# Transitoin time and the phonetics of fingerspelling*

Jonathan Keane

February 18, 2011

## 1 Introduction

### 1.1 Fingerspelling in American Sign Language

American Sign Language — ASL — is used by approximately 500,000 to 2 million people in the USA and Canada[1], the majority of which are deaf. As with other signed 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 and Gunsauls, 2003). Fingerspelling is used more frequently in ASL than in other sign languages (Padden, 1991). Fingerspelling is the representation of English words through a series of handshapes, each of which represents a letter in the word. Every letter used in English is given a unique combination of handshape, orientation, and in a few cases movement path[2] (Cormier et al. (2008) among others). These hand configurations are used sequentially to represent an English word. Figure 1 shows the handshapes for ASL. The orientation of each handshape is altered in this figure for ease of learning. In reality, all letters are articulated with the palm facing forward, away form the signer,

*This project simply couldn't have happened without the contributions of many people. First and foremost my advisors Jason Riggle and Diane Brentari, for their encouragement and insight into the many topics touched on here. My colleague and collaborator, Susan Rizzo, who helped maintain my sanity, and did a tremendous amount of work organizing and executing data coding. I also must thank all of the collaborators that helped in this project: Karen Livescu, Greg Shakhnarovich, Raquel Urtasun, Erin Dahlgreen, Katie Henry, and Katherine Mock. This project also couldn't be possible without the help of the two signers who fingerspelled for many hours for us. I gained valuable insight from comments at the 2011 LSA in Pittsburgh, as well as smaller talks I've given at the University of Chicago. Finally, I must thank my friends and family, especially Jonathan Goya for keeping me sane and listening to me babble on about fingerspelling for much longer than they ought to be subject to. Although all of these people helped immensely, any mistakes or omissions are entirely my own.

[1]These numbers range widely across many sources which was documented by Mitchell et al. (2006).

[2]Traditionally movement is said to only be used for the letters -J- and -Z- as well as to indicate some instances of letter doubling. Although in fluent fingerspelling many letters have movement of some type.

except for -H- (in, towards the signer), -G- (in, towards the signer), -P- (down), -Q- (down) and the end of -J- (to the side).[3]

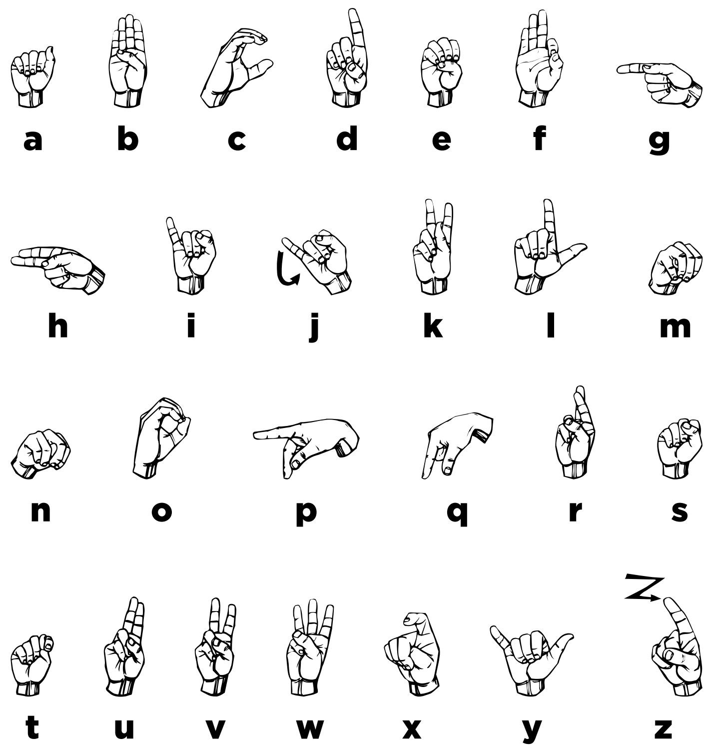



Figure 1: Handshapes for ASL fingerspelling.

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 and Gunsauls (2003). In early research many situated fingerspelling as a mechanism to fill in vocabulary items that are missing in ASL (Padden and Le Master, 1985). On further investigation, it has been discovered that this is not the whole story. 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[4] and G-E-T-O-U-T. An example of the second is

[3]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.

[4]I'm choosing to adopt the clear and elegant typographic conventions of Cormier et al. (2008). Not only are this consistent with separating ASL forms from the text as well as reliably marking the difference between ASL native signs (smallcaps: GROUP) and fingerspelled forms (smallcaps, with hyphens: A-T-L-A-N-T-I-C) but it also is easier to read than many other systems. Following from this, single finger spelled letters will be flanked by hyphens on either side (EG -T-). Finally when I'm discussing transition data following the letter or preceding a letter a dash will indicate which side I'm talking about (T- or -T respecitvely).

a teacher fingerspelling P-R-O-B-L-E-M as in a scientific problem in a science class, because the ASL sign for P-R-O-B-L-E-M has a separate meaning that is not quite compatible with the English sense of the word. 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 and Gunsauls, 2003).

Finally, there is already some literature on the nativization process from fingerspelled form to lexicalized sign (Brentari and Padden, 2001; Cormier et al., 2008). Because it is an integral part of ASL, fingerspelling should be analyzed as any other part of a language. The phonetics and phonology of fingerspelling are in some ways related to ASL, because it uses many of the same articulators, but there are important differences. Thus it is important that we study the phonetics and phonology of fingerspelling as well as of ASL generally. This is not how research has been approached previously; with the exception of (Wilcox, 1992), (Tyrone et al., 1999), Emmorey et al. (2010), and 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. This study is mainly limited by its method of data collection involving infrared emitters and infrared sensitive cameras, but no standard video. Tyrone et al. (1999) looks at fingerspelling in parkinsonian signers, and what phonetic features are compromised in parkinsonian fingerspelling. Emmorey et al. (2010) studied segmentation 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.

## 1.2 The importance of segment duration

If we are studying fingerspelling as any other component of language we must establish basic phonetic features.[5] These basic phonetic features will inform many areas of linguistic research, including automated sign language recognition work that is underway in a variety of places.

