can still be analyzed as two apogees, that just happen to occupy the same time. Why this fusion occurs is outside of the scope of this work, however many (e.g. Wilcox (1992); Emmorey & Petrich (2011)) have noted that fingerspelling often has a rhythmic pattern. We have observed what appear to be consistent rhythmic patterns of holds, although less so for transitions, in our corpus. We speculate that the fusion process might be a way to maintain the rhythm when two apogees are too close together, and do not have conflicting selected fingers so they become fused. It is also interesting that most examples of this fusion happen at the ends of fingerspelled words, where, as we discuss in chapter 3, the transitions are generally shorter. More data and analysis are required to understand this phenomenon fully.

All three of these case studies show evidence in support of the hypotheses (reprinted below) given in chapter 2. With respect to hypothesis 1, we see deviation from targets more with the nonselected fingers. Looking at hypothesis 2, we see that these deviations seem to be preserving or anticipating the configuration of selected fingers in the surrounding apogees. Finally, evidence for hypothesis 3 is that the deviations do not all match the exact extension of the conditioning segment, but are frequently somewhere in between full extension and full flexion.

1. The nonselected fingers are more frequently the targets of coarticulatory pressure (vs. selected fingers).

2. The selected fingers are the sources of coarticulatory pressure.

3. Finger configuration that is due to (phonetic) coarticulatory pressure will differ from configuration due to phonological specification.

The case studies above show support for the hypotheses that follow from the articulatory model of handshape. More robust quantification will allow us to confirm that these patterns hold for the fingerspelling system broadly. The next section will look at pinky extension coarticulation specifically. Analyzing a large corpus of fingerspelling shows that the trends seen in the case studies are representative of the coarticulation of pinky extension in fingerspelling generally.

4.2 A quantitative measure of pinky extension coarticulation

As suggested by the case studies, we have found that there is, indeed, coarticulation with respect to pinky extension. This coarticulation is conditioned by both preceding and following handshapes that include an extended pinky, although there is a clear distinction between handshapes where the pinky is extended and the other fingers are not (-I-, -J-, and -Y-) and those where the pinky is extended along with other fingers (-B-, -C-, and -F-). Additionally, handshapes where the pinky is selected and flexed (-A- and -S-) have less pinky extension coarticulation than handshapes where the pinky is not selected.

4.2.1 Methods

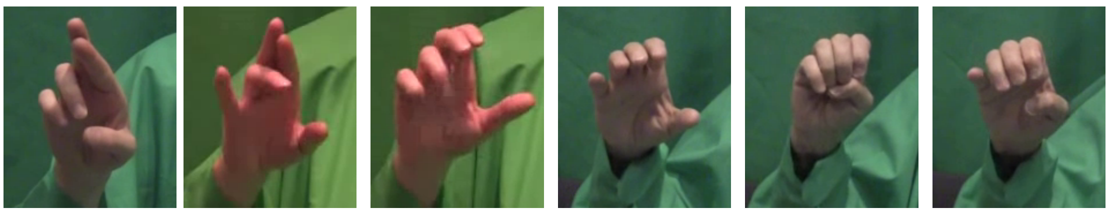
We used the same large corpus of fingerspelled words that was analyzed in chapter 3, although it required further annotation in order to test hypotheses about hand configuration.

Using the timing data annotated so far, we extracted still images of every apogee. The frame that was chosen in the case of single frame holds (or instantaneous apogees) was the one frame that was annotated as the apogee. The frame that was chosen in the case of multiframe holds was the frame that was exactly in the middle of the hold. Because of the criterion for timing annotation (that the hand configuration be stable), choosing the image from the beginning, middle, or end of a given hold should not matter (see section 3.2 for more detailed description). This image was associated with the corresponding apogee data in the database which not only allowed for exploratory data analysis, but was also the basis of our resulting hand configuration annotations: The still images were then used to annotate a number of different features of hand configuration. The major guiding principle in this feature annotation was to keep the task as simple and context free as possible. This has two major goals:

Simplicity — The first principle is simplicity. We wanted each annotation task to be as simple as possible. This allows the training to be simple, and the task to be incredibly quick. Rather than attempting to annotate features of hand configuration as a whole using recent annotation methods (Eccarius & Brentari, 2008; Liddell & Johnson, 2011b,a; Johnson & Liddell, 2011b), we use binary decision tasks that involving looking at an image of an apogee and deciding if some feature of the hand configuration is one of two values. This makes the actual annotation very, very quick, so a number of annotators can be used for every apogee, which allows us to check agreement, rate annotator accuracy, and even then possibly derive some amount of certainty or gradience about the particular phenomenon (although this gradience will not be explored or used in the current study). All individual annotations will be analyzed in the subsequent sections, although we will use models that allow for this type of repeated measurements of the same apogees, as well as modeling the variation associated with individual annotators.

We defined a pinky as extended if the tip of the pinky was above a plane perpendicular to the palmar plane, at the base of the pinky finger (the MCP joint) and the proximal interphalangeal joint (PIP) was more than half extended. Note that the canonical -E- shape would not have pinky extension (fig 4.7e), although some did exhibited coarticulation (fig 4.7f). A more nuanced definition might be needed for further work but this is sufficient to identify apogees where the pinky is not in a closed, flexed configuration. With this metric the handshapes for -B-, -F-, -I-, -J-, -Y-, and sometimes -C- would have extended pinkies, and the rest of the FS-letters would not. Figure 4.7c shows a

-C- without pinky extension, figure 4.7d shows one with pinky extension. Given this definition annotators were shown images of every apogee, and determined if the pinky was extended or not. Of course, as with all phonetic realizations, pinky extension is not actually binary. A variety of measures of the amount of extension (either for the finger overall, or individual joints) could be used, however these are all much more complicated to annotate than a simple binary decision, requiring much more annotator training and time per annotation.



(a) -R- [−ext] (b) -R- [+ext] (c) -C- [−ext] (d) -C- [+ext] (e) -E- [−ext] (f) -E- [+ext]

Figure 4.7: Apogees from (a) D-I-N-O-S-A-U-**R**, (b) C-H-**R**-I-S, (c) Z-A-**C**-K, (d) E-X-P-E-**C**-T-A-T-I-O-N, (e) **E**-V-E-R-G-L-A-D-E-S, and (f) Z-D-R-O-Q-I-**E**

Context-free — Every image was presented with as little context as possible to ensure that the annotations were as objective as possible. Annotators are likely to have a variety of biases about how canonical they expect or do not expect given hand configurations to be. In order to try and reduce the influence of annotator bias, no information was given about the apogee in the image as it was annotated. The FS-letter of the apogee was not included, nor was the word, or any features of the surrounding apogees. Although hand configurations (and orientations) that are near the handshape for a given FS-letter are easy to identify, and thus could still influence annotation decisions, hand configurations that are far from any canonical FS-letter handshape there will be little to distract the annotator from the task at hand (e.g. pinky extension annotation). Additionally even if the annotator knows the hypothesis to be tested (e.g. that certain handshapes in neighboring apogees condition coarticulation), their annotation cannot be biased because they have no way of knowing what the neighboring apogees are. One possible drawback to this method is that in the case of occlusions, it is sometimes impossible to determine some hand configuration features. It is possible that in some of these cases being able to play back the contextual video would provide enough information to

		expected	
		+pinky extension	−pinky extension
observed	+pinky extension	2286	1104
	−pinky extension	56	10077

Table 4.1: **Counts for expected and observed pinky extension** where the columns are handshapes with and without piny extension, and the rows are hand configurations with and without pinky extension. The shaded cells are those where the pinky extension in the hand configuration matches the handshape specification. Here we are using the familiar terminology observed and expected. We use the terms *observed* and *expected*, even though our hypothesis is that there is coarticulation. In other words, we are using these labels in the naive way that we do not expect any apogee that does not (phonologically) have pinky extension in its handshape, to have it (phonetically) in its hand configuration.

determine the appropriate annotation. Although this might be true for a small number of cases, the benefit of reducing annotator bias far outweighs the additional (possible) accuracy in this edge case.

4.2.2 *Results*

Looking at table 4.1 we see that the apogees of handshapes that have pinky extension (-B-, -F-, -I-, -J-, -Y-, and sometimes -C-) by and large have it in the hand configuration as well (2286 apogees, versus 56 apogees with no extension). Of the 56 in this set that do not have pinky extension the majority of them (38) are -C- which has a variant where only the index finger being in the expected extended configuration and the other digits are fully flexed (as if they were nonselected). This variant is also known as baby-C-. For the rest of the apogees (i.e. the handshapes that do not have pinky extension) we see a surprising 1115 apogees have pinky extension, which is almost 10% of all apogees in this set. One source of hand configuration variation is coarticulation. In order to test if the distribution of pinky extension observed is a result of coarticulation, contextual variables around each apogee (e.g. surrounding apogee handshapes, surrounding transition times) need to be investigated.

There are numerous factors that are known or suspected to condition phonetic variation like the variation we see with respect to pinky extension.[2] Two contextual factors are the handshape of the surrounding signs, or in this case apogees, as well as the transition times to and from the surrounding apogees. The hypothesis here is that surrounding apogees that have handshapes with pinky extension will increase the chance of an apogee's hand configuration exhibiting pinky extension even though its handshape does not specify pinky extension. Additionally we hypothesize that if the transition between a conditioning apogee and the apogee we are interested in is faster, this will also increase the chance of pinky extension. In addition to these contextual factors there are other noncontextual factors that might effect rates of pinky extension: the category of the word being fingerspelled (name, noun, non-English) as well as which signer is fingerspelling the word.

For a first look at the effect of the handshape of surrounding apogees we will check the two possible groups that could condition pinky extension in the hand configuration of apogees that do not have pinky extension in their handshape. The two groups of FS-letters that have pinky extension in their handshapes are -I-, -J-, and -Y- as well as -B-, -C-, and -F-. For apogees with handshapes that do not have pinky extension (all FS-letters but -B-, -C-, -F-, -I-, -J-, and -Y-) we see that apogees that have an -I-, -J-, or -Y- on either side of them have more instances with pinky extension than those that have any other letter on either side, including -B-, -C-, and -F- (see figure 4.8).

Additionally, we predict that faster fingerspelling will result in more coarticulation when a conditioning handshape is present. The previous and following transitions can be used as a predictors as well as the previous and following transitions in combination with the previous and following handshapes respectively. This, however, restricts the analysis to apogees that had both a previous and following apogee, that is word-internal apogees, which as we saw in the timing analysis above, behave as a class with respect to their hold durations (in other words: the first and last apogees were held for significantly and considerably longer, but the word-medial apogees all had by and large the same hold durations). Additionally, we can include the hold durations as predictors from the previ-

2. (Cheek, 2001) for environment; (Mauk, 2003) for speed and environment; (Lucas *et al.*, 2002) for grammatical category, among many others. See chapter 2 for a more detailed discussion.

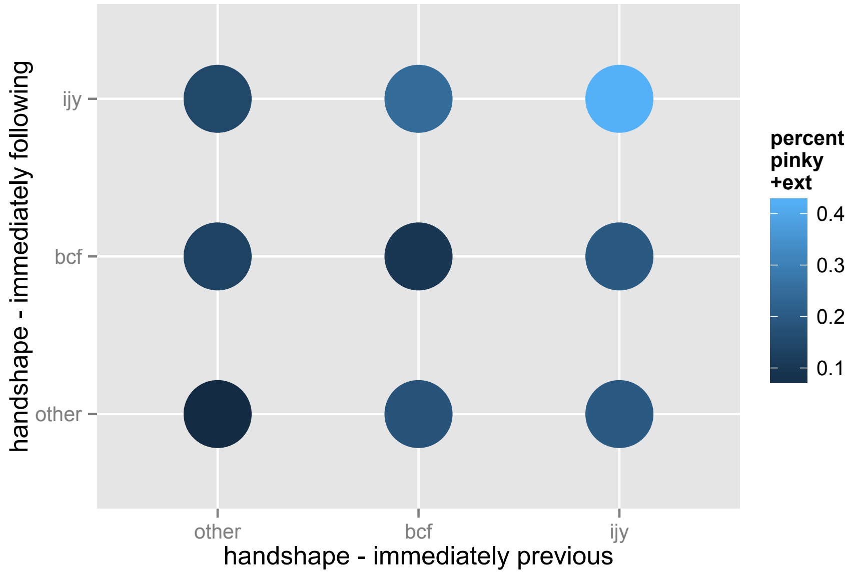


Figure 4.8: A plot showing the percent of apogees with hand configurations that have pinky extension, despite their handshapes not specifying pinky extension, based on surrounding handshapes. Darker colors represent a higher percentage of pinky extension.

ous timing analysis. As we saw from the timing analysis previously, there is a relationship between overall word rate, and hold duration (although this does not hold for transition durations). This means that it is possible that there is some correlation between the time based predictors: when the previous transition time is long, it is more likely that the following transition time is long (or the hold duration is short). For this data, there is a moderate amount of correlation between the previous and following transition times (Kendall's $\tau = 0.55$), and a smaller amount of correlation between previous or following transitions times and hold durations (Kendall's $\tau = -0.35$ and $\tau = -0.32$ respectively). This correlation can impact regressions, although this will in general lead to inflated standard errors, which will lead to falsely rejecting the influence of the predictor (an inflation of type II error). As a check against this models were fit with other methods, and the results are generally the same (see appendix E for comparisons of the different models).

A hierarchical logistic regression model (as a reminder, these are also known as logistic mixed effects regressions) was fit with pinky extension as the outcome variable. This model is similar to the models that were used in chapter 3, although instead of using a linear relationship between the predictors and the outcome, a logistic relationship is used. This relationship is necessary for binary outcomes: because the values of the predictors are only one of two options (here, extended or flexed) what is predicted is not simply the outcome, but rather the log-odds of one of the two categories being true. (Gelman & Hill, 2007; Jaeger, 2008; Baayen *et al.*, 2008)

We used all of the individual annotations for all of the apogees from the same corpus that was analyzed in chapter 3, with a few exclusions. Each apogee was annotated by at least two different annotators, and some were annotated by more than that. Words that contained self-reported errors were removed form the analysis. Additionally any apogee that was annotated as one of the fused apogees discussed in section 4.1 was removed from the analysis. All apogees in the first and last position of words were also removed. This leaves us with 29,499 annotations for 13,523 apogees.

The predictors in the model were as follows:

- apogee of interest handshape group: -B-, -C-, -F-, -I-, -J-, or -Y-; -A- or -S-; -E- or -O-[3]; other — abbreviation: presGroup
- hold duration (zscore of the box-cox power transform[4] of the duration) — abbreviation: holdDur
- previous apogee handshape group -B-, -C-, or -F-; -I-, -J-, or -Y-; other — abbreviation: prevGroup
- previous transition (zscore of the box-cox power transform of the transition time) — abbreviation: prevTrans

3. The FS-letters -A- and -S- were separated from -E- and -O- because they differ in the level of extension expected in the canonical forms. -E- and -O- both have pinkies that are about half extended in the canonical form. For this reason, and as will be discussed later, we expect that -E- and -O- will pattern separately from -A- and -S- as well as the other FS-letters.

4. The box-cox power transformation is used to transform a skewed, non-normal continuous predictor into a non-skewed, normal predictor. The power of the transformation is determined by the skewness and non-normality of the underlying predictor. The predictor is then raised to that power. This transformation (in addition to scaling and centering) makes the predictors more interpretable in scale with other predictors. Models with untransformed and unscaled predictors were also fit, and resulted in the same predictors being significant.

- following apogee handshape group -B-, -C-, or -F-; -I-, -J-, or -Y-; word boundary; other — abbreviation: follGroup

- following transition (zscore of the box-cox power transform of the transition time) — abbreviation: follTrans

- word type: noun; foreign; name — abbreviation: wordtype

- *interaction* apogee of interest × hold duration

- *interaction* previous handshape × previous transition time × hold duration

- *interaction* following handshape × following transition time × hold duration

In this model there were also grouping variables that allowed intercepts and slopes to vary as follows: intercepts were allowed to vary for signer, word, annotator, and FS-letter of the apogee of interest (also called current apogee), as well as for each apogee in the dataset. Additionally the slope of the previous transition time, following transition time, and hold duration effects were allowed to vary based on the signer and based on the FS-letter of the apogee (see table 4.2 for full model details and coefficient plot in figure 4.9).

The signer and FS-letter of the apogee of interest intercept and slope adjustments allows us to see individual (signer) variation within fingerspelling, as well as the inherent variation within each of the phonological specifications for FS-letters.

Intercept adjustments for each apogee in the dataset are necessary because each apogee is annotated by at least 2 different annotators. Because the annotators see the same image when they annotated the same apogee, this is a form of repeated measurements of the same apogee. These repeated measurements allows for a more accurate picture because if an individual annotator makes a mistake, it is unlikely that other annotators will make the same mistake as well. Although these mistakes will produce noise, the number of annotations total (and with increased annotations per individual apogee) will outweigh this noise.

Intercept adjustments for annotator allow us to see the variation among the annotators, and their overall propensity for annotating pinkies extended or flexed. Intercept variations for words allows for the variation because of the specific word that contributes to the likelihood of pinky extension. This, combined with the repeated measurements of individual apogees allows us to model (and remove in our ultimate coefficient estimates) individual annotator trends.

Grouping variable adjustments were as follows:

- Intercept adjustments for signer, as well as slope adjustments for:
  - previous transition time
  - following transition time
  - hold duration
- FS-letter of the apogee of interest (also called current apogee), as well as slope adjustments for:
  - previous transition time
  - following transition time
  - hold duration
- annotator
- word
- each apogee (or item) in the dataset

Using this model, we determined that the following have a significant effect on pinky extension: handshape of the apogee (of interest), handshape of the previous apogee, handshape of the following apogee, and the interaction of following handshape and local transition time. Specifically, the following were correlated with an increased probability of pinky extension in the hand configuration:

- if the apogee of interest was a -B-, -C-, -F-, -I-, -J-, or -Y- (and thus the handshape had pinky extension),
- if the previous or following apogee was an -I-, -J-, or -Y-,
- to a lesser extent (than the following effect immediately above) if the previous or following apogee was a -B-, -C-, or -F-,
- if the wordtype is a name[5],
- if both the following apogee's handshape was -I-, -J-, -Y-, -B-, -C-, or -F- and the following transition time was shorter,
- if both the previous apogee's handshape was -B-, -C-, or -F- and the previous transition time was shorter,
- if the following apogee's handshape was -I-, -J-, -Y-, -B-, -C-, or -F- and the hold duration was longer,
- to a lesser extent (than the following effect immediately above) if the previous apogee's handshape was -I-, -J-, -Y-, -B-, -C-, or -F- and the hold duration was longer,
- finally, the three-way interaction of hold duration, following transitions time, and following group magnifies the two two-way effects of hold duration and following group as well as following transition time and following group.

The following were correlated with a decreased probability of pinky extension:

- if the apogee of interest was an -A- or -S- (but not -E- or -O-) (in other words, those FS-letters that have a selected, and fully flexed pinky finger), which is expected given the articulatory model of handshape,

5. This effect is a little surprising, although it is incredibly small (0.47), and only just outside of the 99% confidence interval. Additionally, in other models this effect is not robust, so should be taken with a grain of salt.

- if the apogee of interest was an -E- or -O- and the hold duration is longer.

All other effects are not significant (again, see the coefficient plot in figure 4.9 and table 4.2 for full model details).

	coeficient (standard error)
(Intercept)	$-4.61(0.67)^{***}$
presGroupas	$-4.40(1.24)^{***}$
presGroupbcfijy	$11.46(0.80)^{***}$
presGroupeo	$2.77(1.21)^{*}$
holdDur	$0.18(0.12)$
prevGroupbcf	$1.47(0.18)^{***}$
prevGroupijy	$3.06(0.17)^{***}$
prevTrans	$-0.26(0.13)^{*}$
follGroupbcf	$0.98(0.22)^{***}$
follGroupijy	$1.60(0.16)^{***}$
follTrans	$0.02(0.17)$
wordtypename	$0.49(0.18)^{**}$
wordtypenonEnglish	$0.12(0.19)$
presGroupas:holdDur	$-0.58(0.35)$
presGroupbcfijy:holdDur	$-0.49(0.24)^{*}$
presGroupeo:holdDur	$-0.86(0.22)^{***}$
prevGroupbcf:prevTrans	$-0.67(0.16)^{***}$
prevGroupijy:prevTrans	$-0.32(0.15)^{*}$
holdDur:prevGroupbcf	$0.67(0.17)^{***}$
holdDur:prevGroupijy	$0.51(0.15)^{***}$
holdDur:prevTrans	$-0.03(0.08)$
follGroupbcf:follTrans	$-0.95(0.20)^{***}$
follGroupijy:follTrans	$-1.98(0.16)^{***}$
holdDur:follGroupbcf	$1.02(0.21)^{***}$
holdDur:follGroupijy	$1.36(0.15)^{***}$
holdDur:follTrans	$0.10(0.07)$
holdDur:prevGroupbcf:prevTrans	$-0.06(0.16)$
holdDur:prevGroupijy:prevTrans	$0.07(0.14)$
holdDur:follGroupbcf:follTrans	$-0.58(0.19)^{**}$
holdDur:follGroupijy:follTrans	$-0.62(0.14)^{***}$
AIC	11249.13
BIC	11688.62
Log Likelihood	-5571.57
Deviance	11143.13
Num. obs.	29499
Num. groups: apogeeId	13523
Num. groups: wordList:word	599
Num. groups: apogeeLetter	26
Num. groups: annotator	19
Num. groups: signer	4
Variance: apogeeId.(Intercept)	3.46
Variance: wordList:word.(Intercept)	1.24
Variance: apogeeLetter.(Intercept)	2.15
Variance: apogeeLetter.follTrans	0.19
Variance: apogeeLetter.prevTrans	0.12
Variance: apogeeLetter.holdDur	0.04
Variance: annotator.(Intercept)	1.28
Variance: signer.(Intercept)	0.69
Variance: signer.follTrans	0.05
Variance: signer.prevTrans	0.01
Variance: signer.holdDur	0.01
Variance: Residual	1.00

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 4.2: Coefficient estimates and standard errors for the full hierarchical logistic model including all predictors for pinky extension

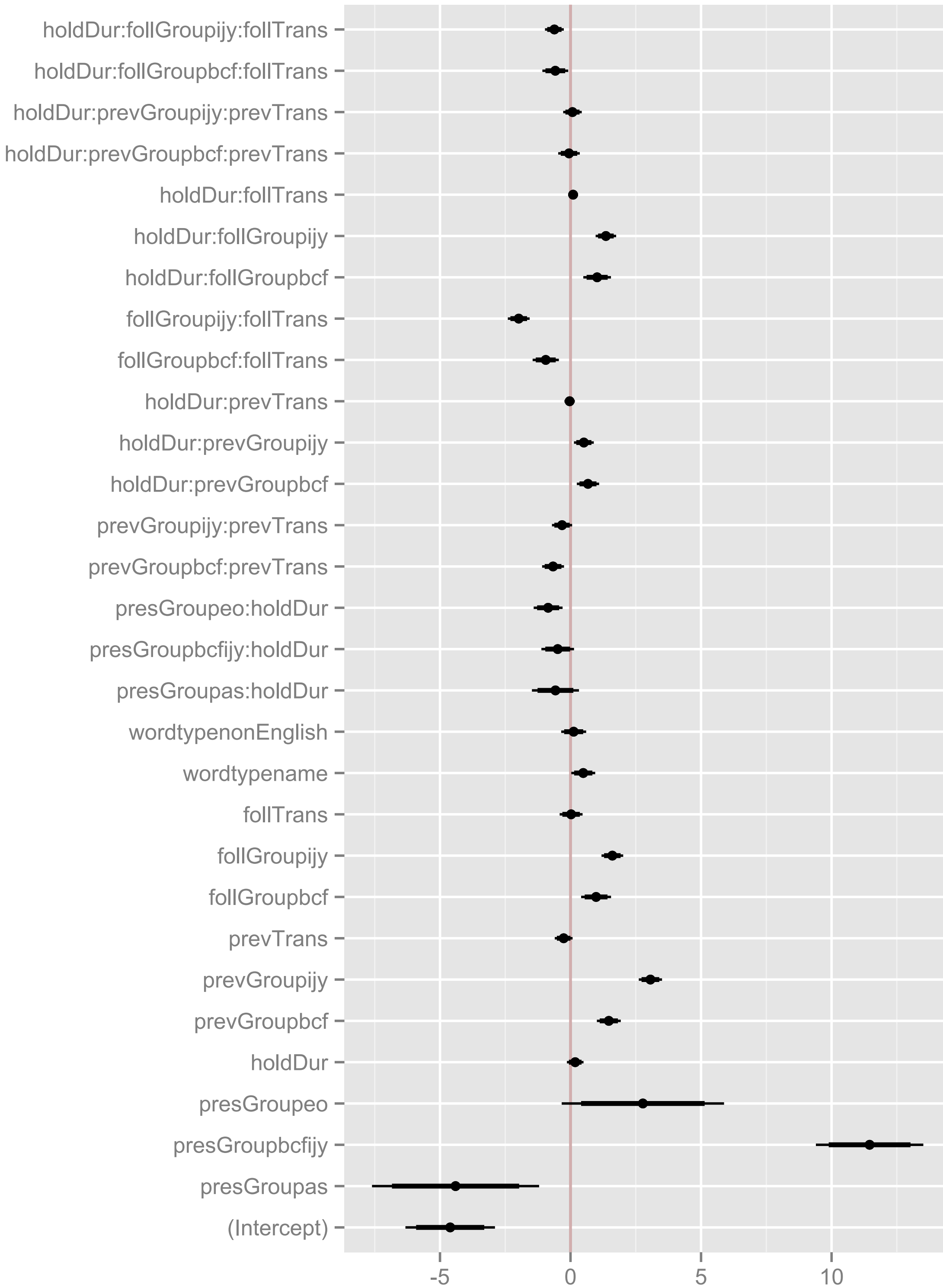


Figure 4.9: **Coefficient plot for the predictors of the hierarchical logistic regression model for pinky extension** Thick lines represent 95% confidence, thin lines 99% confidence, and dots are the estimates of the coefficients (or intercept).