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 (EG individual variation, utterance speed, and familiarity with the target item) as well as a similarly numerous micro factors (EG segment type, preceding

[5] I'm adopting most of my terminology from the field of linguistics in general. This is not uncontroversial, especially because many of the technical terms (EG *phonetics*, *phonology* et c.) are etymologically related to sound and speech. This is not meant to suggest that spoken language is in anyway superior to manual, but rather is used to confirm that, with a few slight modifications, manual languages are analyzable using current linguistic methods, and show very little difference from spoken languages in many areas.

and following segments, articulatory complexity, and stress). (See Klatt (1976) for a review, and Peterson and Lehiste (1960); Lehiste (1972); Oller (1973); Port (1981) for specifics.) 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. As with all phonetic features, duration adds crucial information to the language signal. On it's own it might not provide much information, but when added with other features it adds important information to the language signal. Voice onset time is a similar tiny phonetic feature that has a vast body of research on how it greatly influences the perception of a few segments. 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 and Glass, 2001; Chung and Seneff, 1999; Levinson, 1986).

When treating ASL fingerspelling as any other part of language, an analysis of segment duration becomes important. 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. We expect that a letter sequence that involves two similar handshapes should be somewhat quicker than one that involves two vastly different handshapes. Things are actually a bit more complicated; as will be discussed later at the end of section 3.5, there are phenomenon that seem to follow the Obligatory Contour Principle (OCP). As we have seen in phonetic research elsewhere what comes before and after a segment has an influence on the intermediate segment generally, and its duration specifically. 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. Looking at language this way we see that there are many factors that are general across all modalities; phonetics and phonology are tied in many ways to articulation.

Segment duration in fingerspelling has one large difference from segment duration in speech, in that the static handshape for each letter is generally not held for any length of time, unlike some segments in speech (IE vowels). If we conceptualize fingerspelling as a series of target handshapes that the articulators move through, it becomes clear that the transition time between targets is a more appropriate measure. With respect to fingerspelling, transition time can be substituted for the concept of segment duration with little or no change to the underlying theory. Further discusion on this point is in section 2.2.

Transition time in fingerspelling is a critical jumping off point for the study of the phonetics of fingerspelling and fingerspelling in general, as well as the development of automated sign recognizers. We have seen in spoken language research that individual phonetic features contribute a small bit of information to the language signal, but when each feature is added together a complete signal emerges. To get basic phonetic and phonological information for fingerspelling, we generated a large database of multiple signers fingerspelling various words. Section 2 discusses the methods used to collect this data. Section 3 discusses the results and data obtained from this database. Finally section 4 discusses the theoretical implications, as well as future directions for research in fingerspelling phonetics.

## 2 Methods

We generated a large database of fingerspelled words for multiple concurrent linguistic and computer-vision projects. This is the source of all of the data presented below. It was recorded with the intent to use the data in multiple ways, and thus be as flexible as possible.

### 2.1 Fingerspelling specifications

We recorded two deaf, native ASL signers. The signers were related, which might lead to their fingerspelling being similar, but they also represent a small scale language community by themselves. One signer was a man in his 20s, the other a woman in her 50s. Both signers are bilingual in English.

We constructed a word list with 297 words. 97 are names, 100 are nouns, and 100 are non-English words[6]. These words were chosen to get examples of as many letters in as many different contexts as possible, and are not necessarily representative of the frequency of letter, or letter combinations in English, or even commonly fingerspelled words. The complete wordlist can be found in Appendix A.

[6]These are also called foreign, although that's not entirely accurate, since all fingerspelled words are in some sense foreign to native signers.

We asked the signers to sign at one of two speeds at the beginning of each session. One speed, *careful*, was supposed to be slow, and deliberate.[7] The other speed, *normal*, was supposed to be fluid and conversational, we told the signers to fingerspell naturally, as if they were talking to another native signer.[8] Although the careful is artificially slow for most situations, it is conceivable that it might be used in some situations, especially when communicating with novice signers. It also provides us with indirect evidence of difficulty in production and processing, as will be discussed later (section 3.3).

The video was collected during 4 sessions. Each signer completed 2 sessions, 1 careful speed, and 1 normal speed. Each session lasted between 30 and 50 minutes[9]. 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 signed 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 sign 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.

Once this video was recorded and compressed, and 3–4 human coders identified the apogee of each letter. We defined apogee as the point where the articulators changed direction to proceed on to the next letter (IE 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 letters defied definition in this manner, namely -J- and -Z-, since they have movement. With these two letters coders were asked to just indicate an apogee when they could determine that it was one of these two letters. In order to determine the most likely apogee locations we averaged the apogees from each coder using an algorithm that minimized the mean average distance between the individual coders' apogees. We allowed for misidentified apogees by penalizing missing or extra apogees. Using logs from the recording session we added a best guess at the letter of each apogee using left edge forced alignment. Finally we had someone trained in fingerspelling go through each file and verify that this combined apogee was in the correct location, and the letter associated with it matched the letter being signed. Again, with the two letters involving movement, there was a slight problem: where to put the apogee. To standardize this we coded the apogee for a -J- at the end of the movement of the wrist, where the pinky was pointing out and up; for -Z- we coded the apogee when the index finger reached the end of the top bar

[7]The instructions, given in ASL were to "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.

[8]Again in ASL: "proceed at normal speed and in your natural way of fingerspelling"

[9]Most of the variation was due to the speed: whether the session was a careful session or a normal session

of the -z- as it was traced in the air. This (working) definition of the apogee of -z- might seem odd, but it's entirely practical. In fluent, normal speed fingerspelling the motion of -z-s were often reduced to a single, horizontal stroke. Finally the information from these verified files was imported into a MySQL database to allow for easy manipulation and querying.

## 2.2 Apogees and transition time

There is a persistent idea that segments are like beads strung together, with each being separate and separable. If this were the case it would be extremely easy to describe where the articulation of one segment stops and the next segment begins. This does not seem to be the case with fingerspelling, or to some extent language production generally. There is no easily defined point between the apogees of two adjacent letters where one could say that the articulation of the first has ended, and the articulation of the second has began. The language signal is the result of the articulators moving from target to target.

Our data can be visualized as in figure 2. Each word is made of apogees representing the points in time where the instantaneous velocity of the articulators approached zero, and where the hand most closely resembled the canonical form. Between each apogee is the time it takes to get from one apogee to the next.