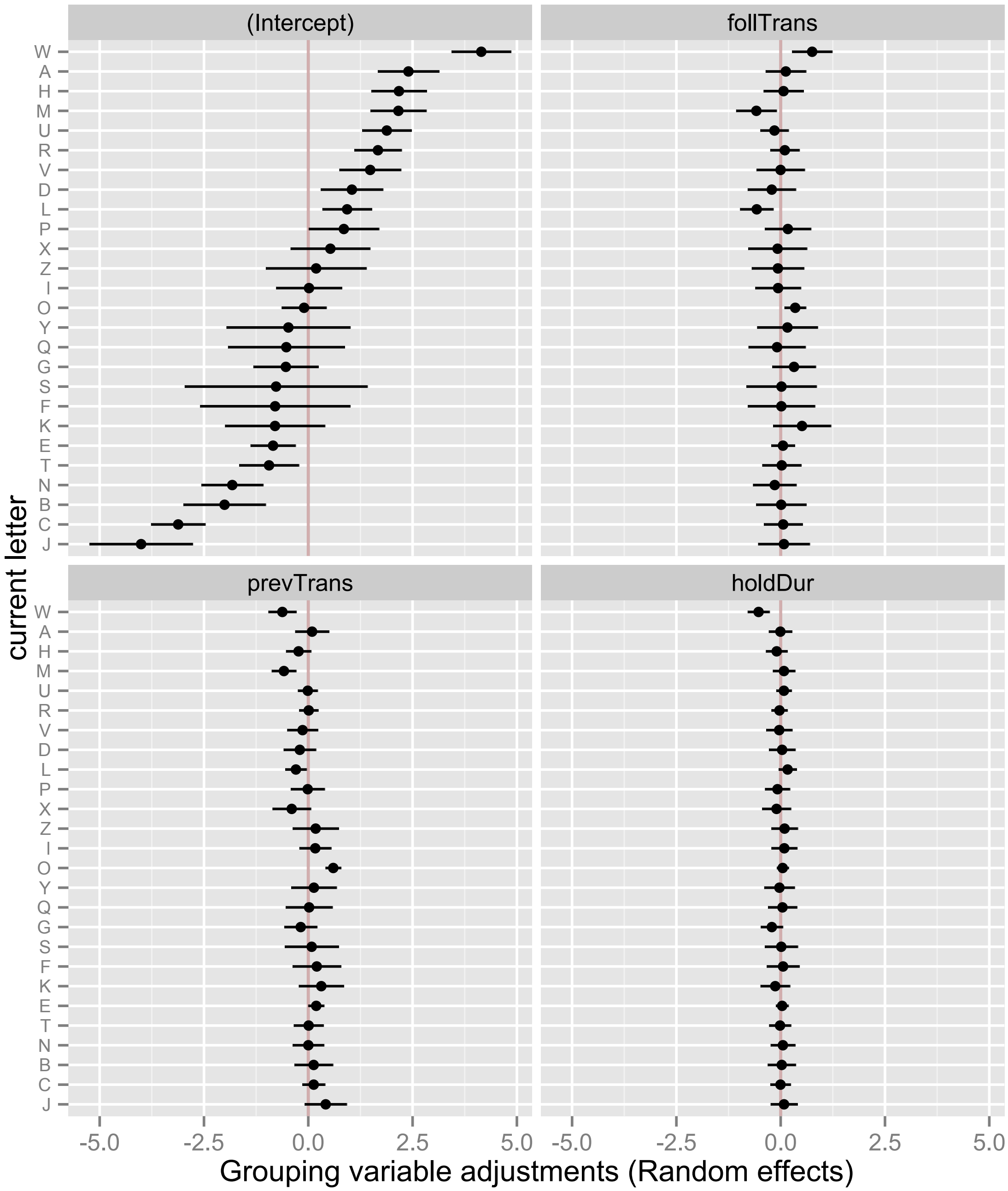


Figure 4.10: **Plot of intercept adjustments (random intercepts) for current letter as well as slope adjustments (random slopes) for following transition time, previous transition time, and hold durations of the hierarchical logistic regression model for pinky extension** As discussed in detail below, of the FS-letters with the pinky nonselected and flexed, some are more likely to have pinky extension (-W-, -H-, -M-, -U-, -R-, and -V-) and some are less likely to have pinky extension (-N- and -T-) than most other letters The levels on the y-axis are current letters, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

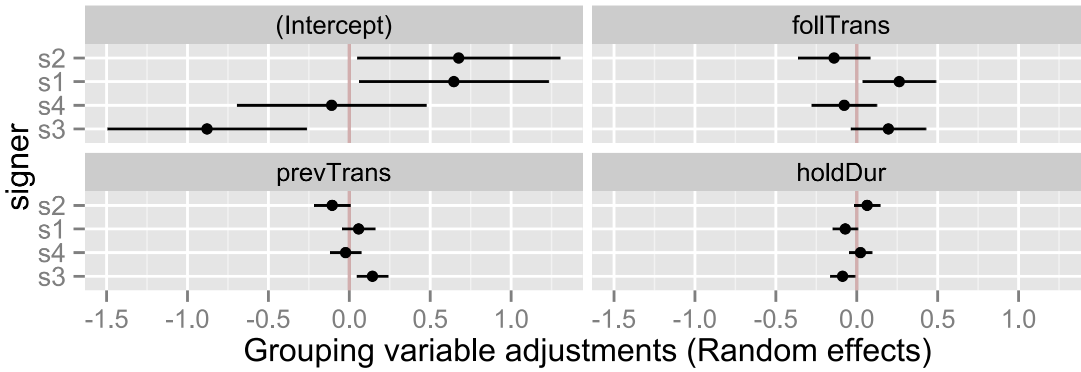


Figure 4.11: **Plot of intercept adjustments (random intercepts) for signer, as well as slope adjustments (random slopes) for following transition time, previous transition time, and hold durations of the hierarchical logistic regression model for pinky extension** As discussed in detail below, there is some intersigner variation (seen in the intercept facet), additionally, there is little variation among signers with respect to the effects of following transition time, previous transition time, and hold duration. The levels on the y-axis are signers, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

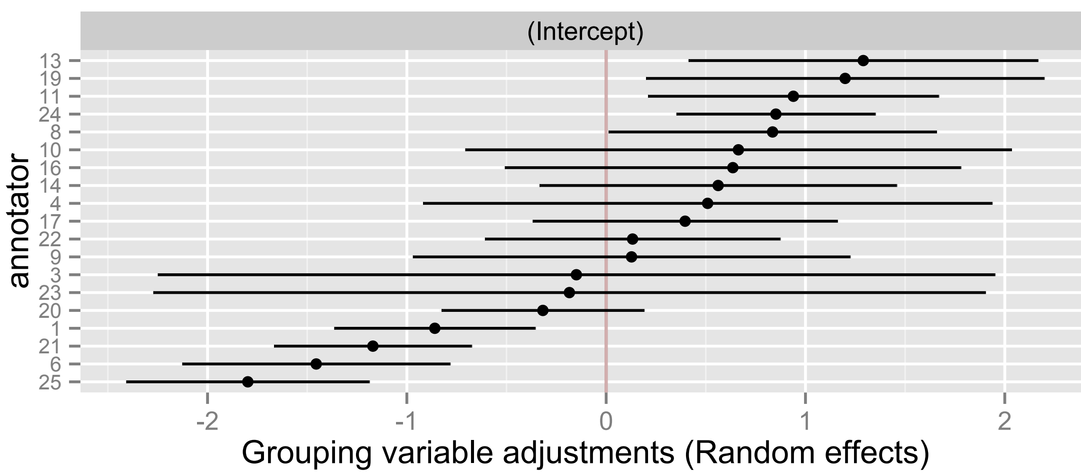


Figure 4.12: **Plot of intercept adjustments (random intercepts) for annotator of the hierarchical logistic regression model for pinky extension** As discussed in detail below, there is some systematic variation of annotators: annotators 25, 6, 21, and 1 are less likely to annotate an apogee as extended, and annotators 13, 19, 11, and 24 are more likely to annotate an apogee as extended. The levels on the y-axis are annotators, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

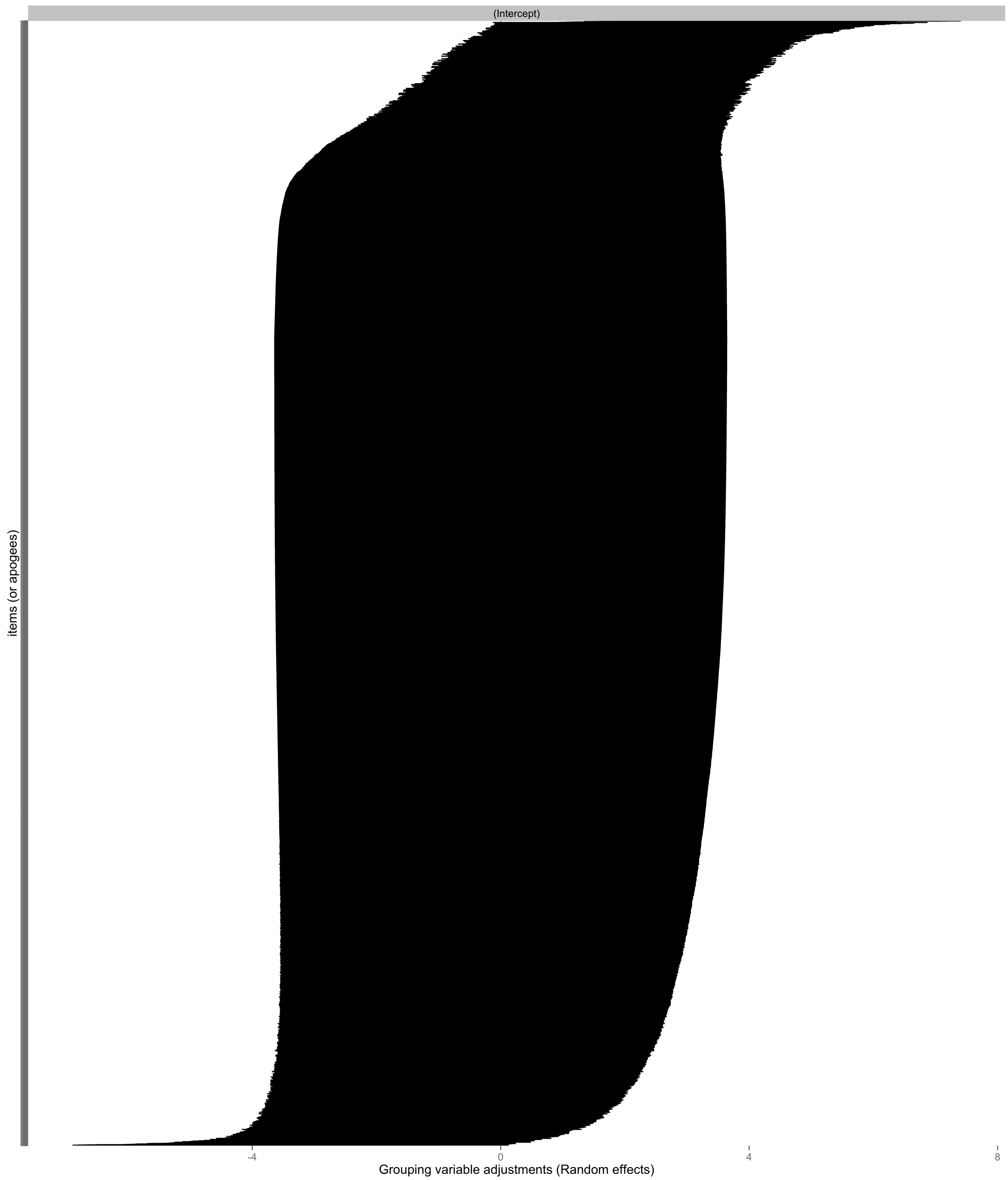


Figure 4.13: **Plot of intercept adjustments (random intercepts) for items (or apogees) of the hierarchical logistic regression model for pinky extension** Although it is difficult to read individual words, as discussed in detail below, there is not much systematic variation of pinky extension between trials. The sigmoidal shape is due to the fact that the intercept adjustments are modeled on a normal distribution. The levels on the y-axis are items (or apogees), and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

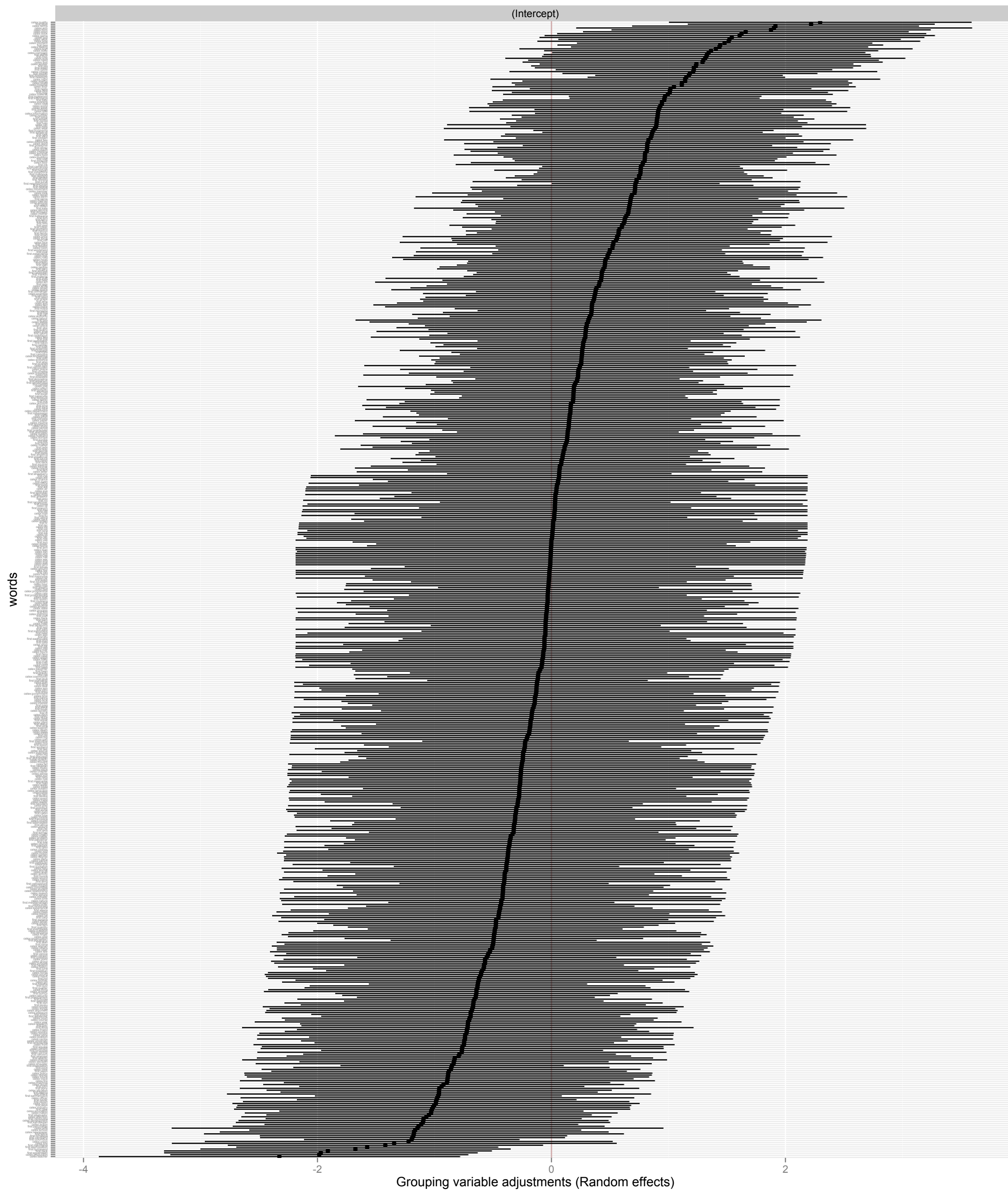


Figure 4.14: **Plot of intercept adjustments (random intercepts) for words nested in word lists of the hierarchical logistic regression model for pinky extension** Because there are a large number of words, there are many levels on the y-axis. Although it is difficult to read individual words, as discussed in detail below, there is not much systematic variation of pinky extension between words. The sigmoidal shape is due to the fact that the intercept adjustments are modeled on a normal distribution. The levels on the y-axis are words (with the word list prefixed to them, to show the nested structure), and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

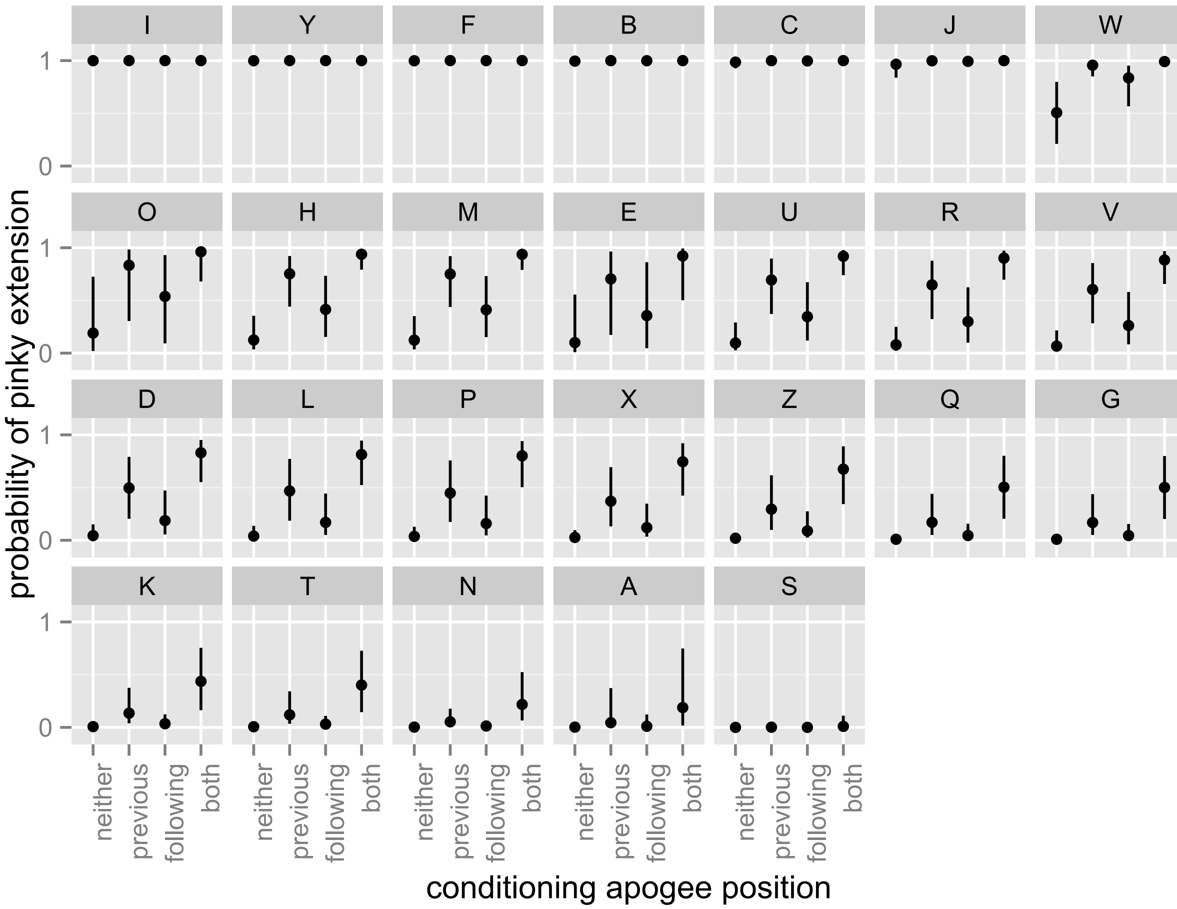


Figure 4.15: **A plot showing the effect of conditioning apogees (-I-, -J-, and -Y-) on the probability of pinky extension at mean transition times and mean hold durations for both previous and following** Dots are model predictions for an apogee with a conditioning apogee in the previous position, following position, both, or neither. The lines are 2 standard deviations on either side. The order of the FS-letters is based on the overall amount of pinky extension.

Model predictions from the regression are visualized in figure 4.15. Here we can see that apogees with handshapes that specify pinky extension (-I-, -J-, -Y-, -F-, -B-, or -C-) almost all have pinky extension in their hand configuration as we expect (they are near ceiling). For apogees of all of the other FS-letters we can see the effect that a conditioning, surrounding apogee (FS-letter: -I-, -J-, or -Y-) has on the probability that an apogee's hand configuration will have an extended pinky. For apogees of FS-letters that do not have pinky extension in their handshapes, the probability that the hand configuration is realized with an extended pinky is nearly zero if there is no -I-, -J-, or -Y-before or after. For some of these FS-letters (in particular -W-, -O-, -M-, -H-, -U-, -E-, -R-, and -V-), that probability is higher if there is an -I-, -J-, or -Y- apogee before or after, and increases greatly if there is an -I-, -J-, or -Y- both before and after.

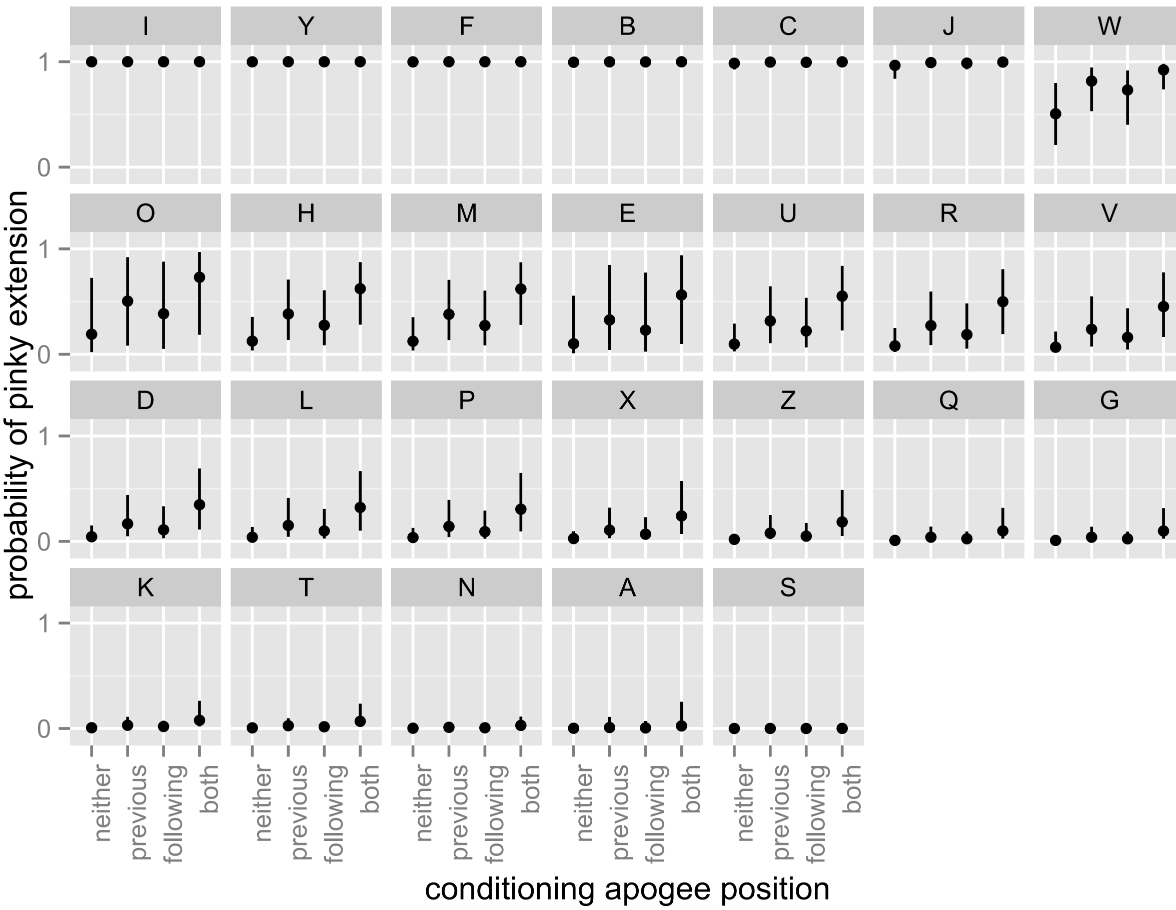


Figure 4.16: **A plot showing the effect of conditioning apogees (-B-, -C-, and -F-) on the probability of pinky extension at mean transition times and mean hold durations for both previous and following** Dots are model predictions for an apogee with a conditioning apogee in the previous position, following position, both, or neither. The lines are 2 standard deviations on either side. This is the same style of plot as figure 4.15, with the only difference being that the conditioning handshape here is a -B-, -C-, or -F-.

We have found that although an -I-, -J-, or -Y- on either side of an apogee conditions coarticulatory pinky extension, a -B-, -C-, or -F- conditions pinky extension less strongly (see figure 4.16). The generalization is that when a pinky is extended along with other fingers (especially the ring and middle fingers), there is less coarticulated pinky extension in surrounding apogees. Although this seems like an odd distinction, it is quite natural when we look at the physiology of the hand. There are three extensors involved in finger (excluding thumb) extension: *extensor indicis proprius* (for the index finger), *extensor digiti minimi* (for the pinky finger), and *extensor digitorum communis* (for all of the fingers) (Ann, 1993). When extended with the other fingers there are two extensors acting on the pinky, where as when it is extended alone there is only a single extensor. Additionally when the pinky is extended and the ring finger is flexed, it must act against the *juncturae tendinum* which connects the pinky to the ring finger. This asymmetry results in slower, less precise pinky extension when the pinky is extended alone, compared to when the other fingers are extended with it. We suggest that it is this muscular asymmetry that accounts for the fact that -I-, -J-, and -Y- condition coarticulation more than -B-, -C-, and -F-.

Figure 4.17 visualizes the effect of transition time and the handshape of surrounding -I-, -J-, or -Y- apogees for the FS-letter -L-. As before, the x-axis in this plot is the location of a conditioning handshape and the y-axis is the probability of pinky extension. The horizontal facets (boxes) are the z-score of the log transformed local transition time[6]. We can see that for apogees that have a conditioning handshape in either the following or both apogees, the probability of pinky extension is high at short local transition times (negative z-scores), but is much lower when the local transition time is longer (positive z-scores). Apogees that have a previous conditioning handshape do not vary much based on transition time. Finally, apogees that do not have a conditioning handshape in either apogee are near 0 regardless of the transition time. The main point is that if there is a conditioning apogee as the following apogee, the local transition time magnifies the effect of a conditioning handshape when it is short, and attenuates it when it is long.

Although previous and following transition times do not have a large main effect, the interaction between the handshape of the previous and following apogees and the previous and following transition times, respectively, are significant. This interaction is not surprising (quick signing or speech

6. Where 0 represents the mean value, −1 represents a transition that is one standard deviation shorter than the mean, and +1 represents one standard deviation longer than the mean.

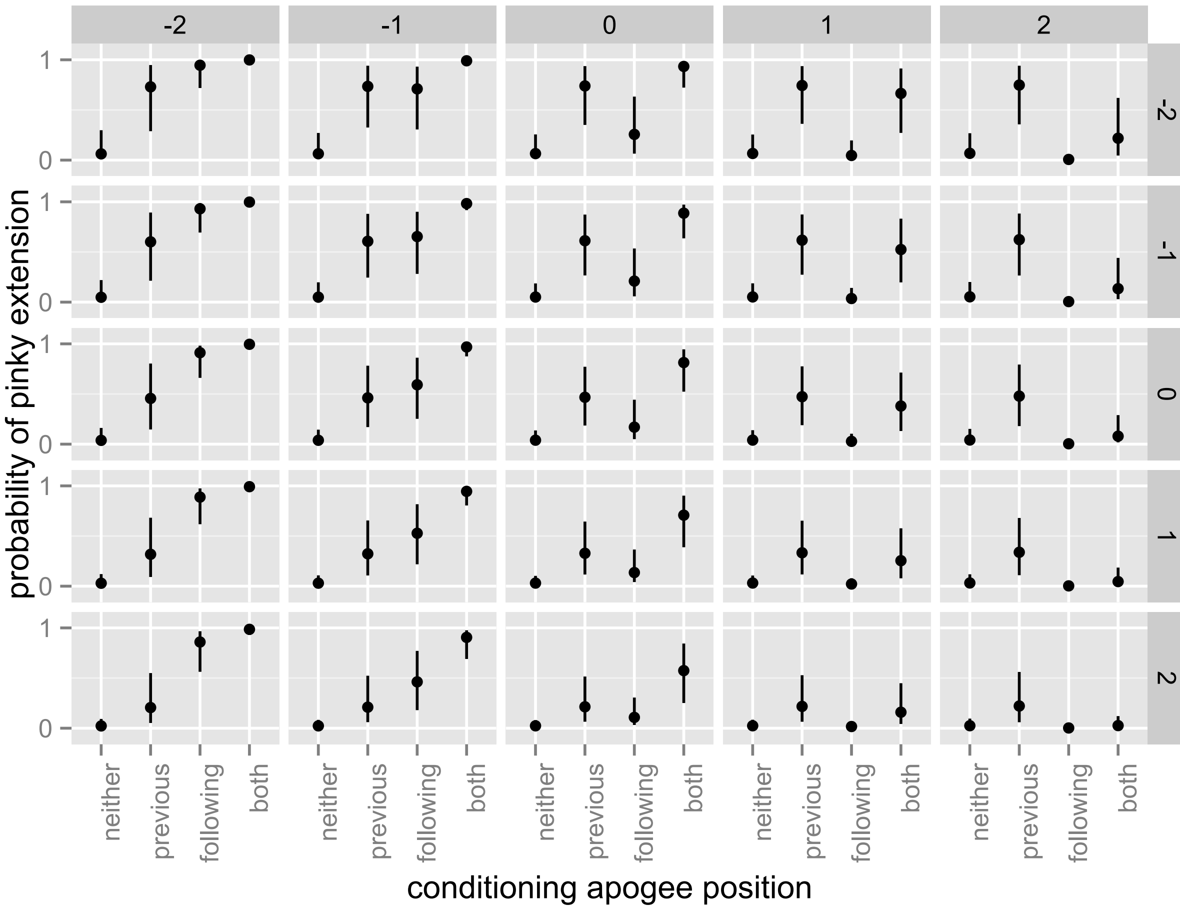


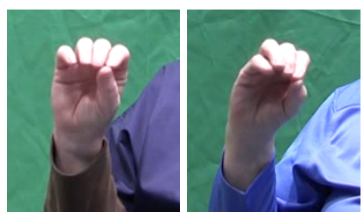
Figure 4.17: **A plot showing the effect of conditioning apogees (-I-, -J-, and -Y-) and transition times on the probability of pinky extension for the FS-letter -L- only at mean hold durations**, faceted by previous and following transition time (z-scores of the log transform, where smaller values are shorter transitions).

results in more coarticulation see (Cheek, 2001) for hand configuration coarticulation in ASL), but it is surprising that there is less interaction between previous handshape and previous transition time (the effect is smaller, and seems to be only for -B-, -C-, and -F- condition FS-letters). One possible explanation for this is that there is an asymmetry between flexion and extension of the pinky. As stated above, the pinky and ring fingers are connected to each other by the *juncturae tendinum* while this ligamentous band cannot exert its own force, it connects the pinky and ring fingers, and will be stretched if the fingers are not in the same configuration (either flexed or extended) (Ann, 1993). For this reason we can expect that pinky extension alone will be slower than pinky flexion alone when the ring finger is also flexed. This is because only the extension is acting against the *juncturae tendinum*, where as flexion would be acting in concert with it. Whereas, pinky flexion is easier when the ring finger is flexed because it relieves the tension on the *juncturae tendinum*, so there is no physiological force that forces the pinky to remain extended. In other words, due to the physiology of the hand we expect to see slower pinky extension, but faster pinky flexion when the ring finger is flexed. This is confirmed in our data: we see an interaction with time for only following apogees. That is, this coarticulation is time dependent only when it is regressive, not when it is progressive.

In order to test if a selected, flexed pinky in the apogee of interest had an effect on the amount of pinky extension, the -A- and -S- as well as -E- and -O- apogees were included as a predictor. Apogees of -A- and -S- showed significantly less pinky extension coarticulation. Looking at the model predictions in figures 4.16 and 4.15, this is clear because both are at the bottom of either plot, with the least pinky extension overall. Additionally, even when there are condition handshapes on either side, they do not show nearly any pinky extension. In fact, for -S-, out of 1284 annotations of 584 apogees not a single annotation was an annotation for pinky extension. For -A-, out of 3388 of 1567 apogees, 22 annotations (across 11 apogees) were annotations that mark the pinky as extended.

This is especially striking when we compare the -A- and -S- handshapes with other handshapes that canonically have a similarly flexed, but nonselected pinky (particularly, -M-, -H-, -U-, -R-, -V-, -D-, -L-, -P-, -X-, -Z-, -Q-, and -G-) which all exhibit much more pinky extension coarticulation. Although the canonical end result of both of these groups of FS-letters is a completely flexed pinky, a subset (-A- and -S-) seem to resist pinky extension coarticulation so strongly.

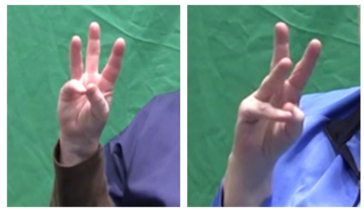
Apogees of -E- and -O- are different: they seem to both have more pinky extension with both being in the top half for overall pinky extension. On the surface this is surprising because the articulatory model predicts that all handshapes that have the pinky selected and flexed should exhibit lower amounts of pinky extension coarticulation. However, the handshape for -E- and -O- has a pinky configuration that is very close to the boundary for extension given our coding scheme (see figure 4.18). Interestingly, with one exception, -E- and -O- are the only FS-letters, when they are not near a conditioning apogee, where confidence intervals of the model predictions of pinky extension overlap 50%. This means that for both of these, the model is especially not confident in predicting the pinky extension of -E- or -O-, which is exactly what we would expect given that the canonical configurations for both of these letters is so close to the boundary for the annotation task. Since the canonical configuration of the pinky for -E- and -O- is so close to the artificial boundary for extension it is not surprising a number of -E- and -O- apogees have pinky extension. This particular phenomenon is an artifact of the coding task. A more gradient measure of pinky extension would allow for this artifact of the coding system we used to be overcome, which would in turn allow for direct testing of how well -E- and -O- apogees followed the patterns of the other pinky selected and flexed handshapes.



(a) -O- (b) -O-

Figure 4.18: Apogees from (a) C-**O**-U-P-L-E, and (b) F-**O**-**O**-D

Another letter stands out as surprising: the FS-letter -W-. It has a large amount of pinky extension, even when not surrounded by conditioning apogees. This is so much so the case that the model prediction for the chance pinky extension when a -W- has no conditioning apogees on either side is greater then 50%. Investigating images of -W- apogees, the reason is clear: many of the examples of -W- look more like a handshape that has been labeled as the 6 handshape in ASL. In traditional phonological specifications the -W- has the index, middle, and ring fingers selected and extended while the thumb and index fingers are nonselected and flexed. The 6 handshape has the thumb and pinky finger selected, in either bent or ring configuration, with the index, middle, and ring fingers nonselected and extended. In the 6 handshape, the pinky and the thumb usually touch at the tip, and are not fully flexed. Whereas, with the traditional -W- handshape the pinky should be fully flexed, with the thumb holding it down. Impressionistic analysis of images of -W- apogees reveals that many look more like the 6 handshape than they do the traditional -W- handshape (see figure 4.19 for examples).



(a) -W- (b) -W-

Figure 4.19: Apogees from (a) T-O-**W**-N, and (b) **W**-I-N-G

There are a few other FS-letters that also exhibit little pinky extension coarticulation: -K-, -T-, and -N-. There is one possible explanation for the FS-letters -T- and -N- having little coarticulation: both are letters that often show up in the digraphs (or fused apogees) -TI- and -NI- which were described in the methods section (3.2) of the timing analysis chapter. These are in some ways extreme examples of coarticulation where two apogees have fused together temporally, although for this coarticulation analysis, all fused apogees were removed from the data because it is not clear how to represent the temporal properties of these fused apogees. Removing these may be artificially lowering our ability to see the coarticulatory effects for -T- and -N- apogees. There are not a huge number of these so far, but this is one area that is ripe for future study.

As discussed in specifics above, individual letters exhibit a large amount of variation. Some letters show more pinky extension than others. More detailed work is needed to investigate if this due to differences in the phonological specifications of each handshape beyond the pinky being selected or not, which might make them more or less susceptible to pinky extension coarticulation. Additionally, the curious case of -W-, as well as the findings in chapter 2, suggests that some of the phonological specifications that have been proposed might need revisiting. See figure 4.10 for a visualization of the intercept and slope adjustments in the model for FS-letters.

There is some intersigner variation: signer 3 has a lower overall probability of pinky extension compared with the other signers, and signers 1 and 2 have a slightly higher overall probability of pinky extension. Although, this variation is much, much smaller than the variation seen in the timing analysis in chapter 3. See figure 4.11 for a visualization of the intercept and slope adjustments in the model for signers.

Although there are annotators that are significantly more or less likely than average to annotate an apogee as extended or flexed (their confidence intervals do not overlap zero), no individual annotator has an estimate of larger than 2 or smaller than -2, compare this with, for example, the amount of variation seen by the FS-letter identity, which ranges from -4 to 4. See figure 4.12 for a visualization of the intercept adjustments in the model for annotators. Additionally, because annotators are included as a grouping variable, the coefficients (and thus the predictions) made by the model remove this variation, and instead predict pinky extension given an average annotator.

There is not much systematic variation across words. See figure 4.14 for a visualization of the intercept adjustments in the model for words. And there is not much systematic variation across individual apogees (items). See figure 4.13 for a visualization of the intercept adjustments in the model for individual apogees (items).

4.2.3 Discussion

We have seen that there does appear to be coarticulation with respect to the pinky finger: an extended pinky in a neighboring apogee will increase the probability that an apogee (that is not otherwise specified for pinky extension) will have pinky extension in its hand configuration. This is exacerbated by transition times that are shorter, and attenuated by transition times that are longer, greatly for conditioning apogees that follow the apogee of interest, but less so for conditioning apogees that precede it.

The set of FS-letters that condition the most coarticulation is initially a bit surprising: it is not all of the FS-letters that have handshapes with pinky extension (-B-, -C-, -F-, -I-, -J-, and -Y-) equally, but rather more so those where the pinky is extended and other fingers are flexed (-I-, -J-, and -Y-). This asymmetry is explained by the physiology of the hand: because when the pinky extensor acts alone it acts slower than when it is used in combination with the common extensor. Thus signers allow pinky extension to overlap across other apogees in order to maintain an overall rhythmic timing.

The fact that there is an interaction between conditioning handshape and time only for apogees following the apogee of interest has a similar explanation. Because the pinky is connected to the ring finger, it will be harder, and thus slower, to extend the pinky when the ring finger is completely flexed. And like before, in order to maintain the overall timing of apogees in fingerspelling, the pinky must be extended earlier, intruding into the hand configuration of earlier apogees that do not have pinky extension in their handshape.

4.3 Conclusions

This chapter has used a quantitative analysis of pinky extension coarticulation as evidence for the differential status of selected versus nonselected (that is, active versus nonactive) fingers in handshapes involved in ASL fingerspelling. Using the articulatory model that was established in chapter 2, as well as principles of articulator activation from articulatory phonology a number of specific hypotheses follow:

A. Because gestures are dynamic, individual handshapes and the articulators that make up the hand will not be static, sequential elements (i.e. discrete FS-letters). Rather, individual articulator gestures, involving all parts of the hand (e.g. digits, wrist), will overlap across several hand configurations (apogees).
B. The hand configuration of a specific instance of a given FS-letter will vary in predictable ways based on the surrounding context.

Broadly speaking both of these are confirmed. First, through the observation of the time course of extension in the case studies, as well as with the broader quantification of pinky extension in a large corpus of fingerspelling, there is considerable gradient activation of the articulators, these periods of activation generally aim at articulatory targets, and sometimes temporally overlap with other gestures. Analyzing the distribution of pinky extension in the corpus, we see that the context surrounding an apogee contributes to whether or not the pinky will be extended. Following the broad hypothesis B above, there are detailed hypotheses about this contextual variation:

1. The nonselected fingers are more frequently the targets of coarticulatory pressure (vs. selected fingers).

2. The selected fingers are the sources of coarticulatory pressure.

3. Finger configuration that is due to (phonetic) coarticulatory pressure will differ from configuration due to phonological specification.

Hypothesis 3 is supported by the case studies, where the deviations of articulators from their target do not all match the exact extension of the conditioning segment, but are frequently somewhere in between full extension and full flexion. Hypothesis 2 is supported by the fact that of the possible conditioning FS-letters -B-, -C-, -F-, -I-, -J-, and -Y- the ones that condition coarticulation the most (-I-, -J-, and -Y-) have the pinky selected and extended. Finally, hypothesis 1 is supported by the fact that FS-letters where the pinky is selected and flexed (-A- and -S-) exhibit less pinky extension than FS-letters where the pinky is similarly flexed, but nonselected. This last part is clear evidence that there are differential categories of activation associated with different articulatory gestures. Although every articulator has some gesture associated with it, some gestures—those of active or selected fingers—are stronger than others—those of nonactive or nonselected fingers. When the gestures overlap (e.g. as the result of a particular gesture and articulator being slower or less controlled) there will be a gradient activation, resulting in coarticulation, like that of the pinky extension modeled here.

This work has built an articulatory model for handshape in sign languages (the articulatory model of handshape), that bridges a gap between phonological models of handshape, and the phonetic realities of hand configuration in signing. This model makes predictions about contextual variation observed with respect to hand configuration, which have been confirmed with data from a large corpus of ASL fingerspelling. Further, this work contributes to articulatory phonology, as well as theories of speech production broadly, by studying the distinction between active and nonactive articulator gestures. Handshape in sign languages is especially well-suited to study this phenomenon because there are many possible combinations of active and nonactive articulators (all five digits), additionally, unlike most articulators for spoken languages, the articulators can be seen and tracked easily without the occlusion of the cheeks and neck.

Chapter 5

Conclusions

The phonetics-phonology interface has not been explored extensively for sign languages. This dissertation serves to rectify that in part. We have seen here the development of a model of the phonetics-phonology interface, as well as a description and analysis of detailed timing properties of fingerspelling, as well as the quantification and analysis of one form of coarticulation seen in fingerspelling.

Articulatory phonology is a type of hybrid form of phonetics and phonology that allows for, and predicts, the type of gradient variation that is seen with coarticulation by using articulator gestures that are activated over time, and can blend together, as the core phonological unit. This model is broadly compatible with dual models that have been proposed for the perception and comprehension (Poeppel *et al.* (2008); Poeppel & Idsardi (2011) among others). The articulatory model of handshape developed here builds on articulatory phonology and models of the phonology of handshape in signed languages to produce a model that links phonological specification to phonetic reality. Not only does this model make predictions about what kinds of phonetic variation we expect to see, but it has also been implemented computationally which allows for a robust testing ground of both phonological specification, as well as the mapping between phonological specifications and phonetic targets. Although this work was limited to ASL fingerspelling, the articulatory model of handshape is generalizable to ASL in general, as well as any signed language.

There has already been some work on the timing properties of fingerspelling. There are a considerable number of studies looking at fingerspelling rate, however, because of the range of methodologies used, populations studied, etc. there is a huge range of reported rates (2.18–6.5 letters per second). This work collected a large corpus of ASL fingerspelling, and analyzed not only the rate of fingerspelling, but also word-internal timing properties. We found a rate (5.84 letters per second) that is in the middle of the range of rates reported in the literature. This rate was by and large replicated using motion capture technology as well. Although it was slightly slower, that seemed to be driven by differences in the measurement of word duration for motion capture data versus regular video data.

On top of the rate data, word-internal timing properties were analyzed, and we found a number of effects on hold duration. First, the rate of fingerspelling has a large effect on the duration of holds: the faster the rate, the shorter the holds. The first and last apogees are held for much longer than word-medial apogees. Of the word medial apogees, holds tend to be the same duration with only slight differences between them. FS-letters with movement are held for longer. FS-letters with orientations that are down or to the side might be held for longer, although this is complicated by the alignment of handshape changes with orientation changes. Additionally there is a large amount of variation between signers for the overall duration of their holds. Transition durations vary more based on the orientation or movement of the apogee before them than for apogees after them. This could be evidence that orientation and handshape changes are aligned to the beginning of the holds for apogees, as opposed to both the beginning and ends of holds. Transitions get considerably shorter in later positions in words. Finally, there is a large amount of variation between signers' transitions as well. Strikingly, the inter-signer variation for holds and transitions does not follow the same pattern: signers with long holds do not necessarily also have long transitions. Rather, there is considerable variation in the ratio of holds to transitions among different signers. This variation might also explain the huge range of rates reported in the literature, since most studies included only a few signers, it is not surprising, given the huge amount of variation, that there is a wide range of rates reported for fingerspelling.

The last part of this work is a quantitative analysis of pinky extension coarticulation. This was used as evidence for the differential status of selected versus nonselected (that is active versus non-active) fingers in handshapes involved in ASL fingerspelling. Returning to the predictions of the articulatory model that was established in chapter 2, as well as principles of articulator activation from articulatory phonology a number of specific hypotheses follow:

A. Because gestures are dynamic individual handshapes and the articulators that make up the hand will not be static, sequential elements (i.e., discrete FS-letters), but rather individual articulator gestures, involving all parts of the hand (e.g., digits, wrist), will overlap across several hand configurations (apogees).
B. The hand configuration of a specific instance of a given FS-letter will vary in predictable ways based on the surrounding context.

Broadly speaking both of these are confirmed. First, through the observation of the time course of extension in the case studies, as well as with the broader quantification of pinky extension. Analyzing the distribution of pinky extension in the corpus, we see that the context surrounding an apogee contributes to whether or not the pinky will be extended. Following the broad hypothesis B above, there are detailed hypotheses about this contextual variation:

1. The nonselected fingers are more frequently the targets of coarticulatory pressure (vs. selected fingers).
2. The selected fingers are the sources of coarticulatory pressure.
3. Finger configuration that is due to (phonetic) coarticulatory pressure will differ from configuration due to phonological specification.

Hypothesis 3 is supported by the case studies, where the deviations of articulators from their target do not all match the exact extension of the conditioning segment, but are frequently somewhere in between full extension and full flexion. Hypothesis 2 is supported by the fact that of the possible conditioning FS-letters -B-, -C-, -F-, -I-, -J-, and -Y- the ones that condition coarticulation the most (-I-, -J-, and -Y-) have the pinky selected and extended. Finally, hypothesis 1 is supported by the fact that FS-letters where the pinky is selected and flexed (-A- and -S-) exhibit less pinky extension than FS-letters where the pinky is similarly flexed, but nonselected. This last part is clear evidence that there is differential categories of activation associated with different articulatory gestures. Although every articulator has some gesture associated with it, some gestures (the active gestures) are stronger than others (the nonactive gestures). When the gestures overlap (e.g. as the result of a particular gesture and articulator being slower or less controlled) there will be a gradient activation, resulting in coarticulation, like that of the pinky extension modeled here.

We have seen that although there is a lot of variation even in the small number of features that we looked at in this work, there is often structure in this variation that is not just random, but rather is based on linguistic properties. With respect to hand configuration the variation is not a simple averaging of the surrounding configurations, but rather is structured: the active (or selected) articulators will be less contextually influenced than articulators that are non-active (or nonselected). With respect to timing, much of the variation is due to inter-signer variation, but there is additionally variation in hold duration based on the phonological orientation or movement of the apogee, as well as variation based simply on the FS-letter of the apogee.

Finally, this work contributes to articulatory phonology, as well as theories of speech production broadly by studying the distinction between active and nonactive articulator gestures. Handshape in sign languages is especially well-suited to study this phenomenon because there are many possible combinations of active and nonactive articulators (all five digits), additionally, unlike most articulators for spoken languages, the articulators can be seen and tracked easily without the occlusion of the cheeks and neck.

5.1 Going beyond fingerspelling

Although fingerspelling is a distinct part of the ASL lexicon, these findings have a few implications for understanding the phonetics and phonology of the rest of ASL, and signed languages in general. As noted above, fingerspelling is distinct from other parts of ASL, because it uses only handshape (and for a very small number of FS-letters orientation and movement) for contrast. The rest of the ASL lexicon uses not only handshape, but also movement, location, orientation, and non-manuals to drive lexical contrasts. The articulatory model of handshape was designed to be a theory of the phonetics-phonology interface that is directly applicable to the rest of ASL, and sign languages generally.

For the timing results, because the other parameters all involve joints that are more proximal, and thus drive the movement of larger articulators (e.g. the elbow moves both the hand and the forearm, the shoulder moves the hand, forearm, and upper arm), segments that contrast across these parameters will likely be slower than segments in fingerspelling which contrast over basically just the joints of the hand. For this reason, the timing properties cannot be straightforwardly generalized to lexical ASL signing, however some of the findings for fingerspelling could hold for signing more broadly. For example, we expect the signer variation found in fingerspelling will be present in comparable amounts for lexical signing. There is some work on prosodic patterns found in ASL and other sign languages, and the positional differences found in fingerspelling are similar to those found in signing: the last sign of an utterance is generally longer than utterance medial signs (Liddell (1978); Wilbur (1999), among many others). It is possible that the pattern found in fingerspelling with respect to holds and transitions in different positions of the word is similar for lexical signs: in fingerspelling, word medial holds are all generally the same duration with only slightly shorter holds in later positions in words, however, the transitions show a significant reduction in duration in later positions in the word. In other words, as the word goes on, signers generally speed up the overall rate of fingerspelling by shortening the transitions but not the holds. This pattern should be tested at the utterance level for ASL and other sign languages: compare the durations of the lexical portions of the signs (since some signs involve movement just holds would not suffice for this definition) to the transitions between these signs. Finally, the methods used to determine fingerspelling location from motion capture data are the beginnings of methods to determine and distinguish the location of signs within the signing space. Of course there are more than two locations, but Hidden Markov Models with more than two states can be implemented and used to detect this distinct locations that the hand is in during lexical signing.

The coarticulation results for pinky extension are specific to ASL fingerspelling, but we expect that they come from ASL general (and possibly sign language general) phonological constraints. There has been some work that shows that similar kinds of handshape coarticulation occur in lexical signing (Cheek, 2001; Mauk, 2003). However, neither had enough (or the precise stimuli) to robustly test the difference between active and nonactive articulators with respect to this coarticulation (although it was noted by Mauk (2003)). Ongoing work is underway to use methods similar to both of these previous studies, with stimuli that have the right stimuli combinations to test if this active/nonactive distinction holds for coarticulation seen in ASL signing generally.

5.2 Broader impacts

The articulatory model of handshape has a number of broader impacts, the main class of which is that, as a computationally implemented model of phonetics and phonology, it will allow for the types of variation that are seen in naturalistic language to be transferred into artificial models of sign production. The module even connects with a 3D hand renderer (`libHand`) to test the output of the model, which can be extended to produce videos of handshape for use as experimental stimuli, to test other models of handshape phonology, etc.

The analysis of timing properties also has a number of broader impacts. Findings from this timing data have been instrumental in additional studies that are ongoing (Keane & Geer, 2014; Geer & Keane, 2014). The temporal analysis discussed in this work was critical to both the formulation of this work, as well as construction of stimuli. Additionally, the timing analysis described here is a critical first step in the analysis of pinky extension coarticulation discussed in chapter 4.

This work establishes general norms for fingerspelling in native ASL users (e.g., transition times, apogee hold durations). Having quantitative norms of specific features of fingerspelling allows for the development of metrics and tests for what types of productions fall outside of the range of typical signers. Norms for typical signers are needed before analyzing how people from different language backgrounds (early learners, late second language learners, etc.) differ in their fingerspelling. This has further impacts on diagnosing language disorders, which has been particularly understudied in ASL signers.

The analysis of pinky extension builds on these further, automatic recognition work for sign languages has frequently found handshape recognition to be a particularly difficult problem. In order to successfully recognize and classify rapidly changing handshapes, researchers must have models of coarticulation like those studied here.

There has been research showing a correlation between fingerspelling ability and literacy (Haptonstall-Nykaza & Schick, 2007; Emmorey & Petrich, 2011). Understanding basic phonetic facts about the production of fingerspelling will allow for more detailed future work on the perception of fingerspelling. Furthermore, understanding how fingerspelling is produced and perceived will enable the study of this correlation in more detail.

Finally, because the task of fingerspelling is unlike many of the common tasks that humans use their hands for, the motor movements involved are not well represented in the literature on motor control. There is literature on grasping, as well as the hand operating against a rigid surface (eg, a keyboard, a musical instrument), but little literature on the rapid, fine motor movements required to form the handshapes necessary for fingerspelling in free space.

Appendix A

AMOHS source

A.1 hc.py

```
##### Error classes #####
class digitError(Exception):
    pass


class jointError(Exception):
    pass


##### checking functions that make sure values are sane
##### variables defining various specifications #####
jointWeight = {"wrist": 4,
            "cm":3,
            "mcp":3,
            "pip":2,
            "ip":1,
            "dip":1}


##### handshape class and recursion #####
class armconfiguration:
    """Representation for arm configuruations"""
    def __init__(self, hand, wrist):

        if isinstance(wrist, joint):
                wrist = wrist
```

```
        else:
                if ((type(wrist) is list) or (type(wrist) is tuple))
                   and len(wrist) == 3:
                        wrist = joint(dfFlex=wrist[0], dfRot=wrist[1],
                           dfPro=wrist[2])
                else:
                        raise digitError("The wrist joint needs a list
                           or tuple with exactly 3 degrees of freedom
                           specified, got %s instead." % (str(wrist)))
        if wrist.df != 3:
                raise digitError("The wrist joint needs 3 degrees of
                   freedom, got %s instead." % (str(wrist.df)))
        else:
                self.wrist = wrist
        self.hand = hand

    def __repr__(self):
        return "%s(wrist=%r, hand=%r)" % (self.__class__.__name__, self
           .wrist, self.hand)

    def __str__(self):
        return """armconfiguration:
wrist: %s
hand: %s
""" % (self.wrist, self.hand)

    def __sub__(self, other):
```

```
        if self.wrist and other.wrist: wristDiff = self.wrist - other.
           wrist
        if self.hand and other.hand: handDiff = self.hand - other.hand
        return armconfigurationDelta(wrist=wristDiff, hand=handDiff)

class armconfigurationDelta(armconfiguration):
    def totalDegreesDifferent(self):
        degDiff = sum([self.wrist.totalDegreesDifferent(),
                       self.hand.totalDegreesDifferent()])
        return degDiff

    def weightedDegreesDifferent(self):
        degDiff = sum([self.wrist.totalDegreesDifferent()*jointWeight["
           wrist"],
                       self.hand.weightedDegreesDifferent()])
        return degDiff

class handconfiguration:
    """Representation for hand configruations"""
    def __init__(self, index, middle, ring, pinky, thumb):
        self.index = index
        self.middle = middle
        self.ring = ring
        self.pinky = pinky
        self.thumb = thumb

    def __repr__(self):
```

```
        return "%s(index=%r, middle=%r, ring=%r, pinky=%r, thumb=%r)" %
            (self.__class__.__name__, self.index, self.middle, self.
           ring, self.pinky, self.thumb)
    def __str__(self):
        return """Handconfiguration:
index: %s
middle: %s
ring: %s
pinky: %s
thumb: %s""" % (self.index, self.middle, self.ring, self.pinky, self.
   thumb)

    def __sub__(self, other):
        if self.index and other.index: indexDiff = self.index - other.
           index
        if self.middle and other.middle: middleDiff = self.middle -
           other.middle
        if self.ring and other.ring: ringDiff = self.ring - other.ring
        if self.pinky and other.pinky: pinkyDiff = self.pinky - other.
           pinky
        if self.thumb and other.thumb: thumbDiff = self.thumb - other.
           thumb
        return handconfigurationDelta(index=indexDiff, middle=
           middleDiff, ring=ringDiff, pinky=pinkyDiff, thumb=thumbDiff)

class handconfigurationDelta(handconfiguration):
    def totalDegreesDifferent(self):
        degDiff = sum([self.index.totalDegreesDifferent(),
```

```
                        self.middle.totalDegreesDifferent(),
                        self.ring.totalDegreesDifferent(),
                        self.pinky.totalDegreesDifferent(),
                        self.thumb.totalDegreesDifferent()])
        return degDiff

    def weightedDegreesDifferent(self):
        degDiff = sum([self.index.weightedDegreesDifferent(),
                        self.middle.weightedDegreesDifferent(),
                        self.ring.weightedDegreesDifferent(),
                        self.pinky.weightedDegreesDifferent(),
                        self.thumb.weightedDegreesDifferent()])
        return degDiff

class finger:
    """A finger"""
    def __init__(self, MCP, PIP, DIP):
        # ensure the that MCP is a joint instance, and has 2 degrees of
            freedom specified.
        if isinstance(MCP, joint):
                MCP = MCP
        else:
                if ((type(MCP) is list) or (type(MCP) is tuple)) and
                    len(MCP) == 2:
                        MCP = joint(dfFlex=MCP[0], dfAbd=MCP[1])
                else:
```

```
            raise digitError("The MCP joint needs a list or
                tuple with exactly 2 degrees of freedom
               specified, got %s instead." % (str(MCP)))
if MCP.df != 2:
        raise digitError("The MCP joint needs 2 degrees of
           freedom, got %s instead." % (str(MCP.df)))
else:
        self.MCP = MCP

# ensure the that PIP is a joint instance, and has 1 degree of
   freedom specified.
if isinstance(PIP, joint):
        PIP = PIP
else:
        PIP = joint(PIP)
if PIP.df != 1:
        raise digitError("The PIP joint needs 1 degree of
           freedom, got %s instead." % (str(PIP.df)))
else:
        self.PIP = PIP
# ensure the that DIP is a joint instance, and has 1 degree of
   freedom specified.
if isinstance(DIP, joint):
        DIP = DIP
else:
        DIP = joint(DIP)
if DIP.df != 1:
```

```
                raise digitError("The DIP joint needs 1 degree of
                    freedom, got %s instead." % (str(DIP.df)))
        else:
                self.DIP = DIP

    def __repr__(self):
        return "%s(MCP=%r, PIP=%r, DIP=%r)" % (self.__class__.__name__,
            self.MCP, self.PIP, self.DIP)

    def __str__(self):
        return """
  MCP: %s
  PIP: %s
  DIP: %s""" % (self.MCP, self.PIP, self.DIP)

    def __sub__(self, other):
        if self.MCP and other.MCP: MCPDiff = self.MCP - other.MCP
        if self.DIP and other.DIP: PIPDiff = self.PIP - other.PIP
        if self.PIP and other.PIP: DIPDiff = self.DIP - other.DIP
        return fingerDelta(MCP=MCPDiff, PIP=PIPDiff, DIP=DIPDiff)

class fingerDelta(finger):
    def totalDegreesDifferent(self):
        degDiff = sum([self.MCP.totalDegreesDifferent(),self.PIP.
           totalDegreesDifferent(),self.DIP.totalDegreesDifferent()])
        return degDiff

    def weightedDegreesDifferent(self):
```

```
        degDiff = sum([self.MCP.totalDegreesDifferent()*jointWeight["
           mcp"],self.PIP.totalDegreesDifferent()*jointWeight["pip"],
           self.DIP.totalDegreesDifferent()]*jointWeight["dip"])
        return degDiff

class thumb:
    """the thumb"""
    def __init__(self, CM, MCP, IP):
        # ensure the that CM is a joint instance, and has 3 degrees of
           freedom specified.
        if isinstance(CM, joint):
                CM = CM
        else:
                if ((type(CM) is list) or (type(CM) is tuple)) and len(
                   CM) == 3:
                        CM = joint(dfFlex=CM[0], dfAbd=CM[1], dfRot=CM
                           [2])
                else:
                        raise digitError("The CM joint needs a list or
                           tuple with exactly 2 degrees of freedom
                           specified, got %s instead." % (str(CM)))
        if CM.df != 3:
                raise digitError("The CM joint needs 2 degrees of
                   freedom, got %s instead." % (str(CM.df)))
        else:
                self.CM = CM
        # ensure the that MCP is a joint instance, and has 1 degree of
           freedom specified.
```

```
        if isinstance(MCP, joint):
                MCP = MCP
        else:
                MCP = joint(MCP)
        if MCP.df != 1:
                raise digitError("The MCP joint needs 1 degree of
                    freedom, got %s instead." % (str(MCP.df)))
        else:
                self.MCP = MCP
        # ensure the that IP is a joint instance, and has 1 degree of
            freedom specified.
        if isinstance(IP, joint):
                IP = IP
        else:
                IP = joint(IP)
        if IP.df != 1:
                raise digitError("The IP joint needs 1 degree of
                    freedom, got %s instead." % (str(IP.df)))
        else:
                self.IP = IP

    def __repr__(self):
        return "%s(CM=%r, MCP=%r, IP=%r)" % (self.__class__.__name__,
            self.CM, self.MCP, self.IP)

    def __str__(self):
        return """
CM: %s
```

```
  MCP: %s
  IP: %s""" % (self.CM, self.MCP, self.IP)

    def __sub__(self, other):
        if self.CM and other.CM: CMDiff = self.CM - other.CM
        if self.MCP and other.MCP: MCPDiff = self.MCP - other.MCP
        if self.IP and other.IP: IPDiff = self.IP - other.IP
        return thumbDelta(CM=CMDiff, MCP=MCPDiff, IP=IPDiff)

class thumbDelta(thumb):
    def totalDegreesDifferent(self):
        degDiff = sum([self.MCP.totalDegreesDifferent(),self.IP.
           totalDegreesDifferent(),self.CM.totalDegreesDifferent()])
        return degDiff

    def weightedDegreesDifferent(self):
        degDiff = sum([self.MCP.totalDegreesDifferent()*jointWeight["
           mcp"],self.IP.totalDegreesDifferent()*jointWeight["ip"],self
           .CM.totalDegreesDifferent()*jointWeight["cm"]])
        return degDiff

##### abstract articulator classes #####
class joint:
    """a joint object"""
    def __init__(self, dfFlex=None, dfAbd=None, dfRot=None, dfPro=None)
        :
        if dfFlex and type(dfFlex) is not int:
```

```
            raise jointError("The value for flexion must be a
               single integer. Got %s instead." % (str(dfFlex))
               )
    if dfAbd and type(dfAbd) is not int:
            raise jointError("The value for abduction must be a
                single integer. Got %s instead." % (str(dfAbd))
               )
    if dfRot and type(dfRot) is not int:
            raise jointError("The value for rotation must be a
               single integer. Got %s instead." % (str(dfRot)))
    if dfPro and type(dfPro) is not int:
            raise jointError("The value for pronation must be a
                single integer. Got %s instead." % (str(dfPro))
               )
    self.dfFlex = dfFlex
    self.dfAbd = dfAbd
    self.dfRot = dfRot
    self.dfPro = dfPro
        # Count the number of degrees of freedom that are being
           used to return the dfs.
    self.df = sum([int(item != None) for item in (self.dfFlex,self.
       dfAbd,self.dfRot,self.dfPro )])

def __sub__(self, other):
    dfFlexDiff = None
    dfAbdDiff = None
    dfRotDiff = None
    dfProDiff = None
```

```
        if self.dfFlex is not None and other.dfFlex is not None:
            dfFlexDiff = self.dfFlex - other.dfFlex
        if self.dfAbd is not None and other.dfAbd is not None:
            dfAbdDiff = self.dfAbd - other.dfAbd
        if self.dfRot is not None and other.dfRot is not None:
            dfRotDiff = self.dfRot - other.dfRot
        if self.dfPro is not None and other.dfPro is not None:
            dfProDiff = self.dfPro - other.dfPro
        return jointDelta(dfFlex=dfFlexDiff, dfAbd=dfAbdDiff, dfRot=
            dfRotDiff, dfPro=dfProDiff)

    def __repr__(self):
        return "%s(dfFlex=%r, dfAbd=%r, dfRot=%r, dfPro=%r)" % (self.
            __class__.__name__, self.dfFlex, self.dfAbd, self.dfRot,
            self.dfPro)

    def __str__(self):
        return """dfFlex: %s, dfAbd: %s, dfRot: %s, dfPro: %s""" % (
            self.dfFlex, self.dfAbd, self.dfRot, self.dfPro)

class jointDelta(joint):
    def totalDegreesDifferent(self):
        if self.dfFlex is None:
            dfFlexDiff = 0
        else:
            dfFlexDiff = abs(self.dfFlex)
        if self.dfAbd is None:
            dfAbdDiff = 0
```

```
        else:
            dfAbdDiff = abs(self.dfAbd)
        if self.dfRot is None:
            dfRotDiff = 0
        else:
            dfRotDiff = abs(self.dfRot)
        if self.dfPro is None:
            dfProDiff = 0
        else:
            dfProDiff = abs(self.dfPro)
        degDiff = sum([dfFlexDiff,dfAbdDiff,dfRotDiff,dfProDiff])
        return degDiff

##### testing #####
index = finger(MCP=(0,-15), PIP=0, DIP=0)
middle = finger(MCP=(30,0), PIP=90, DIP=0)
ring = finger(MCP=(0,0), PIP=0, DIP=0)
pinky = finger(MCP=(0,0), PIP=0, DIP=0)
thmb = thumb(CM=(0,0,0), MCP=0, IP=0)
wrist = (0,0,0)
hc1 = handconfiguration(index, middle, ring, pinky, thmb)
arm1 = armconfiguration(hc1, wrist)

index = finger(MCP=(0,0), PIP=0, DIP=0)
middle = finger(MCP=(90,0), PIP=90, DIP=0)
ring = finger(MCP=(90,0), PIP=0, DIP=0)
pinky = finger(MCP=(0,0), PIP=0, DIP=0)
thmb = thumb(CM=(0,0,0), MCP=0, IP=0)
```

```
wrist = (0,0,0)
hc2 = handconfiguration(index, middle, ring, pinky, thmb)
arm2 = armconfiguration(hc1, wrist)

hcDiff = hc1-hc2
armDiff = arm1-arm2
```