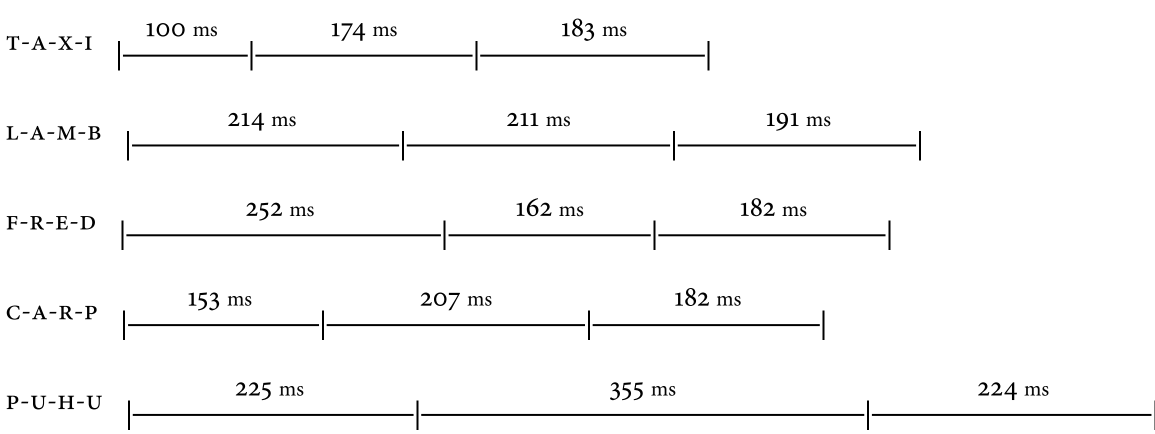



Figure 2: visualizations of T-A-X-I, L-A-M-B, F-R-E-D, C-A-R-P, and P-U-H-U (signer: **s1** speed: **normal**)

As with all experimental data with this many points of data there will always be some outliers. There was, for instance, a problem with the equipment in the middle of fingerspelling one word, but because the signer felt like they had completed the whole word, they pressed the green button, giving us two peaks that had alleged transitionss of over 12 seconds each. In order to get rid of obvious errors like this we excluded data that was further than 2 standard deviations from the mean, based on the speeds that they were signed at (only 140 points out of 12,102 total). There were also three words that included spaces in the presentation to the signers: El Salvador, San Francisco, and Oak Park. There were considerable pauses where the spaces were in both of the signers. Because of this we excluded these three words from our data, but they suggest that we need to investigate compound words further to figure out if this pause is a result of the compound, or the orthography. This resulted in a data set of 11,962 apogees, with letters in nearly every combination of signer, speed, and wordtype. The only exception to this is there are no foreign words with -Q- in them.

## 3 Results

As discussed in section 1.2 we expect that based on a variety of factors there will be differences in transition time. For now we will only look at macro factors involved in transitions time. First we expect the careful speed to have longer transition times than normal speed. Second, we expect our signers will differ to some extent in how quickly they sign due to individual variation. We also expect that the category of the word might affect the transition time, with letters from unfamiliar, non-English words having longer transition times. A quick look at an analysis of variance (ANOVA) (table 1) and we see that all of these are borne out. There are significant differences between the mean transition times for word types, speed, and signers. There are also interactions between word type and speed, speed and signer, as well as all three (speed-signer-word type). The interaction between word type and speed indicates that different word types are treated differently in the two different speeds. The interaction between speed and signer indicates that the signers didn't vary in the exact same way between their two speeds. There is no interaction between word type and signer, which says that both signers treated the different word types the same. We will explore each of these in the following sections.

One thing to note, the data used in the ANOVA is log transformed. This makes the data a bit more normally distributed, and thus makes the ANOVA more accurate. Because it's easier to interpret time in msec than log(msec), we've used the untransformed data in all of the boxplots that follow.

Finally, we also expect that words of different lengths might be fingerspelled at different speeds. In conjunction with this we also expect different positions in the word might exhibit different transition times. Both of these effects have been shown in speech, and similarly are fruitful to take into account when attempting to implement automatic recognition of fingerspelling. Besides these functional uses, both of these differences

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| wordtype | 2 | 33.19 | 16.59 | 220.34 | 0.0000 |
| speed | 1 | 1425.48 | 1425.48 | 18927.22 | 0.0000 |
| signer | 1 | 13.90 | 13.90 | 184.55 | 0.0000 |
| wordtype:speed | 2 | 13.04 | 6.52 | 86.58 | 0.0000 |
| wordtype:signer | 2 | 0.20 | 0.10 | 1.36 | 0.2570 |
| speed:signer | 1 | 232.93 | 232.93 | 3092.84 | 0.0000 |
| Residuals | 12267 | 923.87 | 0.08 | | |

Table 1: ANOVA table for log(transitiontime)

have the ability to inform us about language-memory interaction and the constraints of human production, or even human language perception.

### 3.1 Differences between speeds

As figure 3 suggests there is a difference between the two speeds. This and the following figures are all box and whisker plots where the line represents the median. The box represents the interquartile range of the data — the middle 50% of the data. The whiskers are set at either 1.5 times the lower or upper quartile, or the most extreme data point seen; whichever is less. Any data points that exceed 1.5 times the lower or upper quartile are plotted as plusses.

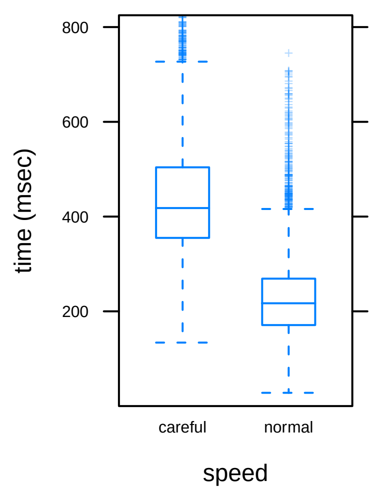
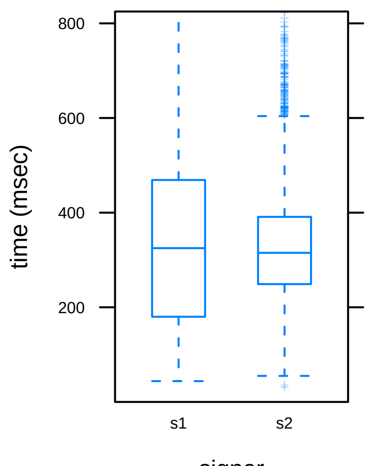