A.2 hs.py

```
import hc

##### Error classes #####
class digitError(Exception):
    pass

class jointError(Exception):
    pass

class abductionError(Exception):
    pass

class oppositionError(Exception):
    pass

##### variables defining phonological specifications #####
digits = {"index", "middle", "ring", "pinky", "thumb"}

phonoJoints = {"ext":180, "midExt":150, "mid":135, "midFlex":120, "flex
   ":90}
```

```
reverseJoints = dict(reversed(item) for item in phonoJoints.items())

phonoAbduction = {"index": {"abducted":20, "neutralAbducted":10, "
   adducted":0, "negativeAbducted":-10},
                  "middle": {"abducted":0, "neutralAbducted":5, "
                     adducted":0, "negativeAbducted":10},
                  "ring": {"abducted":-10, "neutralAbducted":-5, "
                     adducted":0, "negativeAbducted":10},
                  "pinky": {"abducted":-20, "neutralAbducted":-10, "
                     adducted":0, "negativeAbducted":10},
                  # "thumb": {"abducted":45, "neutralAbducted":30, "
                     adducted":20, "negativeAbducted":5}}
                   "thumb": {"abducted":{"opposed": None,
                                        "unopposed": (15, 27, 9)}, #l
                             "neutralAbducted":{"opposed": None,
                                        "unopposed": None},
                             "adducted":{"opposed": (-22,  13, -27), #c
                                        "unopposed": (23,  8, 0)},#g (
                                           a?)
                             "negativeAbducted":{"opposed": (-34, -24,
                                -53), #for t, using traditional methods
                                . Copied from b below, but that needs
                                some refining.
                                        "unopposed": None}#b
                              }
                   }

phonoOpposition = {"opposed":-60, "unopposed":-10}
```

```
reverseOpposition = dict(reversed(item) for item in phonoOpposition.
   items())

phonoOrientations = {"default": (0,0,0), "defaultFS":(-10,0,0), "palmIn
   ":(-75,0,80), "palmDown":(-75,0,0)}
reverseOrientations = dict(reversed(item) for item in phonoOrientations
   .items())

##### checking functions that make sure values are sane
def fingerCheck(members, digits = digits):
    """Checks that members are all in the digits set"""
    # ensure that members is a set
    if members == None:
        members = set()
    elif type(members) is str:
        members = set([members])
    else:
        members = set(members)
    if not digits.issuperset(members):
        raise digitError("At least one of the members provided is not
           in the digits set.")
    return members

def jointCheck(joint, joints = phonoJoints):
    """Checks that joint is in the joints set"""
    if joint == None:
        joint = "ext"
    if not joint in joints:
```

```
        raise jointError("The joint provided is not in the joint set.")
    return joint

def abdCheck(abd, abds = phonoAbduction):
    """Checks that abduction is in the abductoin set"""
    if abd == None:
        abd = "adducted"
    if not abd in abds:
        raise abductionError("The abduction provided ("+str(abd)+") is
           not in the abduction set.")
    return abd

def oppositionCheck(oppos, oppositions = phonoOpposition):
    """Checks that joint is in the joints set"""
    if oppos == None:
        oppos = "opposed"
    if not oppos in oppositions:
        raise oppositionError("The opposition provided is not in the
           opposition set.")
    return oppos

##### handshape class and recursion #####
class arm:
    """Representation of wrist+handshape, to be expanded with elbow and
        shoulder later"""
    def __init__(self, handshape, orientation=None):
        self.handshape = handshape
        if orientation == None:
```

```
            self.orientation = "default"
        else:
            self.orientation = orientation

    def toArmTarget(self):
        wrist = hc.joint(dfFlex=phonoOrientations[self.orientation][0],
            dfRot=phonoOrientations[self.orientation][1], dfPro=
           phonoOrientations[self.orientation][2])
        return hc.armconfiguration(hand=self.handshape.
           toHandconfigTarget() , wrist=wrist)

class handshape:
    """Representation of handshapes using the articulatory model of
       handshape"""
    def __init__(self, selectedFingers, secondarySelectedFingers, thumb
       , nonSelectedFingers):
        self.SF = selectedFingers
        self.SSF = secondarySelectedFingers
        self.thumb = thumb
        if self.SSF and not self.SF.members.isdisjoint(self.SSF.members
           ):
            raise digitError("The members of selected and secodnary
               selected finger groups overlap.")
        self.NSF =  nonSelectedFingers
        if self.NSF and self.SSF:
            self.NSF.members = digits - (self.SF.members | self.SSF.
               members)
        elif self.NSF:
```

```
        self.NSF.members = digits - (self.SF.members)
    #make SSF and NSF are None if there are no members
    if self.SSF and len(self.SSF.members) == 0:
        self.SSF.members = None
    if self.NSF and len(self.NSF.members) == 0:
        self.NSF.members = None

def toHandconfigTarget(self):
    handconfig = {
    "index" : None,
    "middle" : None,
    "ring" : None,
    "pinky" : None,
    "thumb" : None
    }

    for finger in self.SF.members:
        if finger !=  "thumb":
            handconfig[finger] = hc.finger(
                MCP=hc.joint(dfFlex=phonoJoints[self.SF.MCP.value],
                         dfAbd=phonoAbduction[finger][self.SF.abd.
                            value]),
                PIP=hc.joint(dfFlex=phonoJoints[self.SF.PIP.value])
                   ,
                DIP=hc.joint(dfFlex=phonoJoints[self.SF.PIP.value])
            )
        else:
            handconfig[finger] = hc.thumb(
```

```
                MCP=hc.joint(dfFlex=phonoJoints[self.SF.MCP.value])
                    ,
                IP=hc.joint(dfFlex=phonoJoints[self.SF.PIP.value]),
                CM=hc.joint(
                        dfFlex=phonoAbduction[finger][self.SF.abd.
                           value][self.thumb.oppos.value][0],
                        dfAbd=phonoAbduction[finger][self.SF.abd.
                           value][self.thumb.oppos.value][2],
                        dfRot=phonoAbduction[finger][self.SF.abd.
                           value][self.thumb.oppos.value][1])
            )
    if self.SSF is not None:
        for finger in self.SSF.members:
            if finger != "thumb":
                handconfig[finger] = hc.finger(
                    MCP=hc.joint(dfFlex=phonoJoints[self.SSF.MCP.
                        value],
                          dfAbd=phonoAbduction[finger][self.SSF.abd.
                             value]),
                    PIP=hc.joint(dfFlex=phonoJoints[self.SSF.PIP.
                        value]),
                    DIP=hc.joint(dfFlex=phonoJoints[self.SSF.PIP.
                        value])
                )
            else:
                    handconfig[finger] = hc.thumb(
                    MCP=hc.joint(dfFlex=phonoJoints[self.SSF.MCP.
                        value]),
```

```
                IP=hc.joint(dfFlex=phonoJoints[self.SSF.PIP.
                    value]),
                CM=hc.joint(
                        dfFlex=phonoAbduction[finger][self.SSF.
                            abd.value][self.thumb.oppos.value
                            ][0],
                        dfAbd=phonoAbduction[finger][self.SSF.
                            abd.value][self.thumb.oppos.value
                            ][2],
                        dfRot=phonoAbduction[finger][self.SSF.
                            abd.value][self.thumb.oppos.value
                            ][1])
            )

if self.NSF is not None:
    for finger in self.NSF.members:
        if self.NSF.joints.value == "ext":
            NSFAbd = "neutralAbducted"
            NSFAbd = "abducted"
        else:
            NSFAbd = "adducted"
        if finger !=  "thumb":
            handconfig[finger] = hc.finger(
                MCP=hc.joint(dfFlex=phonoJoints[self.NSF.joints
                    .value],
                      dfAbd=phonoAbduction[finger][NSFAbd]),
                PIP=hc.joint(dfFlex=phonoJoints[self.NSF.joints
                    .value]),
```

```
                DIP=hc.joint(dfFlex=phonoJoints[self.NSF.joints
                    .value])
                )
        else:
            handconfig[finger] = hc.thumb(
                MCP=hc.joint(dfFlex=phonoJoints[self.NSF.joints
                    .value]),
                IP=hc.joint(dfFlex=phonoJoints[self.NSF.joints.
                    value]),
                CM=hc.joint(
                        dfFlex=phonoAbduction[finger][NSFAbd]["
                            unopposed"][0],
                        dfAbd=phonoAbduction[finger][NSFAbd]["
                            unopposed"][2],
                        dfRot=phonoAbduction[finger][NSFAbd]["
                            unopposed"][1])
                )
    # Check!
    return  hc.handconfiguration(handconfig["index"], handconfig["
        middle"], handconfig["ring"], handconfig["pinky"],
        handconfig["thumb"] )

def __repr__(self):
    return "%s(selectedFingers=%r, secondarySelectedFingers=%r,
        thumb=%r, nonSelectedFingers=%r)" % (self.__class__.__name__
        , self.SF, self.SSF, self.thumb, self.NSF)

def __str__(self):
```

```python
        return """Handshape:
Selected Fingers: %s
Secondary Selected Fingers: %s
Thumb: %s
Non Selected Fingers: %s
""" % (self.SF, self.SSF, self.thumb, self.NSF)


class selectedFingers:
    """The selected fingers"""
    def __init__(self, members, MCP, PIP, abd):
        # check the members
        try:
            members = fingerCheck(members)
        except digitError:
            print("Selected finger digit error.")
            raise
        self.members = members
        # ensure the that MCP is a joint instance
        if isinstance(MCP, joint):
            self.MCP = MCP
        else:
            self.MCP = joint(MCP)
        # ensure the that PIP is a joint instance
        if isinstance(PIP, joint):
            self.PIP = PIP
        else:
            self.PIP = joint(PIP)
```

```
        # duplicate the PIP configuration to the DIP, this should be
            refined
        self.DIP = self.PIP
        if isinstance(abd, abduction):
            self.abd = abd
        else:
            self.abd = abduction(abd)

    def __repr__(self):
        return "%s(members=%r, MCP=%r, PIP=%r, abd=%r)" % (self.
            __class__.__name__, self.members, self.MCP, self.PIP, self.
            abd)

    def __str__(self):
        return """
  members: %s
  MCP: %s
  PIP: %s
  abd: %s""" % (self.members, self.MCP, self.PIP, self.abd)

class secondarySelectedFingers:
    """The secondary selected fingers"""
    def __init__(self, members=None, MCP=None, PIP=None, abd=None):
        # check the members
        try:
            members = fingerCheck(members)
        except digitError:
            print("Selected finger digit error.")
```

```
        raise
    self.members = members
    # ensure the that MCP is a joint instance
    if isinstance(MCP, joint):
        self.MCP = MCP
    else:
        self.MCP = joint(MCP)
    # ensure the that PIP is a joint instance
    if isinstance(PIP, joint):
        self.PIP = PIP
    else:
        self.PIP = joint(PIP)
    # duplicate the PIP configuration, this should be refined
    self.DIP = self.PIP
    if isinstance(abd, abduction):
        self.abd = abd
    else:
        self.abd = abduction(abd)
    # if members is empty, set all to None:
    if len(members) == 0:
        self.MCP = None
        self.PIP = None
        self.abd = None

def __repr__(self):
    return "%s(members=%r, MCP=%r, PIP=%r, abd=%r)" % (self.
       __class__.__name__, self.members, self.MCP, self.PIP, self.
       abd)
```

```
    def __str__(self):
        return """
  members: %s
  MCP: %s
  PIP: %s
  abd: %s
""" % (self.members, self.MCP, self.PIP, self.abd)

class thumb:
    """the thumb"""
    def __init__(self, oppos=None):
        if isinstance(oppos, opposition):
            self.oppos = oppos
        else:
            self.oppos = opposition(oppos)

    def __repr__(self):
        return "%s(oppos=%r)" % (self.__class__.__name__, self.oppos)

    def __str__(self):
        return """
  Opposition: %s
""" % (self.oppos)

class nonSelectedFingers:
    """the non selected fingers"""
    def __init__(self, joints=None, members = set()):
```

```
        try:
            members = fingerCheck(members)
        except digitError:
            print("Nonselected finger digit error.")
            raise
        self.members = members
        # ensure the that joints is a joint instance
        if isinstance(joints, joint):
            self.joints = joints
        else:
            self.joints = joint(joints)

    def __repr__(self):
        return "%s(joints=%r, members=%r)" % (self.__class__.__name__,
           self.joints, self.members)

    def __str__(self):
        return """
  members: %s
  joints: %s
""" % (self.members, self.joints)

##### abstract articulator classes #####
class joint:
    """a joint object"""
    def __init__(self, value):
        try:
            value = jointCheck(value, joints = phonoJoints)
```

```
        except jointError:
            print("The joint is not in the set of phonologically
               specified joint features.")
            raise
        self.value = value

    def __repr__(self):
        return "%s(value=%r)" % (self.__class__.__name__, self.value)

    def __str__(self):
        return "%s" % (self.value)

class opposition:
    """an oppotision object"""
    def __init__(self, value):
        try:
            value = oppositionCheck(value, oppositions =
               phonoOpposition)
        except oppositionError:
            print("The opposition is not in the set of phonologically
               specified opposition features.")
            raise
        self.value = value

    def __repr__(self):
        return "%s(value=%r)" % (self.__class__.__name__, self.value)

    def __str__(self):
```

```
        return "%s" % (self.value)


class abduction:
    """a abduction object"""
    def __init__(self, value):
        try:
            value = abdCheck(value, abds = phonoAbduction["index"]) #
                the index is hard coded here for the check to work, this
                 is a little weird and should be abstracted.
        except abductionError:
            print("The abduction is not in the set of phonologically
                specified abduction features.")
            raise
        self.value = value

    def __repr__(self):
        return "%s(value=%r)" % (self.__class__.__name__, self.value)

    def __str__(self):
        return "%s" % (self.value)



##### testing #####
foo = handshape(
    selectedFingers = selectedFingers(members = ["index", "middle"],
       MCP=joint("ext"), PIP="ext", abd=abduction("adducted")),
    secondarySelectedFingers = None,
    thumb = thumb(oppos=None),
```

```
    nonSelectedFingers = nonSelectedFingers(joints="flex")
    )
bar = foo.toHandconfigTarget()
baz = arm(handshape=foo, orientation="defaultFS")
qux = baz.toArmTarget()
```

A.3 pm.py

```
import hs
import funcs
import csv
from os import path

class notationError(Exception):
    pass

fingerCodingKeyFile =path.join(funcs.resources_dir,'fingerCodingKey.csv
   ')
fingerCodingKey = funcs.read_csv_data(fingerCodingKeyFile)
fingerCodingCols = funcs.dictToCols(fingerCodingKey)
bsfingerCodingCols = funcs.dictColMapper(fingerCodingKey, "base symbol"
   )

jointCodingKeyFile =path.join(funcs.resources_dir,'jointCodingKey.csv')
jointCodingKey = funcs.read_csv_data(jointCodingKeyFile)
jointCodingCols = funcs.dictToCols(jointCodingKey)
psfjointCodingCols = funcs.dictColMapper(jointCodingKey, "psf")
ssfjointCodingCols = funcs.dictColMapper(jointCodingKey, "ssf")
nsfjointCodingCols = funcs.dictColMapper(jointCodingKey, "nsf")
```

```
abdCodingKeyFile = path.join(funcs.resources_dir,'abdCodingKey.csv')
abdCodingKey = funcs.read_csv_data(abdCodingKeyFile)
abdCodingCols = funcs.dictToCols(abdCodingKey)
psfabdCodingCols = funcs.dictColMapper(abdCodingKey, "psf")
ssfabdCodingCols = funcs.dictColMapper(abdCodingKey, "psf") #ssf is the
    same as the psf for abduction.

def shortToMember(string):
    map = {'I': 'index',
           'M': 'middle',
           'R': 'ring',
           'P': 'pinky',
           'T': 'thumb'
           }
    out = [map[x] for x in list(string)]
    return out

##### prosodic model notation class #####
class selectedFingers:
    """a class for selected fingers based on the PM notation system in
       Eccarius and Brentari 2008 of the type 1T-^@;1T-@;#"""
    def __init__(self, string):
        stringList = list(string)

        # Selected finger symbols
        # fingers
        symbolUp = stringList.pop(0).upper()
```

```
if symbolUp not in set(fingerCodingCols["base symbol"]):
    raise notationError("Unknown base symbol in selected
       fingers")
else:
    if symbolUp != "T":
        self.fing = symbolUp
        try:
            symbolUp = stringList.pop(0)
        except IndexError:
            symbolUp = None
    else:
        self.fing = None
# thumb
if symbolUp: symbolUp = symbolUp.upper()
if symbolUp != "T":
    self.thumb = None
else:
    self.thumb = symbolUp
    try:
        symbolUp = stringList.pop(0)
    except IndexError:
        symbolUp = None
# opposition
if symbolUp != "-":
    self.oppos = None
else:
    self.oppos = symbolUp
    try:
```

```
        symbolUp = stringList.pop(0)
    except IndexError:
        symbolUp = None
# abduction
if symbolUp: symbolUp = symbolUp.lower()
if symbolUp not in set(abdCodingCols["psf"]):
    self.abd = None
else:
    self.abd = symbolUp
    try:
        symbolUp = stringList.pop(0)
    except IndexError:
        symbolUp = None
# joint
if symbolUp: symbolUp = symbolUp.lower()
if symbolUp not in set(jointCodingCols["psf"]):
    if symbolUp == None:
        self.joint = None
    else:
        raise notationError("Unknown joint symbol in selected
           fingers")
else:
    self.joint = symbolUp
# test to ensure there's no string left.
if len(stringList) > 0:
    raise notationError("There's still unparsed string left in
       the selected finger substring.")
```

```
class secondarySelectedFingers:
    """a class for secondary selected fingers based on the PM notation
       system in Eccarius and Brentari 2008 of the type 1T-^@;1T-@;#"""
    def __init__(self, string):
        stringList = list(string)
        # Secondary selected finger symbols
        # fingers
        symbolUp = stringList.pop(0).upper()
        if symbolUp not in set(fingerCodingCols["base symbol"]):
            raise notationError("Unknown base symbol in selected
               fingers")
        else:
            if symbolUp != "T":
                self.fing = symbolUp
                try:
                    symbolUp = stringList.pop(0)
                except IndexError:
                    symbolUp = None
            else:
                self.fing = None
        # thumb
        if symbolUp: symbolUp = symbolUp.upper()
        if symbolUp != "T":
            self.thumb = None
        else:
            self.thumb = symbolUp
            try:
                symbolUp = stringList.pop(0)
```

```
    except IndexError:
        symbolUp = None
# opposition
if symbolUp != "-":
    self.oppos = None
else:
    self.oppos = symbolUp
    try:
        symbolUp = stringList.pop(0)
    except IndexError:
        symbolUp = None
# abduction doesn't exist in PM notation for secondary selected
    fingers, but should be and is accounted for here.
if symbolUp: symbolUp = symbolUp.lower()
if symbolUp not in set(abdCodingCols["psf"]):
    self.abd = None
else:
    self.abd = symbolUp
    try:
        symbolUp = stringList.pop(0)
    except IndexError:
        symbolUp = None
# joint
if symbolUp: symbolUp = symbolUp.lower()
if symbolUp not in set(jointCodingCols["psf"]):
    if symbolUp == None:
        self.joint = None
    else:
```

```
                raise notationError("Unknown joint symbol in secondary
                    selected fingers")
        else:
            self.joint = symbolUp
        # test to ensure there's no string left.
        if len(stringList) > 0:
            raise notationError("There's still unparsed string left in
                the secondary selected finger substring.")


class nonSelectedFingers:
    """a class for non selected fingers based on the PM notation system
        in Eccarius and Brentari 2008 of the type 1T-^@;1T-@;#"""
    def __init__(self, string):
        stringList = list(string)
        # joint
        symbolUp = stringList.pop(0)
        if symbolUp not in set(jointCodingCols["nsf"]):
            if symbolUp is None:
                self.joint = None
            else:
                raise notationError("Unknown joint symbol in
                    nonselected fingers")
        else:
            self.joint = symbolUp
        # test to ensure there's no string left.
        if len(stringList) > 0:
            raise notationError("There's still unparsed string left in
                the nonselected finger substring.")
```

```
class pmHandshape:
    """a class based on the PM notation system in Eccarius and Brentari
        2008 of the type 1T-^@;1T-@;#"""
    def __init__(self, string):
        strings = string.split(";")
        self.SF = selectedFingers(strings.pop(0))
        try:
            stringUp = strings.pop(0)
            if stringUp in set(jointCodingCols["nsf"]):
                self.SSF = None
                self.NSF = nonSelectedFingers(stringUp)
            else:
                self.SSF = secondarySelectedFingers(stringUp)
                try:
                    stringUp = strings.pop(0)
                    self.NSF = nonSelectedFingers(stringUp)
                except IndexError:
                    self.NSF = None
        except IndexError:
            self.SSF = None
            self.NSF = None
        if len(strings) > 0:
            raise notationError("There's still unparsed string left: "+
               str(strings))

    def toAMhandshape(self):
        # set default value for the thumb: opposed
```

```
oppos = "opposed"
# translate the selected fingers
if self.SF.fing:
    sfMem = shortToMember(bsfingerCodingCols[self.SF.fing]['
       fingers'])
if self.SF.thumb and self.SF.thumb == "T" :
    try:
        sfMem.append("thumb")
    except UnboundLocalError:
        sfMem = ["thumb"]
if self.SF.oppos and self.SF.oppos == "-":
    oppos = "unopposed"
else:
    oppos = "opposed"
if self.SF.abd:
    sfAbd = psfabdCodingCols[self.SF.abd]['abd']
else:
    sfAbd = None
if self.SF.joint:
    sfMCP = psfjointCodingCols[self.SF.joint]['MCP']
    sfPIP = psfjointCodingCols[self.SF.joint]['PIP']
else:
    sfMCP = psfjointCodingCols['empty']['MCP']
    sfPIP = psfjointCodingCols['empty']['PIP']
sf =  hs.selectedFingers(members = sfMem, MCP=sfMCP, PIP=sfPIP,
    abd=sfAbd)
# translate the secondary selected fingers
if self.SSF:
```

```
        if self.SSF.fing:
            ssfMem = shortToMember(bsfingerCodingCols[self.SSF.fing
                ]['fingers'])
        if self.SSF.thumb and self.SSF.thumb == "T" :
            try:
                ssfMem.append("thumb")
            except UnboundLocalError:
                ssfMem = ["thumb"]
        if self.SSF.oppos and self.SSF.oppos == "-":
            oppos = "unopposed"
        else:
            oppos = "opposed"
        if self.SSF.abd:
            ssfAbd = ssfabdCodingCols[self.SSF.abd]['abd']
        else:
            ssfAbd = None
        if self.SSF.joint:
            ssfMCP = ssfjointCodingCols[self.SSF.joint]['MCP']
            ssfPIP = ssfjointCodingCols[self.SSF.joint]['PIP']
        else:
            ssfMCP = ssfjointCodingCols['empty']['MCP']
            ssfPIP = ssfjointCodingCols['empty']['PIP']
        ssf = hs.secondarySelectedFingers(members = ssfMem, MCP=
            ssfMCP, PIP=ssfPIP, abd=ssfAbd)
    else:
        ssf = None
    # translate the nonselected fingers
    if self.NSF:
```

```
            if self.NSF.joint:
                ssfJoints = nsfjointCodingCols[self.NSF.joint]['MCP']
            else:
                ssfJoints = None
            nsf = hs.nonSelectedFingers(joints=ssfJoints)
        else:
            nsf = None
        thumb = hs.thumb(oppos=oppos)
        AMhandshape = hs.handshape(selectedFingers = sf,
           secondarySelectedFingers = ssf, thumb = thumb,
           nonSelectedFingers = nsf )
        return AMhandshape

##### test #####
foo = pmHandshape("1;#")
bar = foo.toAMhandshape()
baz = bar.toHandconfigTarget()

foo1 = pmHandshape("DT@;/")
bar1 = foo1.toAMhandshape()

baz1 = bar1.toHandconfigTarget()
```

A.4 letters.py

```
import hs
import pm
import funcs
```

```
import csv
from os import path

##### Error classes #####
class specificationError(Exception):
    pass

##### Read in csvs with letter specifications #####
lettersFile = path.join(funcs.resources_dir,'lettersFromArtModel.csv')
lettersKey = funcs.read_csv_data(lettersFile)
lettersCols = funcs.dictToCols(lettersKey)
letterCodingCols = funcs.dictColMapper(lettersKey, "letter")

def letterToArm(letter):
    """converts a letter to an articulatory model representation of
       handshape"""
    try:
        let = letterCodingCols[letter]
    except KeyError:
        print("That is not a recognized letter")
        raise
    psf = hs.selectedFingers(
        members = let["psf-members"].split(","),
        MCP=let["psf-mcp"],
        PIP=let["psf-pip"],
        abd=hs.abduction(let["psf-abd"])
        )
    if(let["ssf-members"] == "None"):
```

```
        ssf = None
    else:
        ssf = hs.secondarySelectedFingers(
            members = let["ssf-members"].split(","),
            MCP=let["ssf-mcp"],
            PIP=let["ssf-pip"],
            abd=hs.abduction(let["ssf-abd"])
            )
    if(let["thumb-oppos"] == "None"):
        thmb = None
    else:
        thmb = hs.thumb(oppos=let["thumb-oppos"])
    if(let["nsf-joints"] == "None"):
        nsf = None
    else:
        nsf = hs.nonSelectedFingers(joints=let["nsf-joints"])
    handshape = hs.handshape(
        selectedFingers = psf,
        secondarySelectedFingers = ssf,
        thumb = thmb,
        nonSelectedFingers = nsf
        )
    orientation = let["orientation"]
    return hs.arm(handshape=handshape, orientation=orientation)

def printAllLetters():
    for letter in lettersKey:
         print("###################")
```

```python
        print(letter["letter"])
        print(letterToArm(letter["letter"]).toArmTarget())


def ntuples(lst, n):
    return zip(*[lst[i:]+lst[:i-1] for i in range(n)])


def measureContour(string, method="unweighted"):
    stringTup = tuple(string)
    cost = []
    for pair in ntuples(stringTup,2):
        c =  letterToArm(pair[0]).toArmTarget()-letterToArm(pair[1]).
           toArmTarget()
        if method == "unweighted":
             c = c.totalDegreesDifferent()
        elif method == "weighted":
             c = c.weightedDegreesDifferent()
        else:
             raise specificationError("No recognized method for
                measuring contour.")
        cost.append(c)
    return sum(cost)


def similarity(stringA, stringB, method="unweighted"):
    if len(stringA) != len(stringB):
        raise specificationError("The strings are not of the same
           length, cannot compare without some sort of editing")
    cost = []
    for pair in zip(stringA,stringB):
```

```
        c = letterToArm(pair[0]).toArmTarget()-letterToArm(pair[1]).
            toArmTarget()
        if method == "unweighted":
            c = c.totalDegreesDifferent()
        elif method == "weighted":
            c = c.weightedDegreesDifferent()
        else:
            raise specificationError("No recognized method for
                measuring contour.")
        cost.append(c)
    return sum(cost)

def letterToPM(letter):
    """converts a letter to a prosodic model code"""
    try:
        let = letterCodingCols[letter]
    except KeyError:
        print("That is not a recognized letter")
        raise
    return pm.pmHandshape(let["pmCode"])

##### Tests ######
#ensure that all pm codes are readable
for ltr in lettersCols['letter']:
    try:
        letterToPM(ltr).toAMhandshape()
    except:
```

```
        print("Error with "+ltr+". can't convert from PM notation to AM
            handshape")

#ensure that all articulatory model specifications are readable
for ltr in lettersCols['letter']:
    try:
        letterToArm(ltr)
    except:
        print("error with "+ltr+". can't convert from articulatory
           specifications to AM handshape")

#ensure that all articulatory model specifications are readable
for ltr in lettersCols['letter']:
    try:
        AMarm = letterToArm(ltr)
    except:
        print("error with "+ltr+". can't convert from articulatory
           specifications to AM handshape")
        break
    try:
        PMarm = hs.arm(handshape=letterToPM(ltr).toAMhandshape(),
           orientation=letterCodingCols[ltr]["orientation"])
    except:
        print("Error with "+ltr+". can't convert from PM notation to AM
            handshape")
        break

    AMPMdiff = AMarm.toArmTarget()-PMarm.toArmTarget()
```

```
    if AMPMdiff.totalDegreesDifferent() > 0:
        print("The difference between the PM and AM for "+ltr+" is "+
           str(AMPMdiff.totalDegreesDifferent())+" degrees.")
        print("Articulatory model:")
        print(AMarm.toArmTarget())
        print("Prosodic model:")
        print(PMarm.toArmTarget())
```

A.5 render.py

```
import hc
import funcs
import string # for testing
import letters # for testing

import yaml, csv, math, subprocess
from os import path, makedirs

##### Error classes #####
class specificationError(Exception):
    pass

##### Path to deafult in the base pose to alter #####
baseHCposeFile = path.join(funcs.resources_dir,"
   fsBaseOpticalClosedToOpen.yml")

##### Establish joint angles for the base hand #####
index = hc.finger(MCP=(180,5), PIP=180, DIP=180)
middle = hc.finger(MCP=(180,2), PIP=180, DIP=180)
```

```
ring = hc.finger(MCP=(180,-2), PIP=180, DIP=180)
pinky = hc.finger(MCP=(180,-4), PIP=180, DIP=180)
thmb = hc.thumb(CM=hc.joint(dfFlex=15, dfAbd=9, dfRot=27, dfPro=None),
   MCP=180, IP=180)
wrist = (0,0,0) # the wrist values here are not those in the pose file,
    these need to be changed in the future

baseHC = hc.armconfiguration(hc.handconfiguration(index, middle, ring,
   pinky, thmb), wrist)

def ntz(value):
    """Change a none to zero"""
    if value == None:
        value = 0
    return value

def renderImage(hc, imageOutFile, baseHCposeFile=baseHCposeFile, baseHC
   =baseHC):
    ##### Read in the base pose to alter #####
    baseHCposefile = open(baseHCposeFile, "r")
    baseHCpose = yaml.load(baseHCposefile)
    baseHCposefile.close()

    newHCpose = baseHCpose
    diff = baseHC - hc

    fingMap = {"finger4": 'index',
              "finger3": 'middle',
```

```
        "finger2": 'ring',
        "finger1": 'pinky',
        "finger5": 'thumb'}

fingerJointMap = {"joint1": 'MCP',
            "joint2": 'PIP',
            "joint3": 'DIP'}

thumbJointMap = {"joint1": 'CM',
            "joint2": 'MCP',
            "joint3": 'IP'}

for fingerJoint in baseHCpose['hand_joints']:
    finger = fingerJoint[0:7]
    joint = fingerJoint[7:13]
    if finger[0:-1] != "finger" or joint[0:-1] != "joint":
        continue

    if fingMap[finger] == "thumb":
        jointMove = getattr(getattr(diff.hand, fingMap[finger]),
           thumbJointMap[joint])
        if joint == "joint1":
            # these joint mappings are wrong wrong wrong.
            jointMatrix = [(ntz(jointMove.dfFlex)*math.pi)/180, (
               ntz(jointMove.dfAbd)*math.pi)/180 , (ntz(jointMove.
               dfRot)*math.pi)/180 ]
        elif joint == "joint2" or joint == "joint3":
```

```
            jointMatrix = [(ntz(jointMove.dfFlex)*math.pi)/180, (
               ntz(jointMove.dfAbd)*math.pi)/180 , (ntz(jointMove.
               dfRot)*math.pi)/180 ]
        newJoints = [i - j for i, j in zip(baseHCpose['hand_joints'
           ][fingerJoint], jointMatrix)]
        newHCpose['hand_joints'][fingerJoint] = newJoints
    else:
        jointMove = getattr(getattr(diff.hand, fingMap[finger]),
           fingerJointMap[joint])
        jointMatrix = [(ntz(jointMove.dfFlex)*math.pi)/180, (ntz(
           jointMove.dfAbd)*math.pi)/180 , (ntz(jointMove.dfRot)*
           math.pi)/180 ]
        newJoints = [i - j for i, j in zip(baseHCpose['hand_joints'
           ][fingerJoint], jointMatrix)]
        newHCpose['hand_joints'][fingerJoint] = newJoints

wristMatrix = [(ntz(diff.wrist.dfFlex)*math.pi)/180, (ntz(0)*math.
   pi)/180 , (ntz(diff.wrist.dfPro)*math.pi)/180 ]
newHCpose['hand_joints']['metacarpals'] = [i - j for i, j in zip(
   baseHCpose['hand_joints']['metacarpals'], wristMatrix)]


rootMatrix = [(ntz(0)*math.pi)/180, ((ntz(diff.wrist.dfPro)*math.pi
   )/180)*-(3/4), ((ntz(diff.wrist.dfPro)*math.pi)/180)*(5/4) ]
newHCpose['hand_joints']['carpals'] = [i - j for i, j in zip(
   baseHCpose['hand_joints']['carpals'], rootMatrix)]
```

```
    # make tmp directory if it doesn't exist
    if not path.exists(path.join(funcs.resources_dir,"tmp")):
        makedirs(path.join(funcs.resources_dir,"tmp"))

    poseOutFilePath = path.join(funcs.resources_dir,''.join(["tmp/",
       path.basename(imageOutFile),"poseOut.yml"]))
    print(poseOutFilePath)
    poseOutFile = open(poseOutFilePath, 'w')
    poseOutFile.write("%YAML:1.0\n")
    yaml.dump(newHCpose, poseOutFile)
    poseOutFile.close()



    cmd = [path.join(funcs.resources_dir,"imageGen"), path.join(funcs.
       resources_dir,"hand_model/scene_spec.yml"), poseOutFilePath,
       imageOutFile]
    devnull = open('/dev/null', 'w')
    subprocess.call(cmd, stdout=devnull, stderr=subprocess.STDOUT)

##### Tests ######
#ensure that all articulatory model specifications are readable
if not path.exists("./let"):
   makedirs("./let")
for ltr in letters.lettersCols['letter']:
    try:
        AMarm = letters.letterToArm(ltr)
    except:
```

```
    print("error with "+ltr+". can't convert from articulatory
        specifications to AM handshape")
    break
pth = path.join("./let/",''.join(["am-",ltr,".png"]))
print(pth)
renderImage(AMarm.toArmTarget(), pth)

# try:
#     PMarm = hs.arm(handshape=letters.letterToPM(ltr).
    toAMhandshape(), orientation=letters.letterCodingCols[ltr]["
    orientation"])
# except:
#     print("Error with "+ltr+". can't convert from PM notation to
    AM handshape")
#     break
# renderImage(PMarm.toArmTarget(), path.join("./let/",''.join(["pm
    -",ltr,".png"])))
```

Appendix B

Word lists

B.1 First

B.1.1 Names

1. aberdeen
2. afghanistan
3. africa
4. alan
5. alcapulco
6. alexander
7. amy
8. angelica
9. ann
10. apraxia
11. atlantic
12. bea
13. beijing
14. bill
15. botswana
16. cameroon
17. camilla
18. caribbean
19. carl
20. chris
21. cleveland
22. columbus
23. danny
24. debbie
25. don
26. el salvador
27. enrique
28. everglades
29. excel
30. exxon
31. felix
32. finn
33. flossmoor
34. francesca
35. franklin
36. fred
37. gary
38. gayle
39. george
40. giordano
41. greg
42. himalaya
43. inglewood
44. izzy
45. jacqueline
46. jason
47. jimmy
48. joe
49. john
50. josh
51. kate
52. kelly
53. leo
54. lexus
55. libya
56. mary
57. matt
58. mauritania
59. mediterranean
60. mexico
61. mia
62. mississippi
63. mongolia
64. moscow
65. naomi
66. naperville
67. nic
68. oak park
69. owen
70. pam
71. paraguay
72. quentin
73. quincy
74. quotation
75. rangerover
76. rita
77. russ
78. sam
79. san francisco
80. sara
81. scotland
82. skokie
83. tallahassee
84. tanzania

85. tiffany
86. tobias
87. toby
88. tokyo
89. tom
90. venezuela
91. venice
92. viv
93. will
94. william
95. xavier
96. xerox
97. yellowstone
98. yosemite
99. zack
100. zoe

B.1.2 Nouns

1. appetizers
2. aquarium
3. asphyxiation
4. ataxia
5. axel
6. axis
7. basil
8. bass
9. beef
10. boo
11. box
12. cabin
13. cadillac
14. campfire
15. carp
16. claw
17. cliff
18. cliffhanger
19. deck
20. dinosaur
21. dogfight
22. earthquake
23. equal
24. executive
25. expectation
26. expert
27. expo
28. family
29. fanbelt
30. fanny
31. fern
32. findings
33. fir
34. firewire
35. flea
36. flour
37. furniture
38. glue
39. grape
40. gravity
41. headlight
42. herb
43. ink
44. instrument
45. jade
46. jawbreaker
47. jewelry
48. juice
49. lamb
50. life
51. liquid
52. luggage
53. material
54. mitten
55. mustang
56. neighborhood
57. notebook
58. oval
59. oxen
60. oxygen
61. pony
62. quantity
63. quarry
64. quarter
65. queen
66. question
67. quicksand
68. quilt
69. quiz
70. rest
71. riddle
72. sauce
73. seed
74. sequel
75. silk
76. softserve
77. spice
78. spruce
79. square
80. squirrel
81. staff
82. stool
83. strawberry
84. sun
85. taxi
86. tulip
87. turquoise

88. twizzlers
89. vacuum
90. van
91. waffle
92. weed
93. windshield
94. wing
95. xenon
96. xenophobia
97. xmen
98. xylophone
99. yard
100. zebra

B.1.3 Non-English

1. ahoj
2. anteeksi
3. axon
4. belyeg
5. blahopreji
6. chwilke
7. cie
8. csokifagyit
9. czesc
10. daj
11. dekuji
12. dlaczego
13. dnia
14. egyenesen
15. elnezest
16. feleseg
17. felkelni
18. ferfi
19. fiu
20. hei
21. hlad
22. hogyan
23. hol
24. huomenta
25. huone
26. hyvaa
27. igek
28. igen
29. informacja
30. itt
31. jegy
32. jsou
33. juna
34. kahdeksan
35. kaksi
36. kde
37. kerul
38. kolik
39. korhaz
40. koszi
41. kto
42. kuusi
43. lentokentta
44. maanantai
45. mennyibe
46. miluji
47. mina
48. missa
49. moc
50. navstivil
51. nelja
52. neni
53. nerozumim
54. nigdy
55. nogi
56. nowych
57. ole
58. onko
59. opravdu
60. paljonko
61. palyadvar
62. penzvaltas
63. piec
64. pocalujmy
65. pojd
66. pospeste
67. potrebuji
68. powazaniem
69. powaznie
70. prekladatel
71. procvicovat
72. przepraszam
73. puhu
74. rado
75. rakastan
76. rano
77. rendorseg
78. sina
79. siusiu
80. spotykac
81. surgos
82. szia
83. tancolni
84. toistekan
85. tuhat
86. usta
87. utca
88. vcera
89. viisi
90. vitej

91. voitte
92. wlosy
93. yksi
94. zdrowie
95. zgoda
96. zgubilam
97. zizen
98. zobaczenia
99. zopakovat
100. zyc

B.2 Second

B.2.1 English

1. abdul
2. abhorrence
3. abner
4. accent
5. acme
6. adjunct
7. admin
8. adze
9. akbar
10. albuquerque
11. algae
12. alphabet
13. alumnus
14. analyze
15. anecdote
16. asthma
17. astigmatism
18. attn
19. awkward
20. axle
21. babka
22. banjo
23. banquet
24. barware
25. bathhouse
26. beachcomber
27. bedbug
28. bedfellow
29. beehive
30. bellhop
31. BFF
32. biceps
33. birthstone
34. bizs
35. BMW
36. boxwood
37. boy
38. breakfast
39. bulwark
40. bumkin
41. busywork
42. buzzword
43. calf
44. calque
45. calzone
46. camcorder
47. campground
48. camry
49. captain
50. cartwheel
51. chevre
52. chickenpox
53. chongqing
54. churchgoer
55. clockwork
56. clxvii
57. cobweb
58. cognate
59. comfort
60. compaqs
61. cowgirl
62. crabgrass
63. crewcut
64. croquet
65. cuff
66. cuzco
67. cysts
68. damsel
69. deafness
70. debt
71. dijkstra
72. disgust
73. dishpan
74. disjoint
75. dogcatcher
76. dogvane
77. dowry
78. duvet
79. dwarf
80. envy
81. epcot
82. esterhzys
83. exhibit
84. exquisite
85. faithful
86. farce
87. FCC
88. FDA
89. filmgoer
90. fjord
91. flapjack
92. floyd
93. fuzz
94. gallbladder
95. gauze
96. gingko
97. gizmo
98. gretzky
99. gruyere
100. gumdrop

101. gunpowder
102. guvnors
103. hajj
104. halfwit
105. heavyweight
106. hemline
107. highjack
108. hindquarter
109. hipbone
110. humvee
111. hutzpa
112. hybrid
113. hymn
114. hypo
115. inkjet
116. inlay
117. interlude
118. iqbal
119. iraqi
120. jazz
121. jellyfish
122. jihad
123. kafka
124. kaufman
125. key
126. keyhole
127. khayyam
128. kiwi
129. kmart
130. knight
131. kumquat
132. kvetch
133. latvian
134. laxness
135. leipzig
136. liverwurst
137. lobster
138. logjam
139. luxury
140. lyric
141. maelstrom
142. mamzer
143. manhole
144. mcbride
145. mcfadden
146. mcguire
147. mcpherson
148. mezzo
149. misdemeanor
150. misfit
151. mozzarella
152. munchkin
153. nashville
154. novgorod
155. nyquil
156. oatmeal
157. object
158. offprint
159. okra
160. outdoorsman
161. outfit
162. ovum
163. oxbow
164. perjury
165. pewter
166. picnic
167. pigpen
168. plaza
169. pneumonia
170. ponytail
171. potpie
172. powwow
173. poxvirus
174. presbyterian
175. prescription
176. pressroom
177. pretzel
178. pumpkinseed
179. puppet
180. pyjama
181. qatar
182. qwerty
183. ramjet
184. redcat
185. redhead
186. remnant
187. ribcage
188. ringworm
189. runway
190. SFX
191. shipwreck
192. shotgun
193. sixfold
194. skivvy
195. sled
196. snobbery
197. snowflake
198. snowman
199. snowplow
200. stockcar
201. straightjacket
202. submarine
203. subpoena
204. subversion

205. subzero
206. svelt
207. symphony
208. syndrome
209. taffy
210. tbsp
211. thruway
212. trekker
213. updo
214. uzbekistan
215. velazquez
216. vodka
217. windpipe
218. xrefs
219. yevtushenko
220. ymha
221. zhivago
222. zigzag
223. zrich

B.2.2 Non-English

1. afzwoer
2. ajpartilkom
3. aqqet
4. avbildes
5. avfdte
6. awqa
7. batnihx
8. bouwvak
9. chaqchukamuy
10. chawrasqpua
11. chukqayara
12. czuwaszja
13. ejhv
14. ewx
15. fajx
16. fexkelt
17. fraqtha
18. glukzov
19. grzzaby
20. gxm
21. jbikkix
22. joqros
23. jozfkov
24. jqaqana
25. jserripx
26. jwaluhx
27. kajfov
28. kejjlilkom
29. kuxovwe
30. laqlqu
31. lindqvists
32. lvque
33. magneetijzer
34. majnatli
35. meqjusin
36. mixjin
37. mmccx
38. mouvmes
39. najcwaszy
40. ntefqitilniex
41. olfoqni
42. omxotna
43. podzznam
44. pque
45. psowjer
46. qaqywa
47. qaxqxithomx
48. qhas
49. rnexxejtx
50. scheepvaart
51. schrijvend
52. seddaqkom
53. skejt
54. stevje
55. stofvrij
56. svkv
57. sxeo
58. szz
59. tajglov
60. taqfilkomx
61. taqqulu
62. tarbxet
63. tifqgek
64. titniffidx
65. tlaqqmilha
66. tmaqdarniex
67. tobqu
68. tqaulna
69. ujyariy
70. vpitei
71. vxling
72. wzz
73. xdm
74. xgajra
75. zappaptlekx
76. zavhat
77. zevzecu

B.3 CELEX

1. account
2. act
3. action
4. activity
5. age
6. air
7. amount
8. animal
9. answer
10. area
11. argument
12. arm
13. art
14. attention
15. attitude
16. authority
17. baby
18. back
19. bank
20. bed
21. benefit
22. bit
23. blood
24. body
25. book
26. boy
27. Britain
28. brother
29. building
30. business
31. car
32. case
33. cent
34. centre
35. century
36. chair
37. chance
38. change
39. chapter
40. child
41. church
42. city
43. class
44. clothes
45. club
46. committee
47. community
48. company
49. control
50. cost
51. country
52. couple
53. course
54. court
55. daughter
56. day
57. days
58. deal
59. death
60. decision
61. degree
62. department
63. development
64. difference
65. doctor
66. door
67. doubt
68. earth
69. education
70. effect
71. effort
72. end
73. energy
74. Europe
75. evening
76. event
77. evidence
78. example
79. experience
80. eye
81. face
82. fact
83. family
84. father
85. fear
86. feeling
87. few
88. field
89. figure
90. finger
91. fire
92. fish
93. floor
94. food
95. foot
96. force
97. form
98. friend
99. front
100. future
101. game

102. garden
103. girl
104. glass
105. God
106. government
107. ground
108. group
109. hair
110. hall
111. hand
112. head
113. health
114. heart
115. help
116. history
117. home
118. horse
119. hospital
120. hotel
121. hour
122. house
123. husband
124. idea
125. income
126. industry
127. information
128. interest
129. issue
130. job
131. John
132. kind
133. knowledge
134. Labour
135. land
136. language
137. law
138. leader
139. least
140. leg
141. letter
142. level
143. life
144. light
145. line
146. little
147. London
148. look
149. Lord
150. lot
151. love
152. machine
153. man
154. market
155. material
156. matter
157. meeting
158. member
159. method
160. mile
161. mind
162. minister
163. minute
164. Miss
165. moment
166. money
167. month
168. morning
169. mother
170. mouth
171. movement
172. Mrs
173. music
174. name
175. nation
176. nature
177. need
178. newspaper
179. night
180. number
181. office
182. officer
183. oil
184. once
185. one
186. order
187. organization
188. others
189. paper
190. parent
191. part
192. party
193. people
194. period
195. person
196. picture
197. piece
198. place
199. plan
200. plant
201. play
202. point
203. police
204. policy
205. position

206. pound
207. power
208. president
209. pressure
210. price
211. problem
212. process
213. production
214. programme
215. purpose
216. quality
217. question
218. rate
219. reason
220. relationship
221. report
222. research
223. rest
224. result
225. river
226. road
227. role
228. room
229. rule
230. school
231. sea
232. security
233. sense
234. service
235. sex
236. shop
237. shoulder
238. side
239. sign
240. situation
241. size
242. society
243. son
244. sort
245. sound
246. source
247. South
248. space
249. staff
250. stage
251. state
252. States
253. story
254. street
255. student
256. study
257. subject
258. summer
259. sun
260. support
261. system
262. table
263. tax
264. teacher
265. terms
266. theory
267. thing
268. things
269. thought
270. time
271. top
272. town
273. trade
274. tree
275. trouble
276. truth
277. type
278. union
279. university
280. use
281. value
282. view
283. village
284. voice
285. wall
286. war
287. water
288. way
289. week
290. West
291. while
292. wife
293. will
294. window
295. woman
296. word
297. work
298. worker
299. world
300. year

B.4 Motion capture word list

1. george
2. theory
3. people
4. executive
5. venice
6. vacuum
7. mitten
8. furniture
9. century
10. glass
11. jason
12. husband
13. number
14. trouble
15. quicksand
16. earthquake
17. knowledge
18. jewelry
19. size
20. father
21. effort
22. department
23. night
24. finger
25. building
26. situation
27. plethora
28. morphology
29. hypothalmus
30. psychiatry
31. psychology
32. philosophy
33. thanksgiving
34. chugging
35. higgens
36. higgenbothum
37. theory
38. thought
39. garden
40. trouble
41. mother
42. evening
43. rate
44. front
45. work
46. leader
47. age
48. land
49. foot
50. love
51. head
52. husband
53. others
54. parent
55. government
56. area
57. community
58. building
59. Europe
60. field
61. baby
62. plant
63. effort
64. language
65. John
66. officer
67. terms
68. school
69. bit
70. period
71. stage
72. Mrs
73. Miss
74. man
75. level
76. look
77. doctor
78. couple
79. brother
80. member
81. road
82. once
83. house
84. world
85. face
86. clothes
87. class
88. bed
89. policy
90. back
91. action
92. account
93. student
94. development
95. industry
96. newspaper
97. pressure
98. time
99. office
100. interest
101. horse

102. war
103. source
104. society
105. hair
106. security
107. fish
108. course
109. country
110. support
111. letter
112. research
113. Lord
114. story
115. door
116. boy
117. water
118. price
119. way
120. car
121. attention
122. answer
123. leg
124. end
125. food
126. woman
127. eye
128. problem
129. hotel
130. committee
131. service
132. girl
133. mile
134. sign
135. body
136. law
137. state
138. play
139. purpose
140. table
141. village
142. part
143. rule
144. point
145. university
146. knowledge
147. air
148. situation
149. value
150. act
151. kind
152. blood
153. amount
154. people
155. idea
156. river
157. attitude
158. deal
159. figure
160. one
161. pound
162. home
163. book
164. matter
165. floor
166. power
167. information
168. subject
169. least
170. South
171. movement
172. help
173. name
174. space
175. Britain
176. chance
177. rest
178. quality
179. chapter
180. future
181. game
182. report
183. president
184. shoulder
185. cent
186. minute

Appendix C
Hold duration model comparisons

The large number of single frame holds, violates the linearity assumption of the hierarchical linear model. In order to ensure that this is not driving the results we are seeing, we fit additional models on subsets of the data:

These are visualized using the coefficient plot in C.1, where each model is a different color. Full model outputs are listed in table C.1. As the coefficient plot and table shows, none of the coefficient estimates of the models deviate wildly, meaning that we can have confidence in the full model.

As a reminder, the precitors (and their abbreviations are):

- rate (rateScaled)
- word type (wordtype)
- repetition (repetition)
- current apogee orientation or movement phonological group (currGroup)
- previous apogee's phonological orientation or movement group (prevGroup)
- following apogee's phonological orientation or movement group (follGroup)
- position in the word (position)
- *interaction* rate × word type
- *interaction* word type × repetition
- *interaction* interaction of rate × word type × repetition

And the grouping factors are:

- intercept adjustments for signer (1, 2, 3, or 4), as well as slope adjustments for
  - rate
  - word type,
  - repetition,
- intercept adjustments for word length
- intercept adjustments for current apogee FS-letter
- intercept adjustments for previous apogee FS-letter
- intercept adjustments for following apogee FS-letter
- intercept adjustments for trial
- intercept adjustments for words, which are nested within wordlists

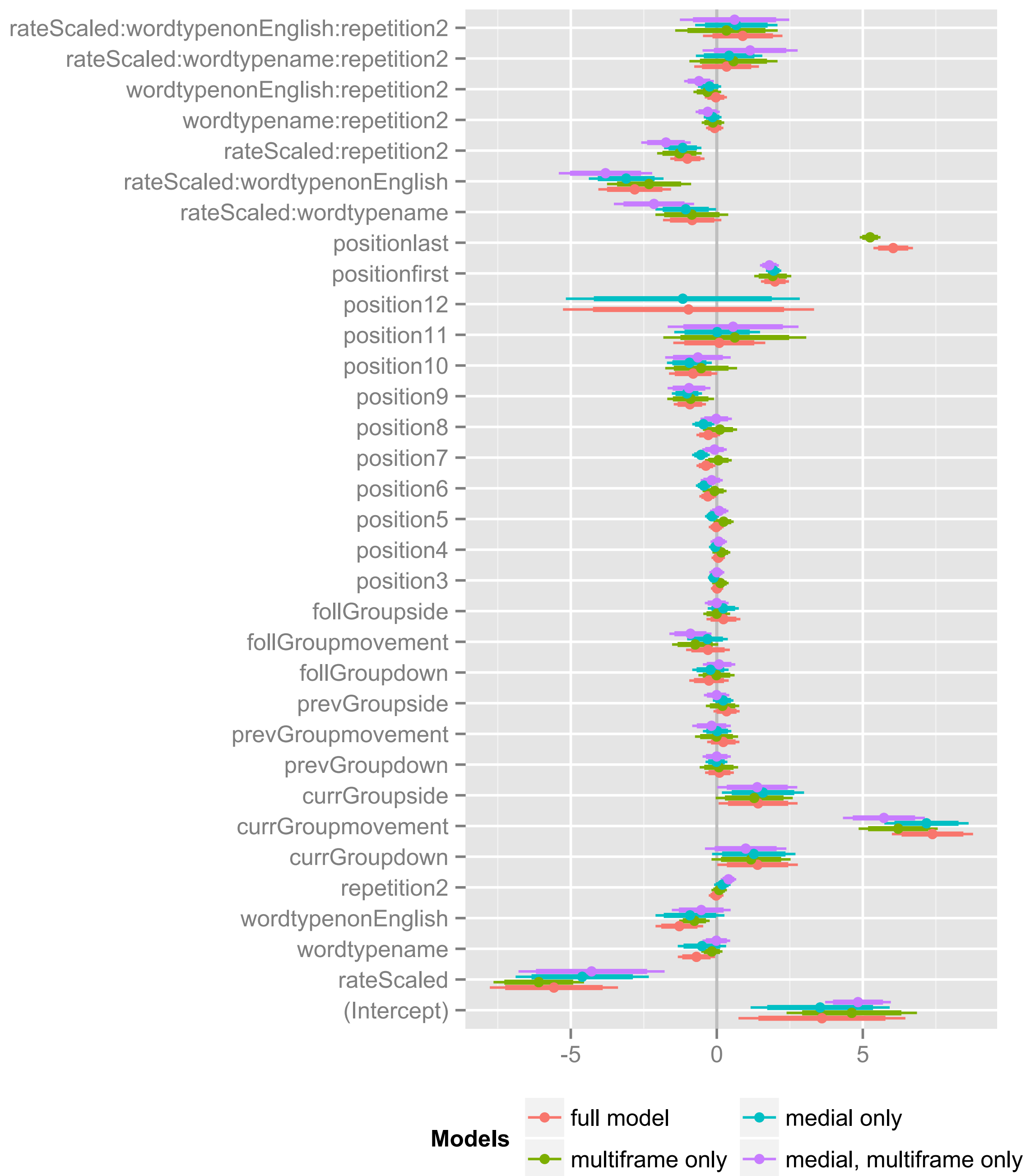


Figure C.1: **Coefficient plot for the predictors of the full hierarchical linear model including all holds, as well as the reduced models (abbreviated: full model): the same model with only multiframe holds (abbreviated: multiframe only), the same model with only word-medial holds (abbreviated: medial only), and the same model with only multiframe word-medial holds (abbreviated: medial, multiframe only)** Thick lines represent 95% confidence, thin lines 99% confidence, and dots are the estimates of the coefficients (or intercept).