Figure 3: transition time between speeds

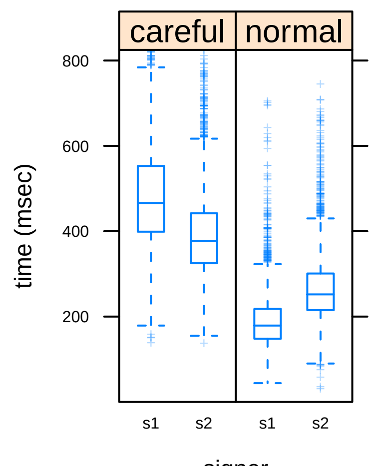
Where the transition time for careful is significantly ($F$ = 18927.22, $p$ < 0.0001) longer than normal by about 211.92 msec. The mean transition time for careful speed is 441.41 msec, and the standard deviation of 124.29 msec, the shortest transition time is 134 msec, and the longest is 988 msec. The mean for normal speed is 229.49 msec, the standard deviation is 86.44 msec, the shortest transition time is 28 msec, and the longest is 906 msec. Interestingly the careful speed is more spread out, indicating that it is less regular than the normal speed. This confirms exactly what we expect, and confirms that fingerspellers, when asked to fingerspell slower, space out each letter. But we have done the due diligence work, and confirmed that this is the case.

## 3.2 Differences between signers

Looking at figure 4a it appears that the signers have about the same mean transition time, but that one (s2) has a much smaller spread. The ANOVA confirms that these two distributions are different ($F = 184.55$, $p < 0.0001$). Signer s1 has a mean of 340.68 msec and a standard deviation of 179.14 msec. Their shortest transition time is 44 msec, and longest 988 msec. Singer s2 has a mean of 331.27 msec, and a standard deviation of 115.47 msec. Their shortest transition time is 28 msec, and longest 968 msec.




(a) transition time between signers




(b) transition time between signers, by speed

Figure 4: speed, and the interaction between speed and signer

It's not quite as simple as saying that the signers just differ in that one has less variation.[10] Looking at the interaction between signer and speed (figure 4b) we see why this is the case. Signer s2's careful speed transition times are shorter and normal speed transition times longer than s1's. When taking speed as well as signer into consideration the difference is significant with a much higher F-value. ($F = 3092.84$, $p < 0.0001$) See table 2 for detailed information about means, standard deviations, shortest, and longest transition times for each speed-signer combination. Here we see that there is considerable difference between the two signers. The mean of signer s2's careful transition times are 90.48 msec shorter, and their normal transition times are 74.91 msec longer than s1's.

Again this is what we expect from the research on spoken language segment duration, however, when we compare the normal/careful ratios of each signer to those in speech we find that there is a difference between

[10]In fact, when looking at density plots and not factoring out speeds it appears the signer s2 has a single speed with a unimodal distribution.

| | mean | sd | shortest | longest |
|---|---|---|---|---|
| s1 careful | 486.52 | 124.41 | 135 | 988 |
| s1 normal | 191.71 | 71.68 | 44 | 897 |
| s2 careful | 396.04 | 106.37 | 134 | 968 |
| s2 normal | 266.62 | 83.57 | 28 | 906 |

Table 2: Mean, standard deviation, shortest, and longest transition time for speed-signer combinations, all in msec

the speech normal/careful ratio and the fingerspelling normal/careful ratio. Port (1981) found that speakers' normal(fast)/careful(slow) ratio ranged from 0.65–0.8. Signer s2's normal/careful ratio is solidly in this range at 0.67, but s1's is quite a bit lower at 0.39. Direct comparison is impossible because a segment of fingerspelling is quite different than a segment of speech. This difference in what constitutes a segment could explain this difference in the normal/careful ratio. A theory of what exactly a segment of fingerspelling is is beyond the scope of this study, but would prove invaluable to the further study of fingerspelling.

### 3.3 Differences between word types

We also found a statistically significant difference in transition time between word types ($F = 220.34$, $p < 0.0001$). Figure 5 shows that this difference is small, and probably only isolated to a difference between non-English word transition times and English word transition times. See table 3 for detailed information about the mean, standard deviation, shortest, and longest transition time for each word type.

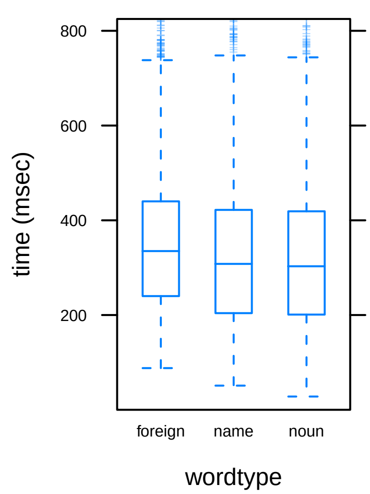



Figure 5: transition time between wordtypes

Taking a closer look, and accounting for what we have already discovered affects transition time, we get

| | mean | sd | shortest | longest |
|---|---|---|---|---|
| foreign | 354.96 | 147.56 | 88 | 967 |
| foreign | 326.92 | 149.18 | 51 | 957 |
| foreign | 326.07 | 153.52 | 28 | 988 |

Table 3: Mean, standard deviation, shortest, and longest transition time for word types, all in msec

figure 6. Here it appears that non-english words have longer transition times, except for signer s1's careful fingerspelling. This is confirmed using TukeyHSD corrected multiple pairwise comparisons. When grouped by signer and speed we cannot reject the null hypothesis that names and nouns come from the same distribution ($p > .18$) for every combination of signer and speed. But we can reject it when comparing non-English words with names and nouns ($p < .0001$) for all combinations, except for signer s1's careful, where we have much lower confidence for nouns and names ($p = .03$ and $p = .003$ respectively). This lower confidence means we can't be sure that we should reject the null hypothesis that the distributions are from the same distribution. But beyond that we see a smaller difference between non-English and English words in both signers in the careful condition, as opposed to the normal condition (see table 4).

| | non-English | English | diff |
|---|---|---|---|
| s1 careful | 496.94 | 481.39 | 15.55 |
| s1 normal | 218.98 | 177.73 | 41.25 |
| s2 careful | 411.84 | 388.25 | 23.59 |
| s2 normal | 293.15 | 253.49 | 39.66 |

Table 4: difference between normal and careful in non-English versus English, all in msec

We can't conclude with any certainty that this is the case, but it appears that what is going on is that unfamiliar words take longer to sign either because of processing, because the letter sequences are unfamiliar and there is a lack of muscle memory for those transitions, or because of some other factor. The reduction in the difference between English and non-English word transition time in careful is probably the result of a ceiling effect produced by the artificial slowness of careful speed. The processing or articulatory difficulty is relieved because the signers are signing so slowly in the careful condition, so there is no slow down on non-English words. Again, detailed information on means, standard deviations, shortest, and longest transition times can be found in table 5.