	full model	multiframe only	medial only	medial, multiframe only
(Intercept)	$3.60(1.11)^{**}$	$4.62(0.87)^{***}$	$3.54(0.92)^{***}$	$4.83(0.44)^{***}$
rateScaled	$-5.58(0.85)^{***}$	$-6.10(0.60)^{***}$	$-4.61(0.89)^{***}$	$-4.29(0.97)^{***}$
wordtypename	$-0.70(0.25)^{**}$	$-0.17(0.14)$	$-0.51(0.32)$	$-0.02(0.18)$
wordtypenonEnglish	$-1.28(0.32)^{***}$	$-0.77(0.20)^{***}$	$-0.92(0.46)^{*}$	$-0.53(0.39)$
repetition2	$-0.02(0.10)$	$0.08(0.10)$	$0.19(0.11)$	$0.40(0.10)^{***}$
currGroupdown	$1.40(0.54)^{**}$	$1.17(0.53)^{*}$	$1.27(0.55)^{*}$	$0.99(0.54)$
currGroupmovement	$7.39(0.54)^{***}$	$6.21(0.53)^{***}$	$7.18(0.56)^{***}$	$5.72(0.54)^{***}$
currGroupside	$1.41(0.52)^{**}$	$1.28(0.51)^{*}$	$1.58(0.55)^{**}$	$1.38(0.53)^{**}$
prevGroupdown	$0.09(0.19)$	$0.07(0.26)$	$-0.01(0.15)$	$0.00(0.19)$
prevGroupmovement	$0.22(0.21)$	$-0.01(0.29)$	$0.01(0.19)$	$-0.18(0.26)$
prevGroupside	$0.33(0.17)$	$0.20(0.22)$	$0.22(0.14)$	$-0.01(0.17)$
follGroupdown	$-0.27(0.26)$	$-0.01(0.24)$	$-0.22(0.24)$	$0.08(0.22)$
follGroupmovement	$-0.30(0.29)$	$-0.74(0.31)^{*}$	$-0.32(0.27)$	$-0.90(0.28)^{**}$
follGroupside	$0.23(0.23)$	$0.00(0.18)$	$0.22(0.21)$	$0.00(0.16)$
position3	$0.01(0.08)$	$0.12(0.11)$	$-0.11(0.08)$	$0.00(0.10)$
position4	$0.05(0.09)$	$0.16(0.12)$	$-0.05(0.08)$	$0.07(0.11)$
position5	$-0.02(0.10)$	$0.23(0.13)$	$-0.16(0.09)$	$0.08(0.12)$
position6	$-0.31(0.11)^{**}$	$-0.08(0.16)$	$-0.44(0.11)^{***}$	$-0.17(0.15)$
position7	$-0.37(0.13)^{**}$	$0.05(0.18)$	$-0.54(0.12)^{***}$	$-0.08(0.16)$
position8	$-0.29(0.16)$	$0.11(0.23)$	$-0.46(0.15)^{**}$	$-0.02(0.21)$
position9	$-0.92(0.21)^{***}$	$-0.90(0.31)^{**}$	$-1.02(0.20)^{***}$	$-0.95(0.29)^{***}$
position10	$-0.81(0.32)^{*}$	$-0.54(0.48)$	$-0.94(0.30)^{**}$	$-0.65(0.44)$
position11	$0.08(0.61)$	$0.62(0.95)$	$0.01(0.57)$	$0.56(0.87)$
position12	$-0.97(1.67)$		$-1.16(1.55)$	
positionfirst	$1.99(0.19)^{***}$	$1.92(0.25)^{***}$	$1.95(0.10)^{***}$	$1.80(0.13)^{***}$
positionlast	$6.04(0.26)^{***}$	$5.25(0.14)^{***}$		
rateScaled:wordtypename	$-0.84(0.39)^{*}$	$-0.85(0.48)$	$-1.06(0.40)^{**}$	$-2.15(0.53)^{***}$
rateScaled:wordtypenonEnglish	$-2.81(0.48)^{***}$	$-2.32(0.56)^{***}$	$-3.10(0.50)^{***}$	$-3.81(0.62)^{***}$
rateScaled:repetition2	$-1.01(0.23)^{***}$	$-1.28(0.30)^{***}$	$-1.17(0.25)^{***}$	$-1.74(0.33)^{***}$
wordtypename:repetition2	$-0.07(0.12)$	$-0.13(0.15)$	$-0.14(0.12)$	$-0.31(0.16)$
wordtypenonEnglish:repetition2	$-0.03(0.15)$	$-0.33(0.18)$	$-0.25(0.15)$	$-0.61(0.20)^{**}$
rateScaled:wordtypename:repetition2	$0.34(0.43)$	$0.57(0.59)$	$0.42(0.44)$	$1.14(0.63)$
rateScaled:wordtypenonEnglish:repetition2	$0.89(0.53)$	$0.33(0.68)$	$0.67(0.55)$	$0.61(0.73)$
AIC	84399.75	51992.31	68103.50	37459.67
BIC	84825.40	52384.62	68511.21	37829.22
Log Likelihood	-42144.87	-25942.15	-33997.75	-18676.84
Deviance	84289.75	51884.31	67995.50	37353.67
Num. obs.	16967	10562	14047	7884
Num. groups: wordList:word	577	575	577	571
Num. groups: trialWR	549	548	549	532
Num. groups: follLetter	27	27	26	26
Num. groups: prevLetter	27	27	27	27
Num. groups: apogeeLetter	26	26	26	26
Num. groups: lengthFact	11	11	11	11
Num. groups: signer	4	4	4	4
Variance: wordList:word.(Intercept)	0.82	0.30	0.95	0.33
Variance: trialWR.(Intercept)	0.43	0.02	0.58	0.04
Variance: follLetter.(Intercept)	0.06	0.01	0.05	0.00
Variance: prevLetter.(Intercept)	0.03	0.05	0.00	0.01
Variance: apogeeLetter.(Intercept)	0.48	0.45	0.52	0.48
Variance: lengthFact.(Intercept)	0.03	0.03	0.02	0.02
Variance: signer.(Intercept)	4.75	2.84	3.25	0.60
Variance: signer.rateScaled	2.71	1.09	2.94	3.28
Variance: signer.wordtypename	0.16	0.01	0.32	0.05
Variance: signer.wordtypenonEnglish	0.29	0.06	0.72	0.48
Variance: signer.repetition2	0.02	0.01	0.03	0.01
Variance: Residual	7.76	7.62	6.68	6.31

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.1: Coefficient estimates and standard errors for the full hierarchical linear model including all holds (abbreviated: full model), as well as the reduced models: the same model with only multiframe holds (abbreviated: multiframe only), the same model with only word-medial holds (abbreviated: medial only), and the same model with only multiframe word-medial holds (abbreviated: medial, multiframe only)

Appendix D
Additional visualizations for motion capture data

D.1 Fingerspelling rate, as measured with motion capture data

D.1.1 Rates from the one HMM *for all signers model*

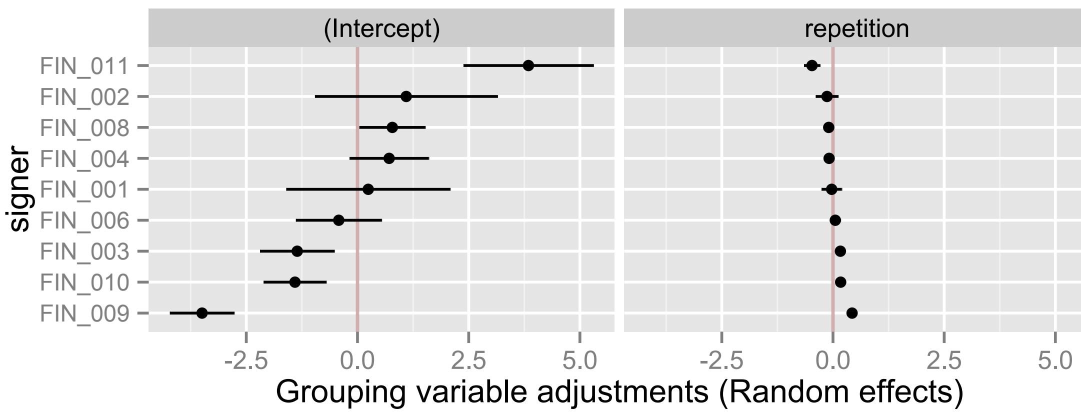


Figure D.1: **Plot of intercept adjustments (random intercepts) for signer, as well as slope adjustments (random slopes) for repetition of the hierarchical linear model for rates, using the all signer HMM model** As discussed in detail above, there is a large amount of intersigner variation (seen in the intercept facet), additionally, there is some variation among signers with respect to the effect of repetition. The levels on the y-axis are signers, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

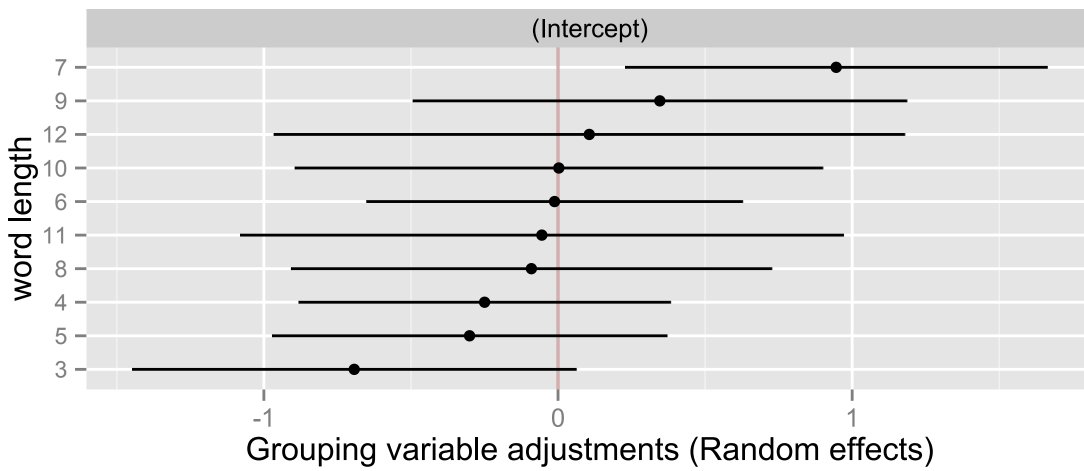


Figure D.2: **Plot of intercept adjustments (random intercepts) for length of the hierarchical linear model for rates, using the all signer HMM model** As discussed in detail above, there is not much systematic variation of rate between word lengths. The levels on the y-axis are the word lengths, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

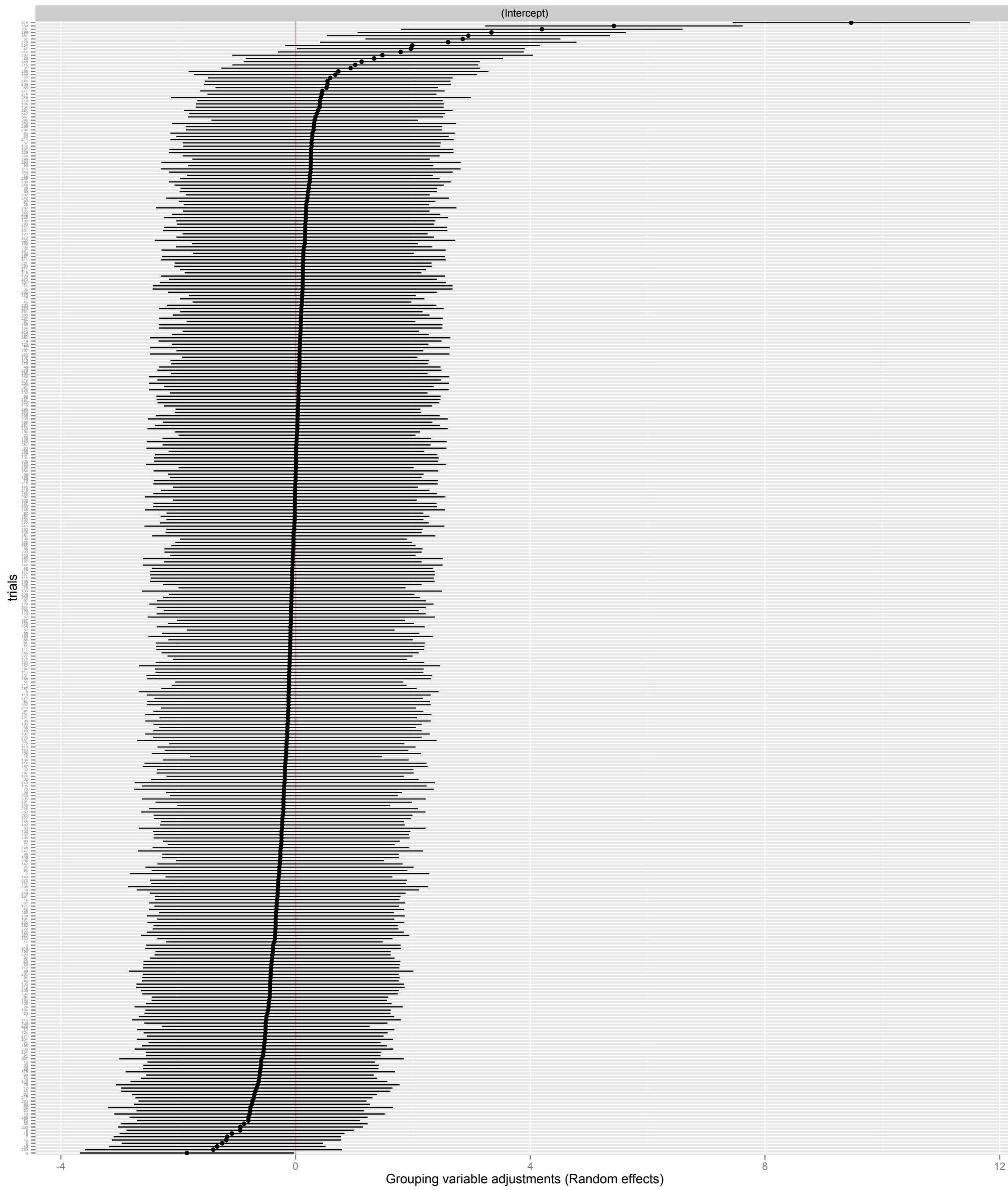


Figure D.3: **Plot of intercept adjustments (random intercepts) for trials of the hierarchical linear model for rates, using the all signer HMM model** Because there are a large number of trials, there are many levels on the y-axis. Although it is difficult to read individual words, as discussed in detail above, there is not much systematic variation of rate between trials. The sigmoidal shape is due to the fact that the intercept adjustments are modeled on a normal distribution. The levels on the y-axis are trial(number)s, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

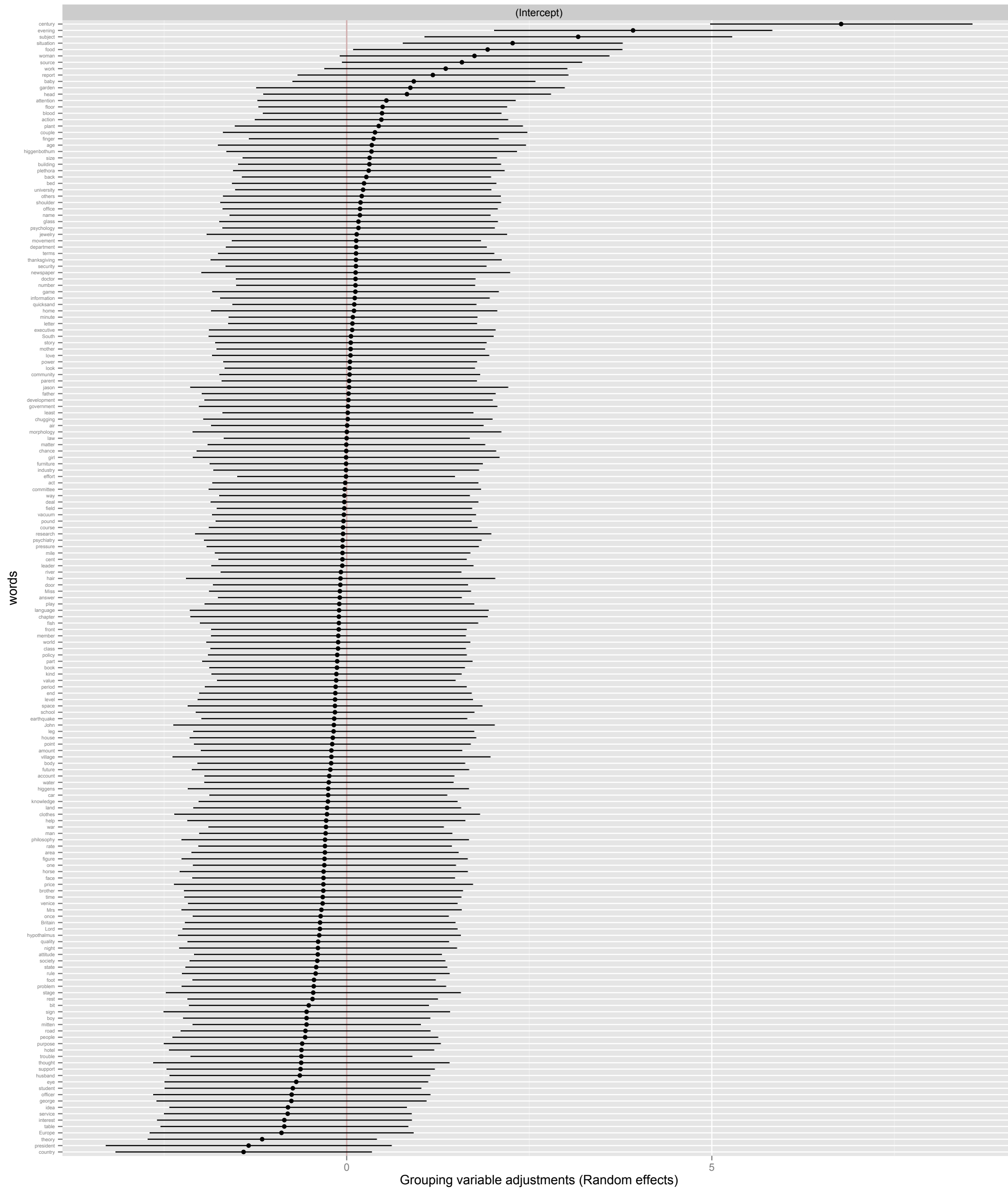


Figure D.4: **Plot of intercept adjustments (random intercepts) for words of the hierarchical linear model for rates, using the all signer HMM model** Because there are a large number of words, there are many levels on the y-axis. Although it is difficult to read individual words, as discussed in detail above, there is not much systematic variation of rate between words. The sigmoidal shape is due to the fact that the intercept adjustments are modeled on a normal distribution. The levels on the y-axis are words (with the word list prefixed to them, to show the nested structure), and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

D.1.2 Rates from the signer-specific HMM model

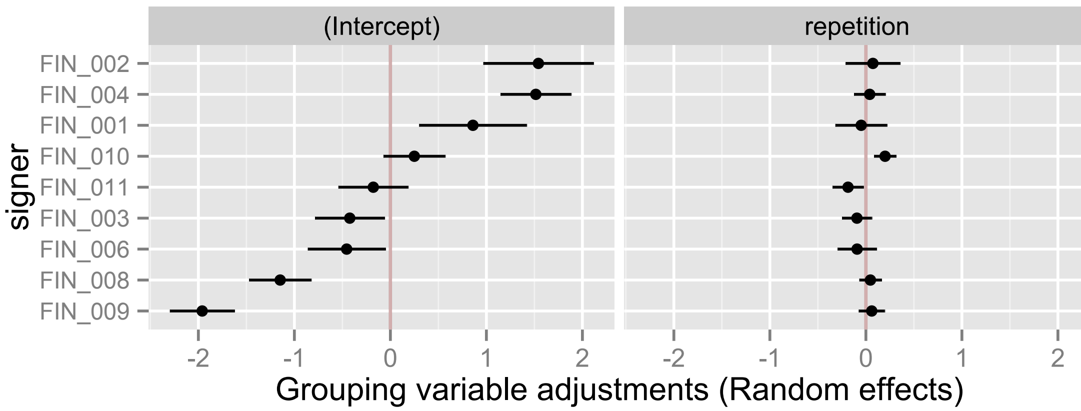


Figure D.5: **Plot of intercept adjustments (random intercepts) for signer, as well as slope adjustments (random slopes) for repetition of the hierarchical linear model for rates, using the signer-specific HMM model** As discussed in detail above, there is a large amount of intersigner variation (seen in the intercept facet), additionally, there is some variation among signers with respect to the effect repetition. The levels on the y-axis are signers, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

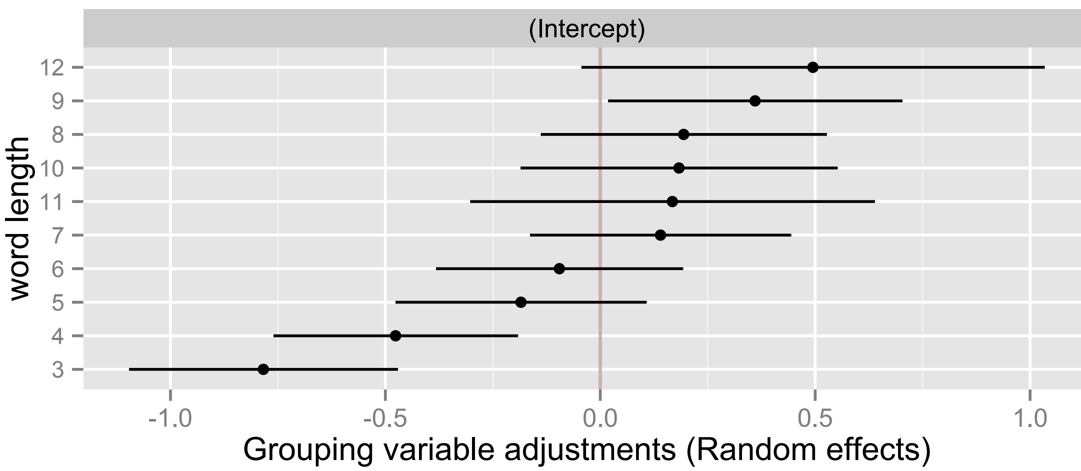


Figure D.6: **Plot of intercept adjustments (random intercepts) for length of the hierarchical linear model for rates, using the signer-specific HMM model** As discussed in detail above, there is not much systematic variation of rate between word lengths. The levels on the y-axis are the word lengths, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

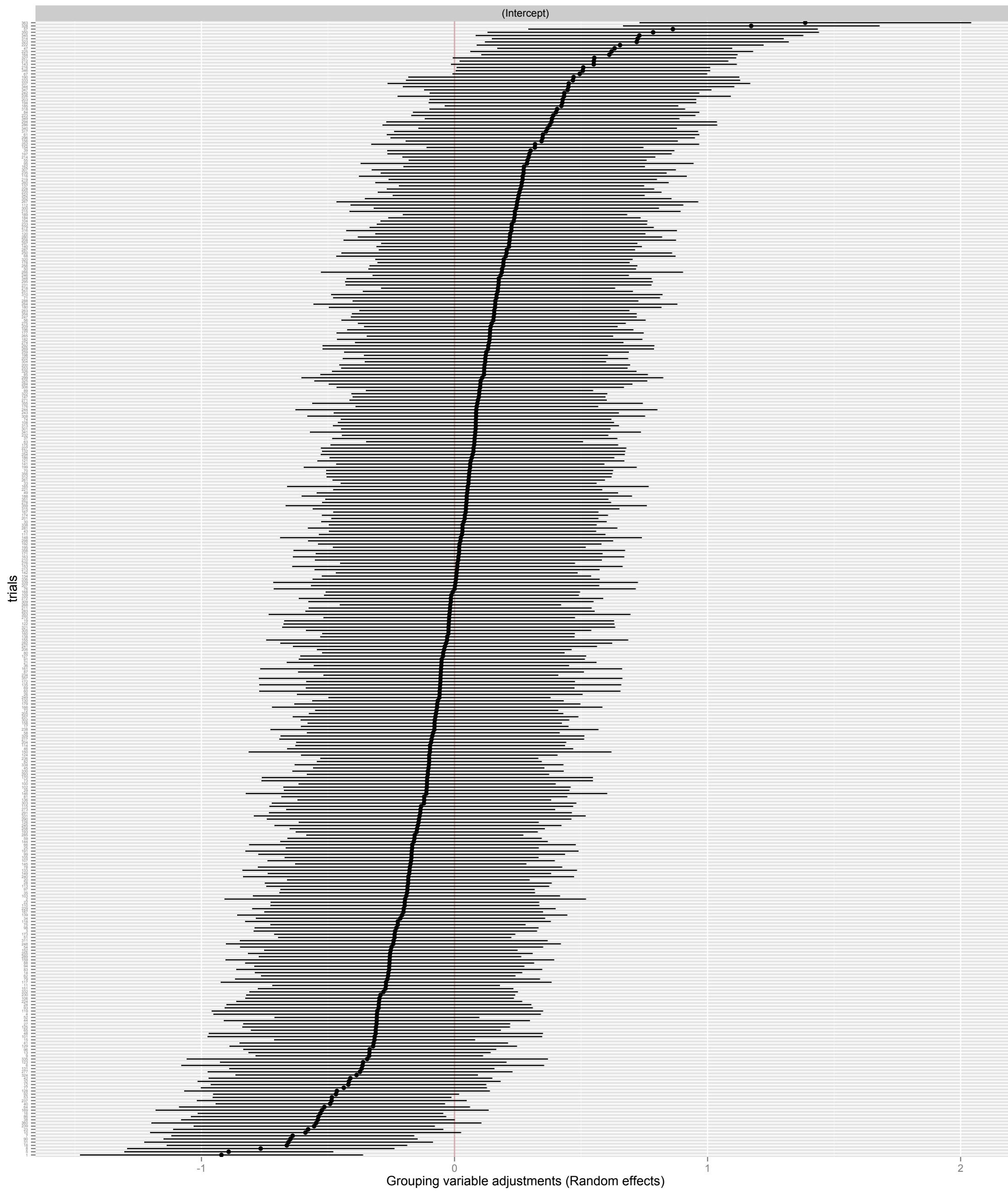


Figure D.7: **Plot of intercept adjustments (random intercepts) for trials of the hierarchical linear model for rates, using the signer-specific HMM model** Because there are a large number of trials, there are many levels on the y-axis. Although it is difficult to read individual words, as discussed in detail above, there is not much systematic variation of rate between trials. The sigmoidal shape is due to the fact that the intercept adjustments are modeled on a normal distribution. The levels on the y-axis are trial(number)s, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

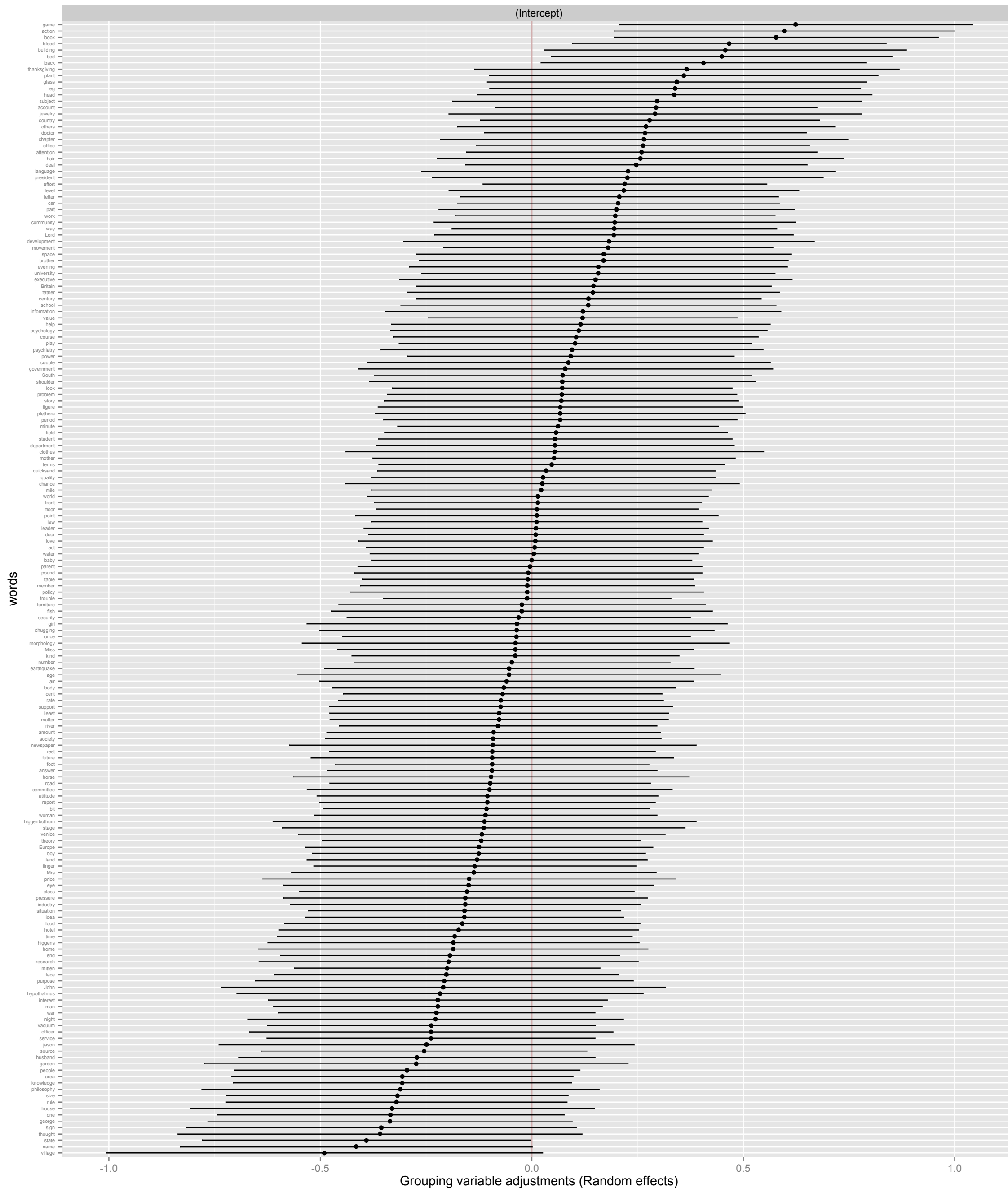


Figure D.8: **Plot of intercept adjustments (random intercepts) for words of the hierarchical linear model for rates, using the signer-specific HMM model** Because there are a large number of words, there are many levels on the y-axis. Although it is difficult to read individual words, as discussed in detail above, there is not much systematic variation of rate between words. The sigmoidal shape is due to the fact that the intercept adjustments are modeled on a normal distribution. The levels on the y-axis are words, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