The difference between English and non-English words seems quite small. One might wonder if this difference is even perceivable. Klatt (1976) surveyed literature (Huggins, 1972; Fujisaki et al., 1975; Klatt and

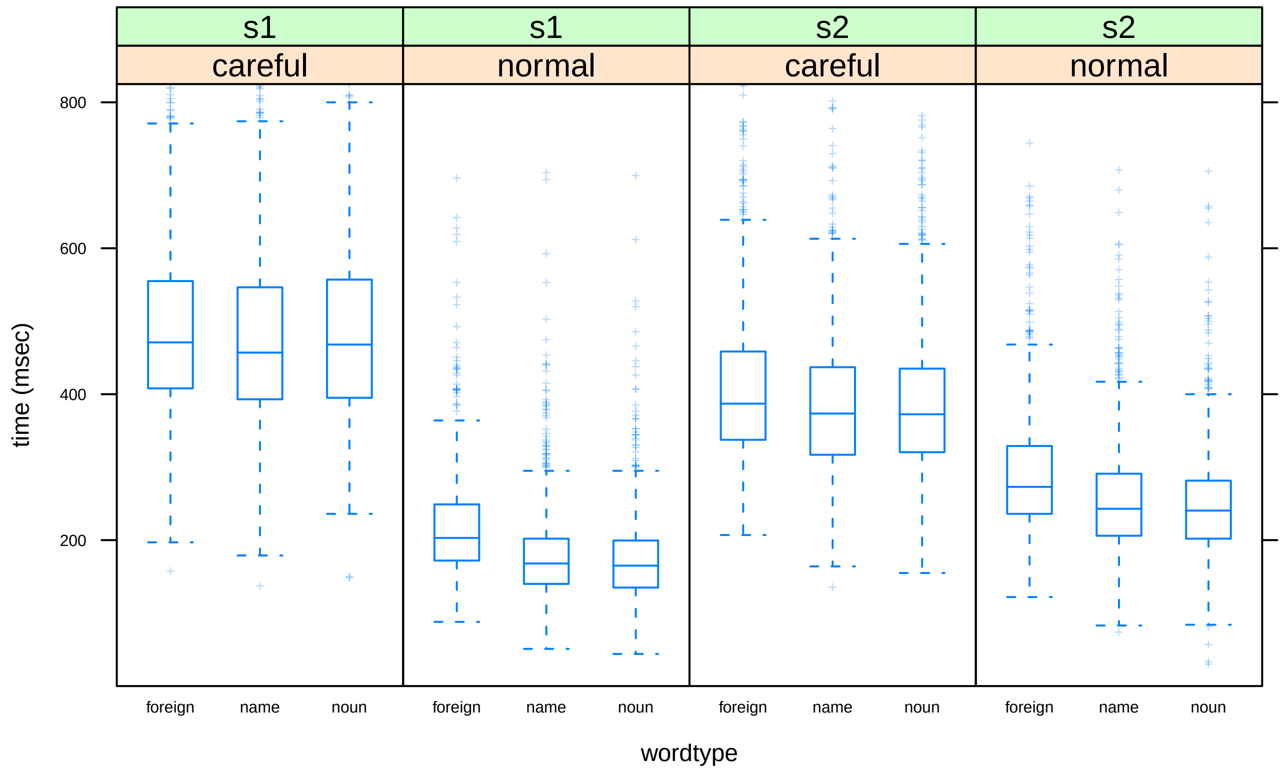


Figure 6: transition time between wordtypes, by speed and signer

| | mean | sd | shortest | longest |
|---|---|---|---|---|
| s1 careful foreign | 496.94 | 126.93 | 156 | 967 |
| s1 careful name | 478.68 | 120.96 | 135 | 953 |
| s1 careful noun | 484.17 | 124.77 | 147 | 988 |
| s1 normal foreign | 218.98 | 74 | 88 | 897 |
| s1 normal name | 179.41 | 66.04 | 51 | 702 |
| s1 normal noun | 176.02 | 66.49 | 44 | 894 |
| s2 careful foreign | 411.84 | 109.55 | 207 | 952 |
| s2 careful name | 386.7 | 102.69 | 134 | 957 |
| s2 careful noun | 389.85 | 105.17 | 155 | 968 |
| s2 normal foreign | 293.15 | 88.21 | 122 | 906 |
| s2 normal name | 257.19 | 79.12 | 72 | 705 |
| s2 normal noun | 249.75 | 76.55 | 28 | 890 |

Table 5: Mean, standard deviation, shortest, and longest transition time for speed-signer-word type combinations, all in msec

Cooper, 1975) discussing just-noticeable difference for segment durations in speech. This survey turns up that speakers notice that speech sounds weird when the segment durations are lengthened by just 20 to 25 msec. While further research is needed confirm that these numbers are accurate in fingerspelling, it is interesting that non-English words are lengthened by more than what we know humans can identify as different in speech.

### 3.4 Individual letters

Now we want to know what information does each letter individually contribute to the transitions times that surround it. Up until now we've been looking at the transition times regardless of what letters are on either side. The time it takes to get from one apogee to the next is determined, in part, by what letters each apogee is. As with all languages there is a massive amount of coarticulation, which can effect apogees further away than just the next one. However, an abstraction to look at what kind of influence a particular letter has on transition time is to look at how long it takes to get from the preceding apogee to the target, and then from the target to the next one. We can group all of the available preceding transitions for a given letter, and all of the available following transitions for a given letter together to see what kind of systematic effect on transition time a particular letter has. Going back to the example of T-A-X-I before (repeated, with the letters filled in, as figure 7), looking at -A- we would add the time between the -T- and the -A- to all the other times that precede -A-S, and we would add the time between the -A- and the -X- to all the other times that follow -A-S.

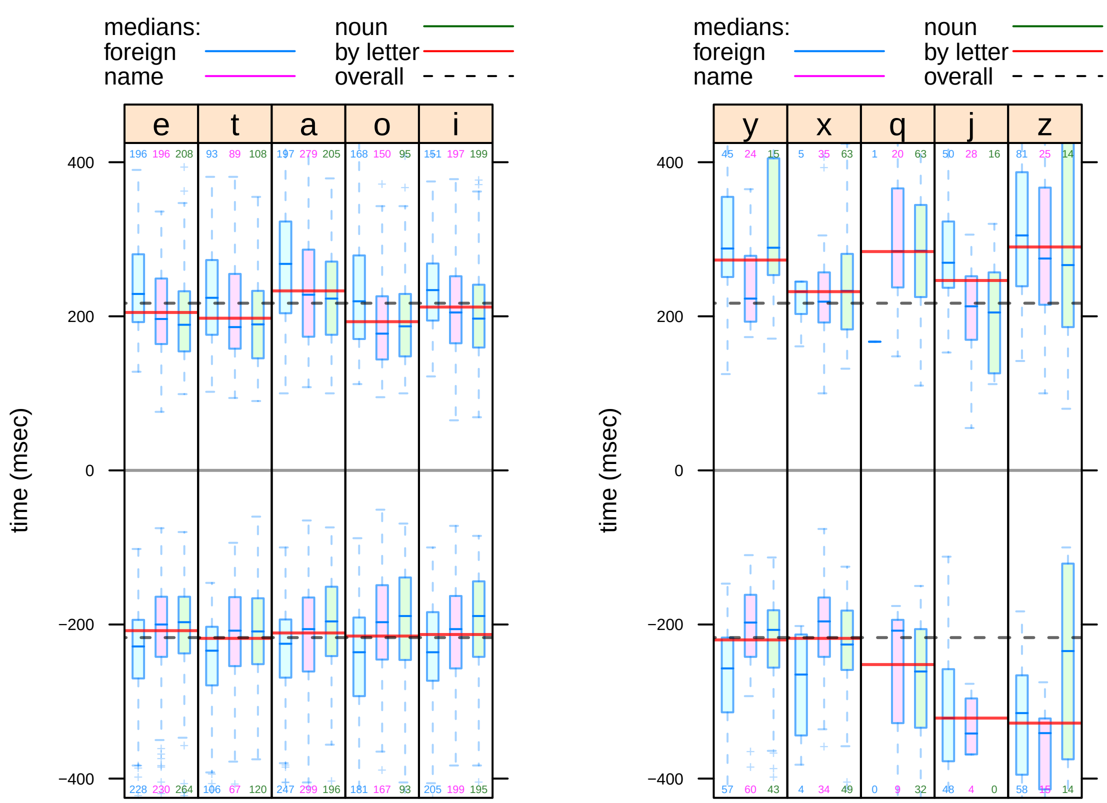
$$T \xrightarrow{\text{100 ms}} A \xrightarrow{\text{174 ms}} X \xrightarrow{\text{183 ms}} I$$

Figure 7: visualization of T-A-X-I (signer: **s1** speed: **normal**)

This abstraction does not perfectly line up with what many would conceptualize as segments. However the period of time surrounding an apogee will be the period of time where the letter of that apogee should have the most influence over what the articulators are doing, and for our purposes serve the function to answer the question how much difference in transition times do individual letters contribute?

We expect that there won't be much of a difference for many letters since much of the ASL fingerspelling literature suggest that native fingerspelling is almost always metronomic; every letter takes the same amount of time to execute (Hanson, 1981; Wilcox, 1992). Indeed, we find that for most letters the transition times stay close to the mean for that particular signer and speed. The only exception to this is that some very low frequency letters (-ZZ-, -Z-, -J-, and -Q-) exhibit much longer transition times than other letters. Figure 8 show the means and deviations of select individual letters split between letters and word types. Total means for each letter across word types are given by the red lines, and a total mean for each speed is given by the dotted grey line. The sample size of letters in each word type is given at the bottom of the plot.

Of the five most frequent letters, none of their means are farther than 25 msecs away from the mean for the speed and, four out of the five are even below the mean (see figure 8). The only one that is above the mean is only 2 msecs above. For this high frequency group of letters the mean is 443.17 msec, and the standard deviation is 137.85 msec. Considering that most letters follow this pattern it's clear that transition time will not be a very good predictor of what letter the segment is alone. The exception to this are a few very low frequency letters like -z-, -j-, -q-, and -x-, which range from 201.42 – 7.26 msec longer than the mean for the speed. For this low frequency group of letters the mean is 575.27 msec, and the standard deviation is 181.21 msec. The low frequency letters are both longer, and show more deviation than the more frequent letters. Although length is not distinctive for most letters, it may prove valuable for identifying these low frequency letters.




(a) for letters -e-, -t-, -o-, -a-, and -i-

(b) for letters -y-, -x-, -q-, -j-, and -z-

Figure 8: transition times based on letter – normal

For the rest of the letters see figure 10. This figure includes all letters, ordered by decreasing frequency in

English,[11]. As with figure 8 above, it shows the means and deviations of individual letters split between speed, and word types. Total means for each letter across word types are given by the red lines, and a total mean for each speed is given by the dotted grey line. The sample size of letters in each word type is given at the bottom of the plot. The longer transition time might be due to some level of signer unfamiliarity with the letters since they are lower frequency, or it could be a perceiver oriented mechanism employed by signers to enhance the communication signal for letters that have an extremely low frequency. The actual mechanism for this longer transition time, is fairly simplistic, notice that the two lowest frequency letters in English (*z*, and *j*) are also the only two letters that have (traditionally been described with) movement in fingerspelling. In the collection of this data we noticed that there are a variety of other letters that seem to also be accompanied by movement, even though they are not classically taught as containing movement. The letters that this movement was noticed on, were almost all low frequency letters: -Q-, -X-, and -Y-. Although these letters weren't invariably accompanied by movement, they often were. Looking at the transition times before and after, we notice that it is exactly these letters that are the longest – not at all surprising if many of them have movement, as well as the obligatory handshape and palm orientation. It takes longer to form a handshape-orientation combination, and then apply a specified movement than it does to just form a handshape-orientation combination.

As for frequency alone, the letter -Y- in fingerspelling seems to buck this trend, being about half way through the frequency rank of letters. But in a list of 2,000 common nouns the letter *y* is the 8th least common letter, rather than the 11th least common letter for all of English. Although more research and corpus development is needed to see exactly what the frequency of letters are for fingerspelling generally, using English nouns as a proxy, it seems that *y* has a lower frequency than in English as a whole.