D.2 Word duration from video

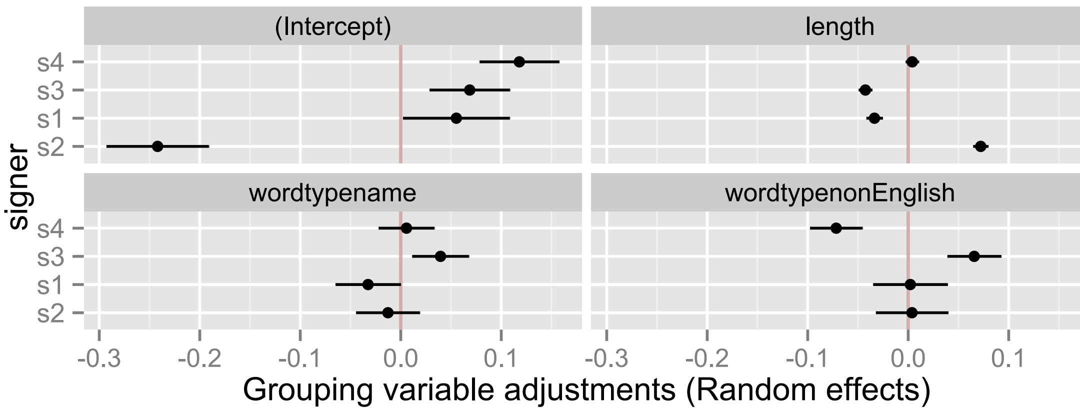


Figure D.9: **Plot of intercept adjustments (random intercepts) for signer, as well as slope adjustments (random slopes) for word type and repetition of the hierarchical linear model for all word durations** As discussed in detail above, there is a large amount of intersigner variation (seen in the intercept facet), additionally, there is some variation among signers with respect to the effects of word type and repetition. The levels on the y-axis are signers, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

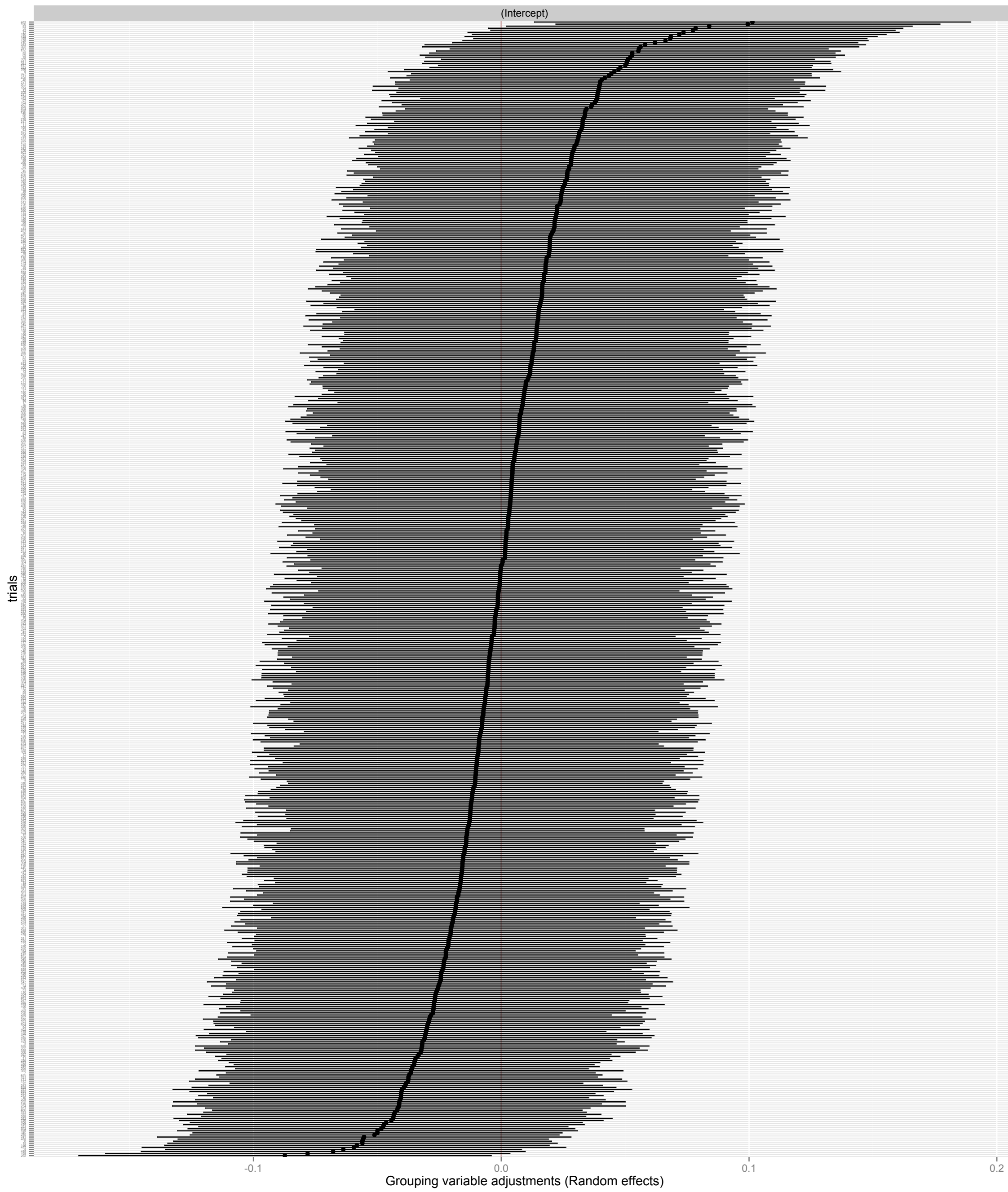


Figure D.10: **Plot of intercept adjustments (random intercepts) for trials of the hierarchical linear model for all word durations** Because there are a large number of trials, there are many levels on the y-axis. Although it is difficult to read individual words, as discussed in detail above, there is not much systematic variation of word duration between trials. The sigmoidal shape is due to the fact that the intercept adjustments are modeled on a normal distribution. The levels on the y-axis are trial(number)s, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

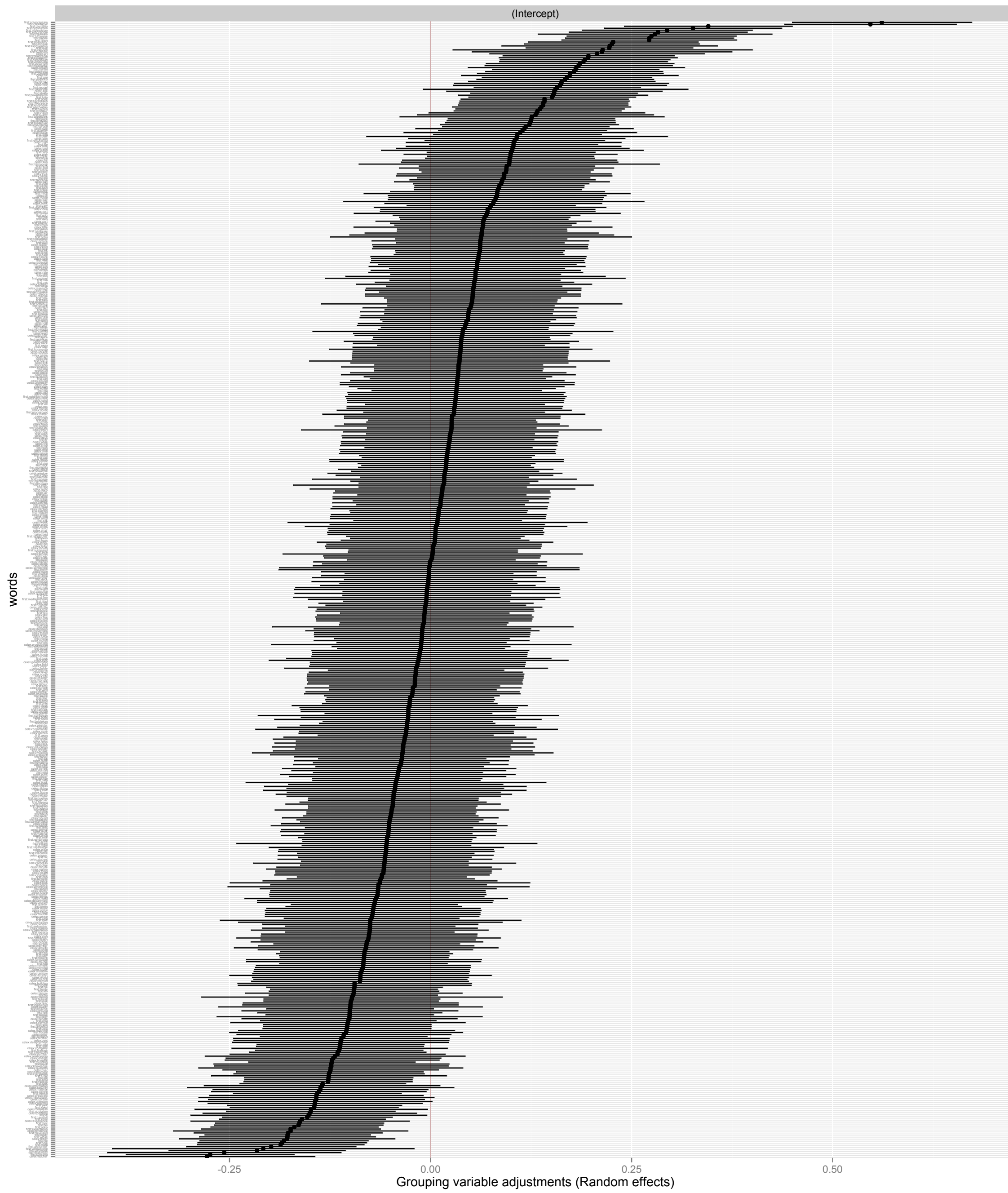


Figure D.11: **Plot of intercept adjustments (random intercepts) for words nested in word lists of the hierarchical linear model for all word durations** Because there are a large number of words, there are many levels on the y-axis. Although it is difficult to read individual words, as discussed in detail above, there is not much systematic variation of word duration between words. The sigmoidal shape is due to the fact that the intercept adjustments are modeled on a normal distribution. The levels on the y-axis are words (with the word list prefixed to them, to show the nested structure), and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

D.3 Word duration from motion capture, all signer HMM

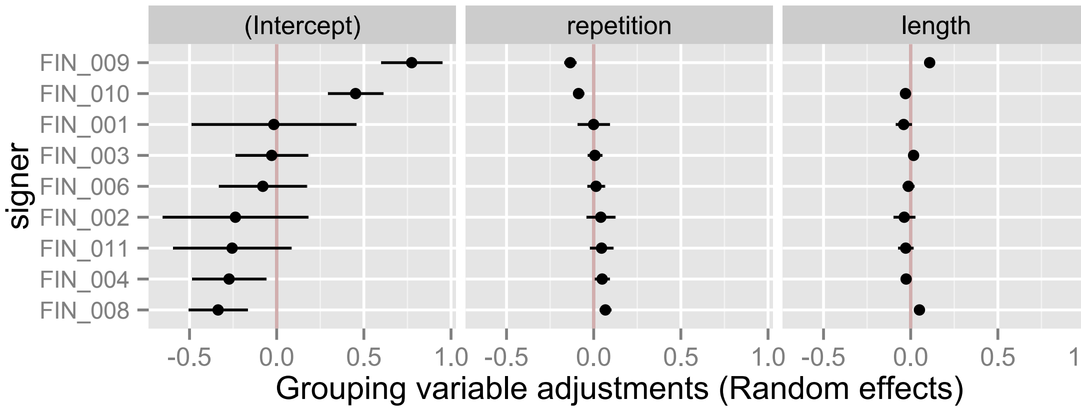


Figure D.12: **Plot of intercept adjustments (random intercepts) for signer, as well as slope adjustments (random slopes) for repetition of the hierarchical linear model for all word durations using motion capture data and the all signer HMM** As discussed in detail above, there is a large amount of intersigner variation (seen in the intercept facet), additionally, there is some variation among signers with respect to the effect of repetition. The levels on the y-axis are signers, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

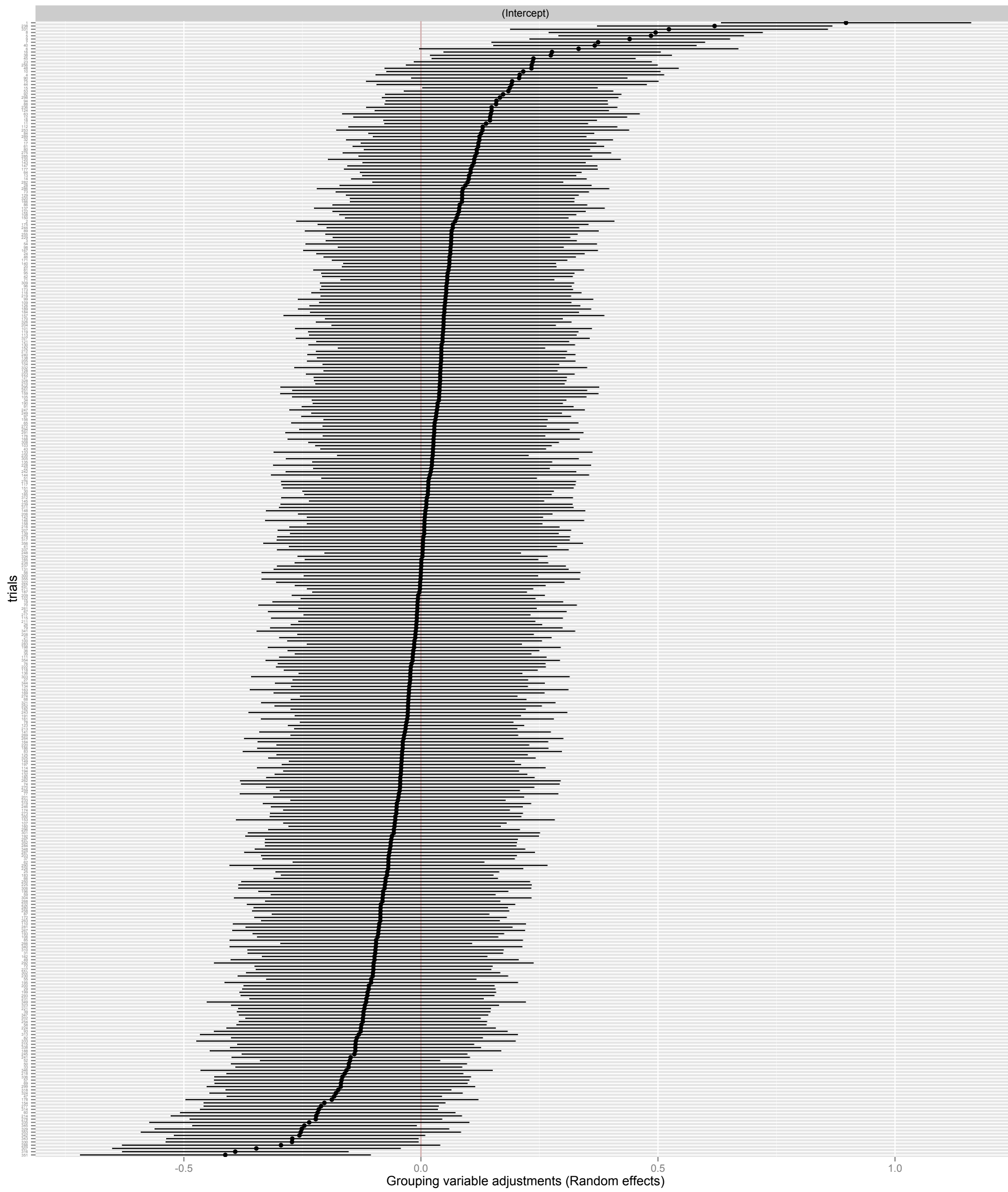


Figure D.13: **Plot of intercept adjustments (random intercepts) for trials of the hierarchical linear model for all word durations using motion capture data and the all signer HMM** Because there are a large number of trials, there are many levels on the y-axis. Although it is difficult to read individual words, as discussed in detail above, there is not much systematic variation of word duration between trials. The sigmoidal shape is due to the fact that the intercept adjustments are modeled on a normal distribution. The levels on the y-axis are trial(number)s, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

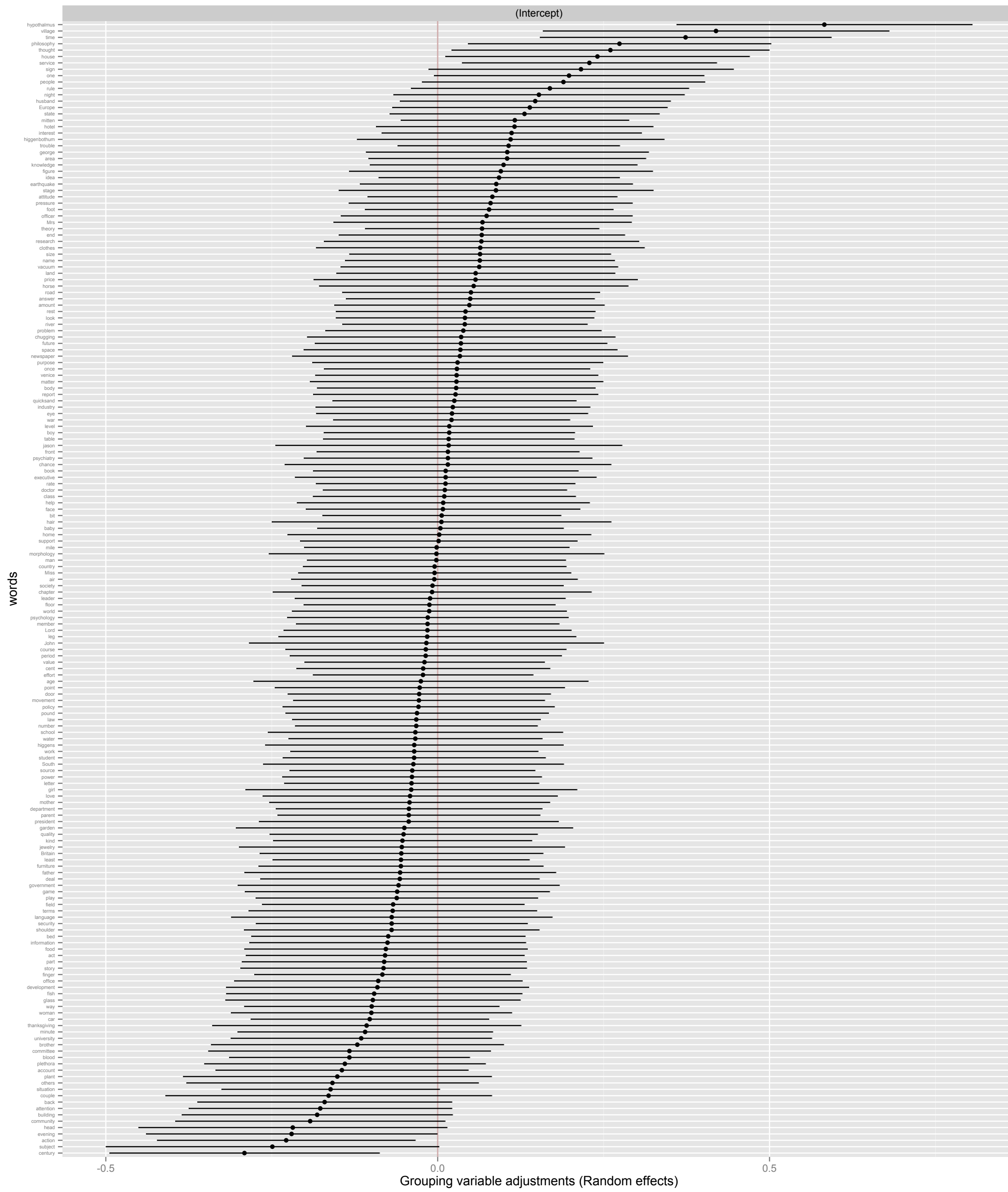


Figure D.14: **Plot of intercept adjustments (random intercepts) for words of the hierarchical linear model for all word durations using motion capture data and the all signer HMM** Because there are a large number of words, there are many levels on the y-axis. Although it is difficult to read individual words, as discussed in detail above, there is not much systematic variation of word duration between words. The sigmoidal shape is due to the fact that the intercept adjustments are modeled on a normal distribution. The levels on the y-axis are words, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

D.4 Word duration from motion capture, signer specific HMM

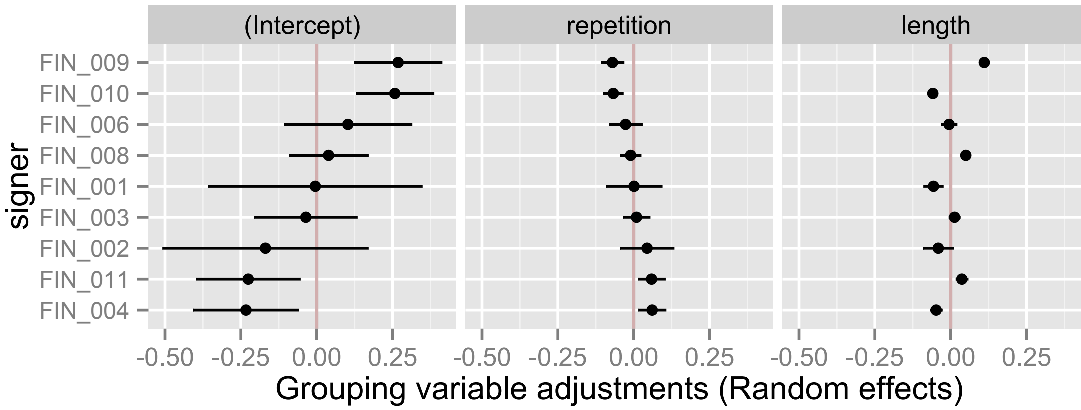


Figure D.15: **Plot of intercept adjustments (random intercepts) for signer, as well as slope adjustments (random slopes) for repetition of the hierarchical linear model for all word durations using motion capture data and the signer-specific HMM** As discussed in detail above, there is a large amount of intersigner variation (seen in the intercept facet), additionally, there is some variation among signers with respect to the effect of repetition. The levels on the y-axis are signers, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

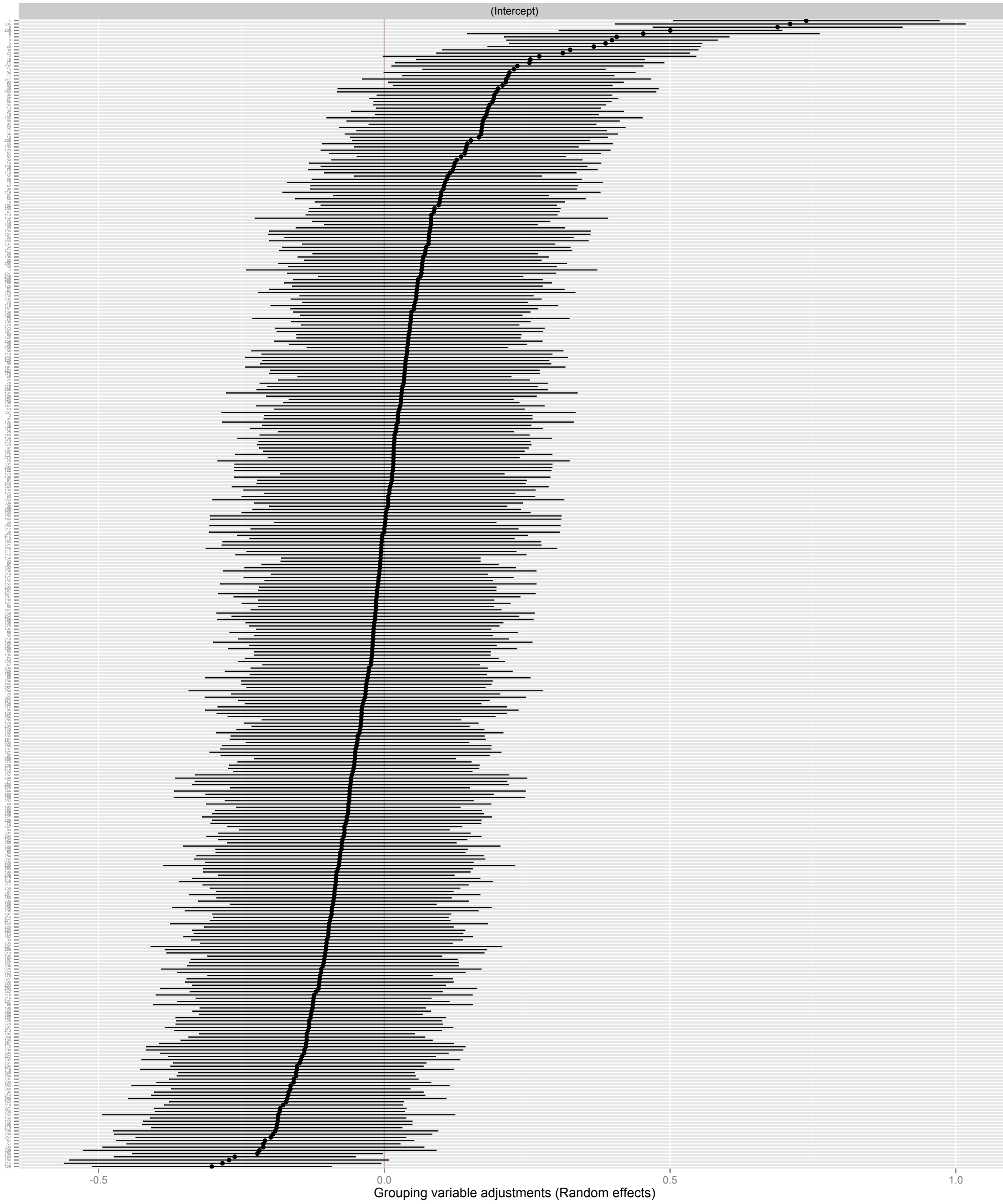


Figure D.16: **Plot of intercept adjustments (random intercepts) for trials of the hierarchical linear model for all word durations using motion capture data and the signer-specific HMM** Because there are a large number of trials, there are many levels on the y-axis. Although it is difficult to read individual words, as discussed in detail above, there is not much systematic variation of word duration between trials. The sigmoidal shape is due to the fact that the intercept adjustments are modeled on a normal distribution. The levels on the y-axis are trial(number)s, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

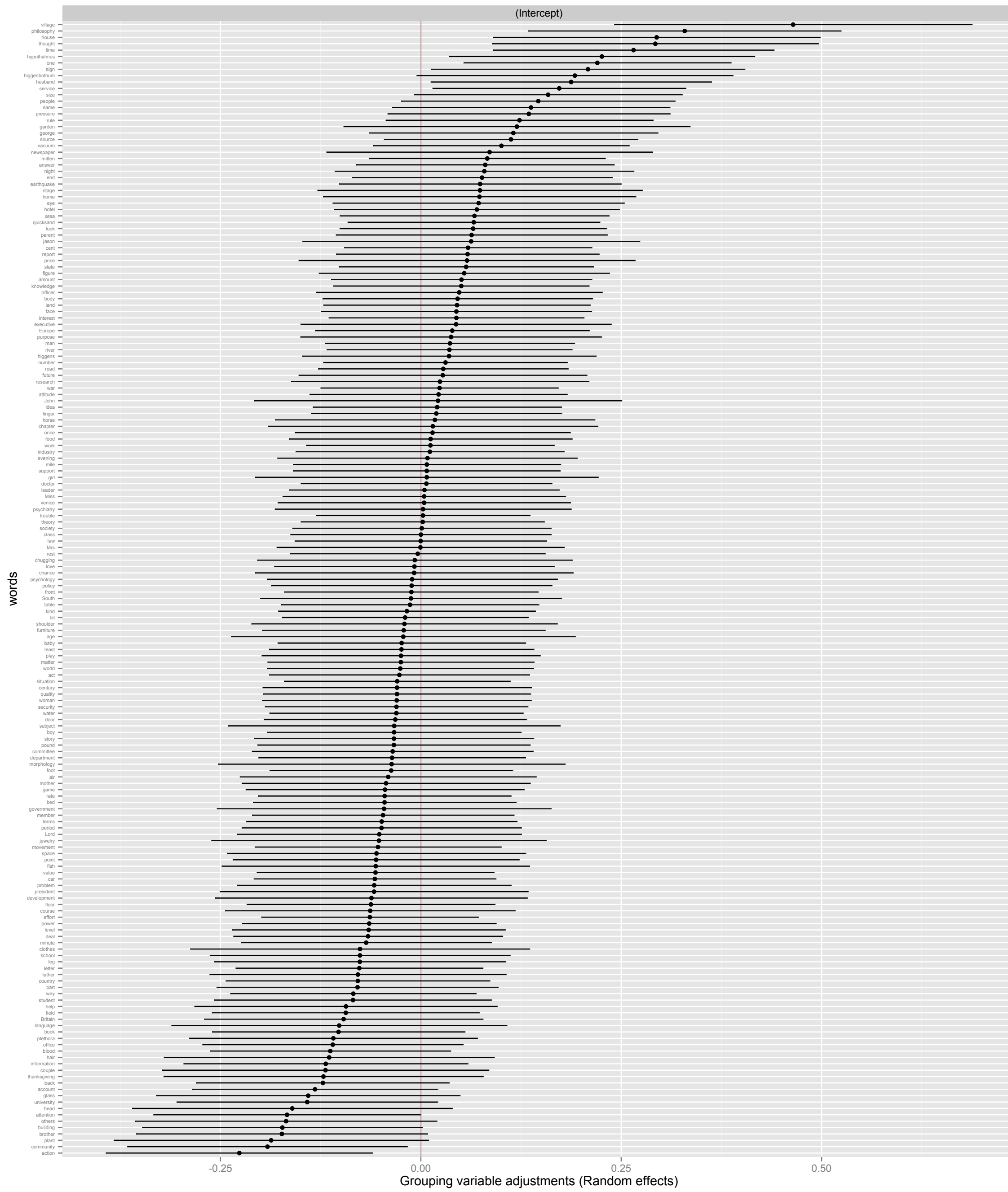


Figure D.17: **Plot of intercept adjustments (random intercepts) for words of the hierarchical linear model for all word durations using motion capture data and the signer-specific HMM** Because there are a large number of words, there are many levels on the y-axis. Although it is difficult to read individual words, as discussed in detail above, there is not much systematic variation of word duration between words. The sigmoidal shape is due to the fact that the intercept adjustments are modeled on a normal distribution. The levels on the y-axis are words, and they are ordered by the magnitude of the intercept adjustment: from smallest on the bottom to largest on the top.