A few low frequency letters stand out; long transition times are a strong predictor of membership in this restricted class. Because this class is small this information can be exploited more easily to identify these letters. Beyond the functional use of this information, we have discovered that letter articulations for at least three letters not usually described as having movement are more complex than their traditional descriptions. Furthermore, once we factor out the numerous factors discussed above, as well as the ones in the following sections we may find that transition time has more variation than it seems here. Further work developing models that will do this is the next step in our research.

[11]with the letter -ZZ- being given an arbitrary low frequency.

### 3.5 Position and word length

We can now take a look at how position and word length affect transition time between letters. A reasonable hypothesis about word length is the transitions in longer words are shorter than the transitions of shorter words, as a way to somewhat normalize individual word length variation.

We will look at normal first, because it is more naturalistic. Impressionistically (see figures 9, 11, and 12[12]) there isn't very much variation between words of different lengths for signer s1. For signer s2 there seems to be a gradual curve up (transition times getting longer) as words get longer. Using multiple comparisons with TukeyHSD correction we find that there are two groups 3–4 letter words and 5–11 letter words[13] All pairwise comparisons between these two groups are different ($p < .05$), with the exception of 5 letter words and 3 letter words, where we cannot reject the null hypothesis that they are from the same distribution ($p = 0.362$). The only other comparison that shows significance is between 9 letter words and 5 letter words ($p = 0.0106$).

Now turning to careful we see that the distributions are again more spread out. There is much more variation within each speaker in careful than there is in normal. Again for signer s1 there isn't much of a trend, although there are a few unpatterned comparisons that are statistically significant. For signer s2 there is much less of a pattern than there was for normal, but again there appears to be a slight increase in transition time as words get longer. There is more noise in these comparisons, but two groups seem to appear 3–7 letter words and 8–11 letter words ($p < .05$).[14]

To look at how position affects transition time length let's look at only speaker s1, normal speed. See figures 11–12 at the end for plots of the transition time in each position separated by word length, signer, and speed. While the overall mean transition time for the lengths of words didn't seem to change very much from one length to the next, we see different patterns within each length based on the position of the letter. (figure 9)

We see words for lengths 3–10[15] In the shorter words the first transition is long, but after that the transitions are fairly stable with little deviation from a slight downward slope, resulting in shorter transition times. When the words get longer (especially 9 letter words, but some others as well), however, there are a few points (transitions 4 — and to a much smaller extent some later transitions) that show much longer transition times

[12] Because of the way that we verified and coded this data there are some instances where words end up with more (and sometimes less) letters fingerspelled than the word contained. This is why some of these plots include an extra letter at the end. This noise is fairly small and so it doesn't skew the data too much in either direction.

[13] The transition times for words that are longer than 11 letters show a lot of variation. This could be evidence that longer words stress the memory systems responsible for managing this kind of information.

[14] Again words longer than 11 letters show a lot of variation.

[15] Again, lengths >10 show extreme variation that just muddied up the patterns here.

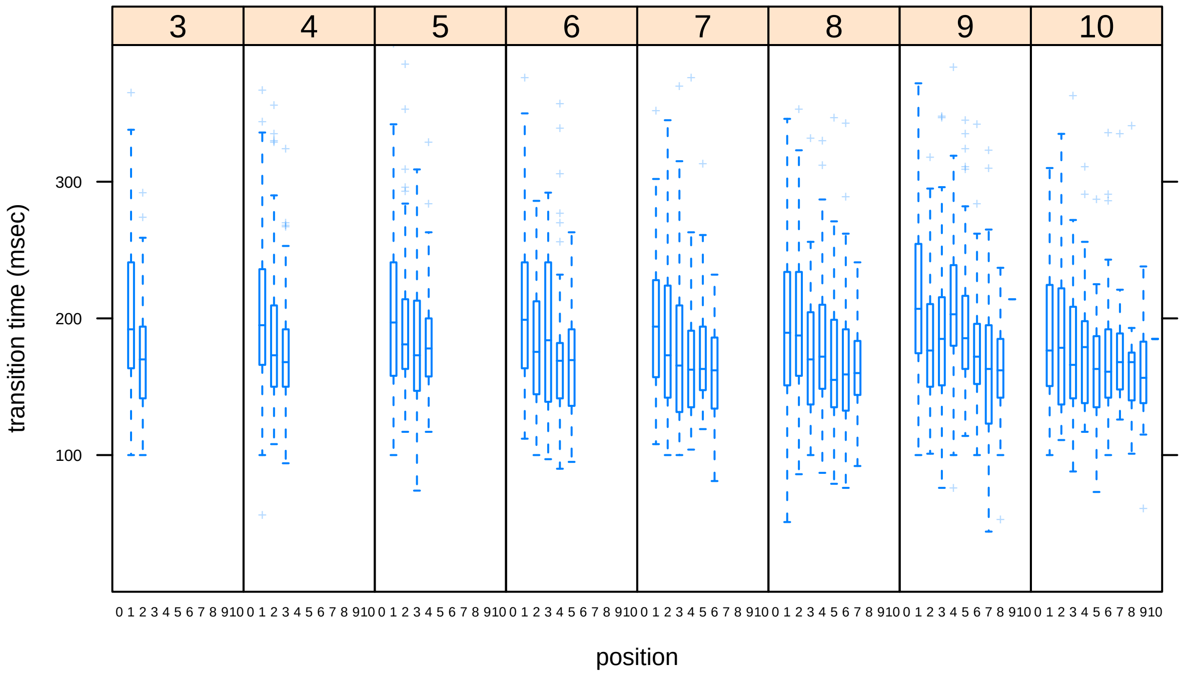



Figure 9: transition times based on transition position, s1, normal

than the general downward slope as the word goes on.

This jump in transition time for longer words can be given many explanations. One possibility is that there is some sort of memory window of 4–6 letters that signers are able to keep in their head at once, and when confronted with a word that is longer than this they need to chunk their production together to allow them to remember what the next few letters are. This chunking might also be explained by articulatory planning, where the articulatory planning process groups 4–6 letters together. This chunking might also make fingerspelling more like ASL phonologically: 4 letters is approximately 3 movements, which is approximately the number of movements it takes to get to the first position, execute movement, and then move on to the next sign for standard lexical signs in ASL. Finally, one more possibility is that this is actually a perceiver oriented mechanism, to allow time for the perceiver to process a few letters before the signer continues on. There is much literature on accommodation, but in fingerspelling specifically there is at least one other example of this perceiver oriented enhancement of the signal, that is epenthetic flairs between two similar handshapes. An example would be the word A-T, both handshapes are very closed, so often a signer will open their hand up much more than is strictly necessary to move to thumb from the side of the hand to between the index and middle finger. This flair serves to differentiate the handshapes for the person viewing them, and actually makes the articulation more energy intensive. These have been documented by Brentari (1998).