Appendix E
Pinky extension model comparisons

Because there is some correlation between the timing predictors, Although the correlation is not perfect, it could result in overly large estimates of standard errors (resulting in large confidence intervals), or erratic estimates of coefficients. In order to ensure that these are not a problem with our full model, three additional models were fit, leaving one of the predictors out of each: one with the hold duration predictor removed (labelled below as hold durs. removed), one with the previous transition time predictor removed (labelled below as prev. trans. removed), and one with the following transition time predictor removed (labelled below as foll. trans. removed) These are visualized using coefficient plots in E.1, where each model is a different color. Full model outputs are listed in table E.1. As the coefficient plot and table shows, none of the coefficient estimates of the models deviate wildly, meaning that we can have confidence in the full model.

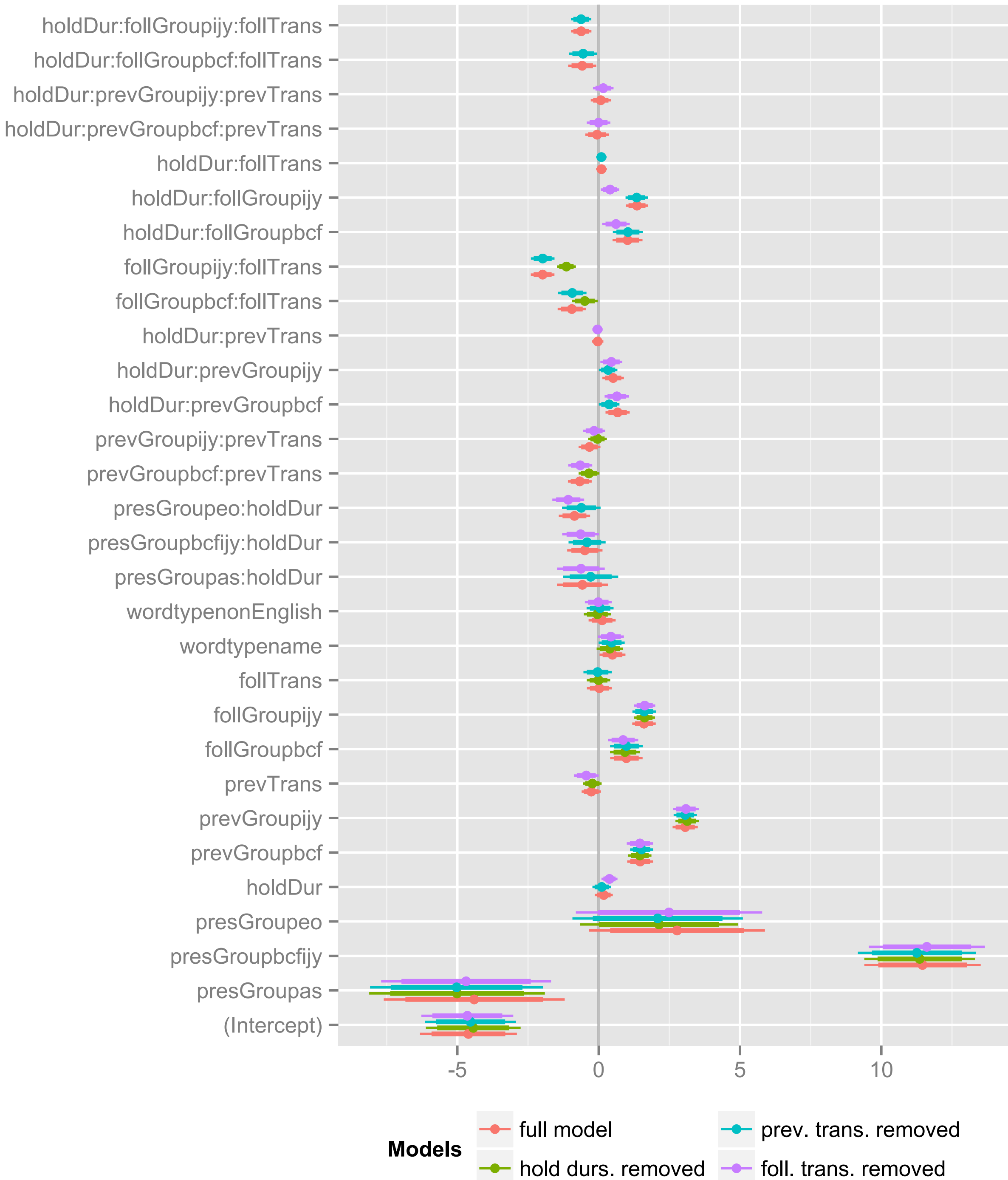


Figure E.1: **Coefficient plot for the predictors of the full hierarchical linear model hierarchical logistic regression model for pinky extension, as well as three reduced models, each leaving out one of the timing predictors** Thick lines represent 95% confidence, thin lines 99% confidence, and dots are the estimates of the coefficients (or intercept).

	full model	hold durs. removed	prev. trans. removed	foll. trans. removed
(Intercept)	$-4.61(0.67)^{***}$	$-4.44(0.65)^{***}$	$-4.53(0.62)^{***}$	$-4.65(0.63)^{***}$
presGroupas	$-4.40(1.24)^{***}$	$-5.01(1.21)^{***}$	$-5.03(1.19)^{***}$	$-4.69(1.17)^{***}$
presGroupbcfijy	$11.46(0.80)^{***}$	$11.36(0.76)^{***}$	$11.26(0.81)^{***}$	$11.61(0.80)^{***}$
presGroupeo	$2.77(1.21)^{*}$	$2.14(1.08)^{*}$	$2.09(1.17)$	$2.49(1.28)$
holdDur	$0.18(0.12)$		$0.11(0.13)$	$0.38(0.11)^{***}$
prevGroupbcf	$1.47(0.18)^{***}$	$1.46(0.16)^{***}$	$1.52(0.16)^{***}$	$1.46(0.18)^{***}$
prevGroupijy	$3.06(0.17)^{***}$	$3.13(0.16)^{***}$	$3.06(0.16)^{***}$	$3.09(0.18)^{***}$
prevTrans	$-0.26(0.13)^{*}$	$-0.22(0.13)$		$-0.44(0.17)^{**}$
follGroupbcf	$0.98(0.22)^{***}$	$0.93(0.21)^{***}$	$0.98(0.22)^{***}$	$0.86(0.21)^{***}$
follGroupijy	$1.60(0.16)^{***}$	$1.62(0.14)^{***}$	$1.61(0.16)^{***}$	$1.63(0.15)^{***}$
follTrans	$0.02(0.17)$	$-0.01(0.16)$	$-0.04(0.19)$	
wordtypename	$0.49(0.18)^{**}$	$0.39(0.18)^{*}$	$0.46(0.18)^{*}$	$0.43(0.18)^{*}$
wordtypenonEnglish	$0.12(0.19)$	$-0.05(0.19)$	$0.05(0.19)$	$-0.01(0.18)$
presGroupas:holdDur	$-0.58(0.35)$		$-0.28(0.38)$	$-0.63(0.33)$
presGroupbcfijy:holdDur	$-0.49(0.24)^{*}$		$-0.41(0.25)$	$-0.64(0.25)^{*}$
presGroupeo:holdDur	$-0.86(0.22)^{***}$		$-0.62(0.27)^{*}$	$-1.08(0.22)^{***}$
prevGroupbcf:prevTrans	$-0.67(0.16)^{***}$	$-0.34(0.14)^{*}$		$-0.65(0.16)^{***}$
prevGroupijy:prevTrans	$-0.32(0.15)^{*}$	$-0.04(0.13)$		$-0.16(0.15)$
holdDur:prevGroupbcf	$0.67(0.17)^{***}$		$0.37(0.14)^{**}$	$0.64(0.17)^{***}$
holdDur:prevGroupijy	$0.51(0.15)^{***}$		$0.33(0.13)^{**}$	$0.45(0.15)^{**}$
holdDur:prevTrans	$-0.03(0.08)$			$-0.04(0.07)$
follGroupbcf:follTrans	$-0.95(0.20)^{***}$	$-0.49(0.18)^{**}$	$-0.94(0.20)^{***}$	
follGroupijy:follTrans	$-1.98(0.16)^{***}$	$-1.14(0.13)^{***}$	$-1.98(0.16)^{***}$	
holdDur:follGroupbcf	$1.02(0.21)^{***}$		$1.03(0.21)^{***}$	$0.61(0.19)^{**}$
holdDur:follGroupijy	$1.36(0.15)^{***}$		$1.34(0.15)^{***}$	$0.40(0.13)^{**}$
holdDur:follTrans	$0.10(0.07)$		$0.10(0.07)$	
holdDur:prevGroupbcf:prevTrans	$-0.06(0.16)$			$0.00(0.16)$
holdDur:prevGroupijy:prevTrans	$0.07(0.14)$			$0.16(0.14)$
holdDur:follGroupbcf:follTrans	$-0.58(0.19)^{**}$		$-0.55(0.19)^{**}$	
holdDur:follGroupijy:follTrans	$-0.62(0.14)^{***}$		$-0.62(0.14)^{***}$	
AIC	11249.13	11496.79	11287.71	11512.73
BIC	11688.62	11753.85	11611.10	11836.12
Log Likelihood	-5571.57	-5717.40	-5604.85	-5717.36
Deviance	11143.13	11434.79	11209.71	11434.73
Num. obs.	29499	29499	29499	29499
Num. groups: apogeeId	13523	13523	13523	13523
Num. groups: wordList:word	599	599	599	599
Num. groups: apogeeLetter	26	26	26	26
Num. groups: annotator	19	19	19	19
Num. groups: signer	4	4	4	4
Variance: apogeeId.(Intercept)	3.46	3.82	3.55	3.95
Variance: wordList:word.(Intercept)	1.24	1.31	1.27	1.23
Variance: apogeeLetter.(Intercept)	2.15	2.01	2.07	1.84
Variance: apogeeLetter.follTrans	0.19	0.09	0.30	
Variance: apogeeLetter.prevTrans	0.12	0.06		0.30
Variance: apogeeLetter.holdDur	0.04		0.08	0.01
Variance: annotator.(Intercept)	1.28	1.19	1.26	1.29
Variance: signer.(Intercept)	0.69	0.74	0.53	0.54
Variance: signer.follTrans	0.05	0.06	0.07	
Variance: signer.prevTrans	0.01	0.02		0.03
Variance: signer.holdDur	0.01		0.00	0.00
Variance: Residual	1.00	1.00	1.00	1.00

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table E.1: Coefficient estimates and standard errors for the full hierarchical logistic model including all predictors for pinky extension (abbreviated: full model), as well as the reduced models: the same model with the hold duration predictor removed (abbreviated: hold durs. removed), the same model with the previous transition time predictor removed (abbreviated: prev. trans. removed), and the same model with the following transition time predictor removed (abbreviated: foll. trans. removed)

References

Ann, Jean, 1993. *A linguistic investigation of the relationship between physiology and handshape*. The University of Arizona dissertation.

Baayen, R. Harald, D.M. Davidson, & Douglas Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 390–412.

Barr, Dale J, Roger Levy, Christoph Scheepers, & Harry J Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68.255–278.

Bates, Douglas, 2010. *Linear mixed model implementation in lme4*. Department of Statistics, University of Wisconsin-Madison.

Battison, Robbin. 1978. *Lexical Borrowing in American Sign Language*. Silver Spring, Maryland: Linstock Press.

Bayley, Robert, Ceil Lucas, & Mary Rose. 2002. Phonological variation in American Sign Language: The case of 1 handshape. *Language Variation and Change* 14.19–53.

Bell-Berti, Fredericka. 1993. *Nasals, Nasalization, and the Velum*, volume 5 of *Phonetics and Phonology*, chapter Understanding velic motor control: studies of segmental context, 63–85. Academic Press, Inc.

Bladon, R Anthony W, & Ameen Al-Bamerni. 1976. Coarticulation resistance in English /l/. *Journal of Phonetics* 4.137–150.

Boersma, Paul. 2009. Cue constraints and their interactions in phonological perception and production. *Phonology in Perception* 15.55–110.

Bornstein, Harry. 1965. *Reading the manual alphabet: A research program for developing a filmed program for teaching the manual alphabet*. Centennial Series 2. Gallaudet College Press.

Bradlow, Ann R. 1995. A comparative acoustic study of English and Spanish vowels. *The Journal of the Acoustical Society of America* 97.1916.

Brentari, Diane. 1998. *A prosodic model of sign language phonology*. The MIT Press.

——, & Petra Eccarius. 2010. *Sign Languages*, chapter Handshape contrasts in sign language phonology, 284–311. Cambridge, UK: Cambridge University Press.

——, & Carol Padden. 2001. *Foreign vocabulary in sign languages: A cross-linguistic investigation of word formation*, chapter Native and foreign vocabulary in American Sign Language: A lexicon with multiple origins, 87–119. Mahwah, NJ: Lawrence Erlbaum.

Browman, Catherine P. 1994. Lip aperture and consonant releases. In *Phonological Structure and Phonetic Form*, ed. by Patricia A. Keating, volume 3 of *Papers in Laboratory Phonology*, 331–353.

——, & Louis Goldstein. 1986. Towards an articulatory phonology. *Phonology yearbook* 3.9–252.

——, & ——. 1992. Articulatory phonology: An overview. Technical report, Haskins Laboratories.

Cheek, Davina Adrianne, 2001. *The Phonetics and Phonology of Handshape in American Sign Language*. University of Texas at Austin dissertation.

Chomsky, Noam, & Morris Halle. 1968. *The Sound Pattern of English*. MIT Press.

Chung, G.Y., & S. Seneff. 1999. A hierarchical duration model for speech recognition based on the angie framework1. *Speech Communication* 27.113–134.

Chung, Hyunju, Eun Jong Kong, Jan Edwards, Gary Weismer, Marios Fourakis, & Youngdeok Hwang. 2012. Cross-linguistic studies of children's and adults' vowel spaces. *The Journal of the Acoustical Society of America* 131.442.

Cohn, Abigail C, 1990. *Phonetic and phonological rules of nasalization*. UCLA dissertation.

Cormier, Kearsy, Adam Schembri, & Martha E. Tyrone. 2008. One hand or two? Nativisation of fingerspelling in ASL and BANZSL. *Sign Language & Linguistics* 11.3–44.

Crasborn, Onno Alex, 2001. *Phonetic implementation of phonological categories in Sign Language of the Netherlands*.

Crystal, Thomas H, & Arthur S House. 1982. Segmental durations in connected speech signals: Preliminary results. *The Journal of the Acoustical Society of America* 72.705–716.

——, & ——. 1988. Segmental durations in connected-speech signals: Current results. *The journal of the acoustical society of America* 83.1553–1573.

Daniloff, Raymond, & R Hammarberg. 1973. On defining coarticulation. *Journal of Phonetics* 1.239–248.

Deng, L, & D Sun. 1994. Phonetic classification and recognition using hmm representation of overlapping articulatory features for all classes of English sounds. In *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on*, volume 1, I–45. IEEE.

Eccarius, Petra, 2002. Finding common ground: A comparison of handshape across multiple sign languages. Master's thesis, Purdue University.

——, 2008. *A constraint-based account of handshape contrast in sign languages*. Purdue University dissertation.

——, Rebecca Bour, & Robert A Scheidt. 2012. Dataglove measurement of joint angles in sign language handshapes. *Sign language and linguistics* 15.39.

——, & Diane Brentari. 2008. Handshape coding made easier a theoretically based notation for phonological transcription. *Sign Language & Linguistics* 11.69–101.

Emmorey, Karen, Allison Bassett, & Jennifer Petrich, 2010. Processing orthographic structure: Associations between print and fingerspelling. Poster at 10th Theoretical Issues in Sign Language Research conference, West Lafayette, Indiana.

——, & Jennifer Petrich. 2011. Processing orthographic structure: Associations between print and fingerspelling. *Journal of Deaf Studies and Deaf Education* 17.194–204.

Farnetani, Edda, & Daniel Recasens. 1999. *Coarticulation: Theory, data and techniques*, chapter Coarticulation models in recent speech production theories, 31–65. Cambridge, UK: Cambridge University Press.

Fenlon, Jordan, Adam Schembri, Ramas Rentelis, & Kearsy Cormier. 2013. Variation in handshape and orientation in British Sign Language: The case of the 1-hand configuration. *Language & Communication* 33.69—91.

Gales, Mark, & Steve Young. 2008. The application of hidden Markov models in speech recognition. *Foundations and Trends in Signal Processing* 1.195–304.

Geer, Leah, 2010. ASL fingerspelling: Comparing native and non-native signers. Unpublished manuscript Gallaudet University; Washington, DC.

——, & Jonathan Keane. 2014. Exploring factors that contribute to successful fingerspelling comprehension. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, ed. by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, & Stelios Piperidis, Reykjavik, Iceland. European Language Resources Association.

Gelman, Andrew. 2013. Commentary: p values and statistical practice. *Epidemiology* 24.69–72.

——, & Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

——, & Francis Tuerlinckx. 2000. Type s error rates for classical and bayesian single and multiple comparison procedures. *Computational Statistics* 15.373–390.

Gick, Bryan, Ian Wilson, Karsten Koch, & Clare Cook. 2004. Language-specific articulatory settings: Evidence from inter-utterance rest position. *Phonetica* 61.220–233.

glmm wiki. 2014. glmm.wikidot.com/faq.

Grosvald, Michael, & David Corina. 2008. Location-to-location coarticulation: A phonetic investigation of American Sign Language. In *Proceedings of the 24th Northwest Linguistics Conference (NWLC)*, 59–66.

Hammarberg, Robert. 1976. The metaphysics of coarticulation. *Journal of Phonetics* 4.353–363.

Hanson, Vicki L. 1981. When a word is not the sum of its letters: Fingerspelling and spelling. In *Teaching American Sign Language as a Second/Foreign Language. Proceedings of the Third National Symposium on Sign Language Research and Teaching. Boston, MA. October 26–30, 1980*, ed. by F. Caccamise, M. Garretson, & U. Bellugi, 176–85.

——. 1982. Use of orthographic structure by deaf adults: Recognition of fingerspelled words. *Applied Psycholinguistics* 3.343–356.

Haptonstall-Nykaza, Tamara S., & Brenda Schick. 2007. The transition from fingerspelling to English print: Facilitating English decoding. *Journal of Deaf Studies and Deaf Education* 12.172–183.

Hoopes, Rob. 1998. *Pinky extension and eye gaze: Language use in Deaf communities*, chapter A preliminary examination of pinky extension: Suggestions regarding its occurrence, constraints, and function, 3–17. Washington, DC: Gallaudet University Press.

Jaeger, T Florian. 2008. Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of memory and language* 59.434–446.

Jantunen, Tommi. 2013. Signs and transitions: Do they differ phonetically and does it matter? *Sign Language Studies* 13.211–237.

Jerde, Thomas E., John F. Soechting, & Martha Flanders. 2003. Coarticulation in fluent fingerspelling. *Journal of Neuroscience* 23.2383–2393.

Johnson, Robert E, & Scott K Liddell. 2011a. Toward a phonetic representation of hand configuration: The thumb. *Sign Language Studies* 12.316–333.

——, & ——. 2011b. Toward a phonetic representation of signs: Sequentiality and contrast. *Sign Language Studies* 11.241–274.

Keane, Jonathan, 2010. Segment duration and the phonetics of fingerspelling ASL. Master's thesis, University of Chicago. defended 18 October 2010 Advisors: Jason Riggle and Diane Brentari.

——. 2014. AMOHS python module. software. Version 0.1.0 zenodo.org/record/11456.

——, Diane Brentari, & Jason Riggle. 2012a. Handshape and coarticulation in ASL fingerspelling. conference presentation. Linguistic Society of America 2012 Annual Meeting.

——, Diane Brentari, & Jason Riggle. 2012b. Handshape assimilation in ASL fingerspelling. conference presentation. Societas Linguistica Europaea, Loanwords as Evidence in Formal Linguistics.

——, Diane Brentari, & Jason Riggle. 2013. Dispelling prescriptive rules in ASL fingerspelling: the case of -E-. poster. Theoretical Issues in Sign Language Research 11; London, UK.

——, Diane Brentari, & Jason Riggle. forthcoming. *The Segment in Phonology and Phonetics*, chapter Segmentation and pinky extension in ASL fingerspelling. Wiley.

——, Erin Dahlgren, & Jason Riggle. 2011. Variation in segment duration in ASL fingerspelling. conference presentation. Linguistic Society of America 2011 Annual Meeting.

——, & Leah Geer. 2014. Holds or transitions: Which is more important? conference presentation. Chicago Linguistics Society 50.

Keating, Patricia A. 1980/1990. The window model of coarticulation: articulatory evidence. *Papers in laboratory phonology I* 26.451–470.

Klatt, Dennis H. 1975. Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech, Language and Hearing Research* 18.686.

Klatt, D.H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America* 59.1208.

Ladd, D Robert. 2011. Phonetics in phonology. *The Handbook of Phonological Theory, Second Edition* 348–373.

Lehiste, I. 1972. The timing of utterances and linguistic boundaries. *The Journal of the Acoustical Society of America* 51.2018.

Levinson, S.E. 1986. Continuously variable duration hidden Markov models for automatic speech recognition. *Computer Speech & Language* 1.29–45.

Liddell, Scott, & Robert E Johnson. 1989. American Sign Language: A phonological base. *Sign Language Studies* 64.195–277.

Liddell, Scott K. 1978. Nonmanual signals and relative clauses in American Sign Language. *Understanding Language Through Sign Language Research* 59–90.

——, & Robert E Johnson. 2011a. A segmental framework for representing signs phonetically. *Sign Language Studies* 11.408–463.

——, & ——. 2011b. Toward a phonetic representation of hand configuration: The fingers. *Sign Language Studies* 12.5–45.

Lindblom, Björn. 1963. On vowel reduction. Technical Report 29, Royal Institute of Technology, Speech Transmission Laboratory, Stockholm.

Livescu, Karen, & J. Glass. 2001. Segment-based recognition on the phonebook task: Initial results and observations on duration modeling. In *Seventh European Conference on Speech Communication and Technology*. Citeseer.

Lucas, Ceil, Robert Bayley, Mary Rose, & Alyssa Wulf. 2002. Location variation in American Sign Language. *Sign Language Studies* 2.407–40.

Mandel, Mark, 1981. *Phonotactics and morphology in American Sign Language*. University of California, Berkeley dissertation.

Mann, Virginia A, & Bruno H Repp. 1980. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics* 28.213–228.

Mauk, Claude E, 2003. *Undershoot in two modalities: Evidence from fast speech and fast signing*. University of Texas at Austin dissertation.

——, & Martha E Tyrone. 2008. Sign lowering as phonetic reduction in American Sign Language. In *Proceedings of the 8th International Symposium on Speech Production*, 185–188.

——, & ——. 2012. Location in asl: Insights from phonetic variation. *Sign Language & Linguistics* 15.128–146.

Meier, Richard P, & Amanda S Holzrichter. 2000. Child-directed signing in American Sign Language.

Mitchell, Ross E., Travas A. Young, Bellamie Bachleda, & Michael A. Karchmer. 2006. How many people use ASL in the united states? why estimates need updating. *Sign Language Studies* 6.306–356.

Moon, Seung-Jae, & Björn Lindblom. 1994. Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical society of America* 96.40.

Nittrouer, Susan, & Michael Studdert-Kennedy. 1987. The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech, Language and Hearing Research* 30.319.

Öhman, Sven EG. 1966. Coarticulation in vcv utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America* 39.151.

——. 1967. Numerical model of coarticulation. *The Journal of the Acoustical Society of America* 41.310.

Oller, D.K. 1973. The effect of position in utterance on speech segment duration in English. *The Journal of the Acoustical Society of America* 54.1235.

Padden, Carol. 1991. *Theoretical issues in sign language research*, chapter The Acquisition of Fingerspelling by Deaf Children, 191–210. The University of Chicago Press.

——, & Darline Clark Gunsauls. 2003. How the alphabet came to be used in a sign language. *Sign Language Studies* 4.10–33.

——, & Barbara Le Master. 1985. An alphabet on hand: The acquisition of fingerspelling in deaf children. *Sign Language Studies* 47.161–172.

Parisot, Anne-Marie, 2003. *Accord et cliticisation: les cas des verbes à forme rigide en langue des signes québécoise*. Montréal, Université du Québec à Montréal, Département de linguistique et de didactique des langues dissertation.

Peirce, Jonathan W. 2007. PsychoPy — Psychophysics software in Python. *Journal of neuroscience methods* 162.8–13.

Peterson, G.E., & I. Lehiste. 1960. Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America* 32.693.

Phillips, Dusty. 2010. *Python 3 object oriented programming*. Packt Publishing Ltd.

Poeppel, David, & William Idsardi. 2011. Recognizing words from speech: the perception-action-memory loop. *Lexical representation: A multidisciplinary approach. New York, NY: Mouton de Gruyter* 171–196.

——, William J Idsardi, & Virginie van Wassenhove. 2008. Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363.1071–1086.

Port, R.F. 1981. Linguistic timing factors in combination. *The Journal of the Acoustical Society of America* 69.262.

Quinto-Pozos, David. 2010. Rates of fingerspelling in American Sign Language. Poster at 10th Theoretical Issues in Sign Language Research conference, West Lafayette, Indiana.

r-sig-mixed-models listserv. 2010. stat.ethz.ch/pipermail/r-sig-mixed-models/2010q2/thread.html#3709.

Rabiner, Lawrence. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77.257–286.

Raphael, Lawrence J. 1972. Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *The Journal of the Acoustical Society of America* 51.1296.

Reich, Peter A., & Margaret Bick. 1977. How visible is visible english? *Sign Language Studies* 14.59–72.

Repp, Bruno H. 1981. Two strategies in fricative discrimination. *Perception & Psychophysics* 30.217–227.

Richardson, Matthew, Jeff Bilmes, & Chris Diorio. 2000. Hidden-articulator markov models: performance improvements and robustness to noise. In *INTERSPEECH*, 131–134.

Sandler, Wendy. 1986. The spreading hand autosegment of American Sign Language. *Sign Language Studies* 50.1–28.

——. 1989. *Phonological Representation of the Sign: Linearity and Nonlinearity in American Sign Language*. Foris Pubs USA.

——, & Diane Lillo-Martin. 2006. *Sign language and linguistic universals*. Cambridge Univ Press.

Seidl, Amanda, Alex Cristià, Amelie Bernard, & Kristine H Onishi. 2009. Allophonic and phonemic contrasts in infants' learning of sound patterns. *Language Learning and Development* 5.191–202.

Stokoe, William C. 1960. *Sign language structure: An outline of the visual communication systems of the American deaf*. Number 8 in Studies in Linguistics Occasional Papers. Buffalo 14, New York: Department of Anthropology and Linguistics, University of Buffalo.

——, Dorothy C. Casterline, & Carl G. Croneberg. 1965. *A Dictionary of American Sign Language on Linguistic Principles*. Washington DC: Gallaudet College Press.

Tufte Forum Users. 2014. Instructions at the point of need. www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0001HD.

Tyrone, Martha E. 2002. Overview of capture techniques for studying sign language phonetics. In *Gesture and sign languages in human-computer interaction*, ed. by Ipke Wachsmuth & Timo Sowa, 15–24. Berlin, Heidelberg: Springer Berlin / Heidelberg SN -.

——, J. Kegl, & H. Poizner. 1999. Interarticulator co-ordination in Deaf signers with Parkinson's disease. *Neuropsychologia* 37.1271–1283.

——, & Claude E Mauk. 2010. Sign lowering and phonetic reduction in American Sign Language. *Journal of phonetics* 38.317–328.

——, Hosung Nam, Elliot L Saltzman, Gaurav Mathur, & Louis Goldstein. 2010. Prosody and movement in American Sign Language: A task-dynamics approach. In *Speech Prosody 2010*, 1–4.

van der Hulst, H. 1995. The composition of handshapes. *Working Papers in Linguistics* 1–17.

Van der Kooij, Els. 2002. Phonological categories in Sign Language of the Netherlands. *The Role of Phonetic Implementation and Iconicity. LOT, Utrecht* .

Šarić, Marin. 2011. Libhand: A library for hand articulation. Version 0.9 libhand.org.

Wallis, John. 1653/1972. *Grammatica linguae anglicanae*. London: Longman. Edited/translated by J. A. Kemp.

Whitworth, Cecily, 2011. *Features, Clusters, and Configurations: Units of Contrast in American Sign Language Handshapes*. Gallaudet University dissertation.

Wilbur, Ronnie. 1999. Stress in ASL: Empirical evidence and linguistic issues. *Language and Speech* 42.229–250.

Wilbur, Ronnie B, & Brenda S Schick. 1987. The effects of linguistic stress on ASL signs. *Language and Speech* 30.301–323.

Wilcox, Sherman. 1992. *The Phonetics of Fingerspelling*. Amsterdam: John Benjamins Publishing Company.

Wilson, Ian, 2006. *Articulatory settings of French and English monolingual and bilingual speakers*. University of British Columbia dissertation.

——, & Bryan Gick. 2006. Articulatory settings of French and English monolinguals and bilinguals. Poster at The Journal of the Acoustical Society of America.

Zakia, Richard D, & Ralph Norman Haber. 1971. Sequential letter and word recognition in deaf and hearing subjects. *Perception & Psychophysics* 9.110–114.