This chunking is mostly gone in the careful condition, where the transition times either decrease predictably, or are completely erratic. (figures 11b and 12b) This is another indication that there is a ceiling effect when signers sign carefully. The absence of a systematic pattern for positions also indicates in another way that the careful speed is not as natural as the normal speed where we see a much more regular pattern.

The systematic alteration of transition time based on the position in the word will distort the transition times of individual letters. Further work is needed to completely model this interaction. This is an important next step in the study of the phonetics of fingerspelling.

Further data is needed to collect enough data in all of these conditions to see just how much of an affect this really is, and to what extent it is generalizable across multiple signers. But with the two signers we have so far, it seems that there is some kind of chunking. We also need to look at the data more carefully to see if we can tease apart evidence for the different explanations given above.

## 4 Implications and future research

There are many directions that this research can be pursued further. Generally, we need to collect more data from more signers in order to confirm that many of these trends are in fact generalizable beyond these two specific, related signers. We also need many more and more diverse instances of words. We do not have any data for some combinations of word type, and letter. Beyond that, we don't have instances of each letter in each position in each word type. We are even more so lacking bigram data for all possible bigrams. Although we have nearly 12,000 individual apogees, it would take at least 13,122 apogees to get all of the possible bigrams for 27 letters (-A-–-Z-, and -ZZ-), in 2 different speeds, 3 different word types, and in 3 different contexts: beginning, medial, end (a huge simplification considering the chunking we saw in section 3.5). We are in the planning stages of collecting more data from more signers, and although we don't have any reason to assume that the results reported here are specific to these two signers, we can't be sure that they won't change with the addition of more data. All of this will continue to facilitate the larger project of automatic fingerspelling recognition.

Because of the sparsity of data, and the numerous factors that seem to affect transition time more complicated statistical models should be developed that allow for multiple comparisons across many predictors.

Models of this type have been used in the social sciences for a while, and increasingly in linguistics specifically, using similar data to what we have here with great success.

The discussion about the effect of position on transition time in section 3.5 brings up an area that needs a lot of future research. It shows that there is a prosody to fingerspelling that needs further investigation. We have intentionally avoided trying to analyze this using spoken language terms because that would require a much more developed theory of fingerspelling prosody than can be accomplished in the current study. But this is a field of exploration that is critical to continue.

As was alluded to in the discussion about the importance of transition time, other phonetic features of fingerspelling must be identified and quantified. This will allow us to start to put all of the pieces together that we need to have a more full understanding of exactly what is going on when one fingerspells, or even when one perceives fingerspelling. This will also contribute invaluable information to automatic fingerspelling recognition.

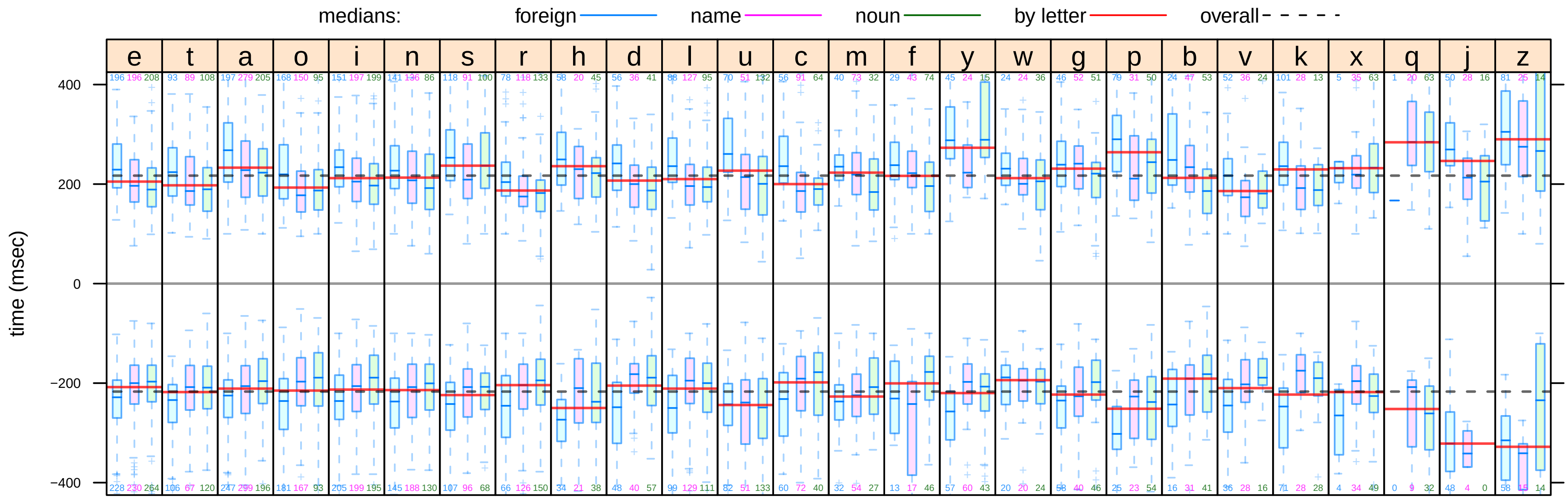
Additionally a metric to measure the movement of specific letters in the data that has been collected should be developed and applied to measure exactly how much movement is happening in letters like -Q-, -X-, and -Y-. This metric of movement could integrate into a system of uniform information density (Jaeger, 2006; Levy and Jaeger, 2007), where signers explicitly enhance a signal for letters that they know are low frequency.

With over 3 hours of fingerspelling we have a vast amount of data that needs further exploration. One area that has progressed in parallel to this research on transition time is an exploration of errors in fingerspelling. Although it would be inappropriate to delve too deeply, one interesting phenomenon that doesn't seem to have any analogue in speech is the discovery of word initial anomalies multiple times in the production of the fingerspelled forms that we elicited. These anomalies almost always had a fully formed handshape, and occurred well before the first letter. The distance between the word initial anomaly and the apogee representing the actual beginning of the word was much greater than the 250 msec that the first transitions usually lasted (Rizzo, 2010). Accounting for this, and many other errors should make this transition time data more accurate. At the same time, understanding what these errors and anomalies look like will inform an automated recognition system what ought to be disregarded.

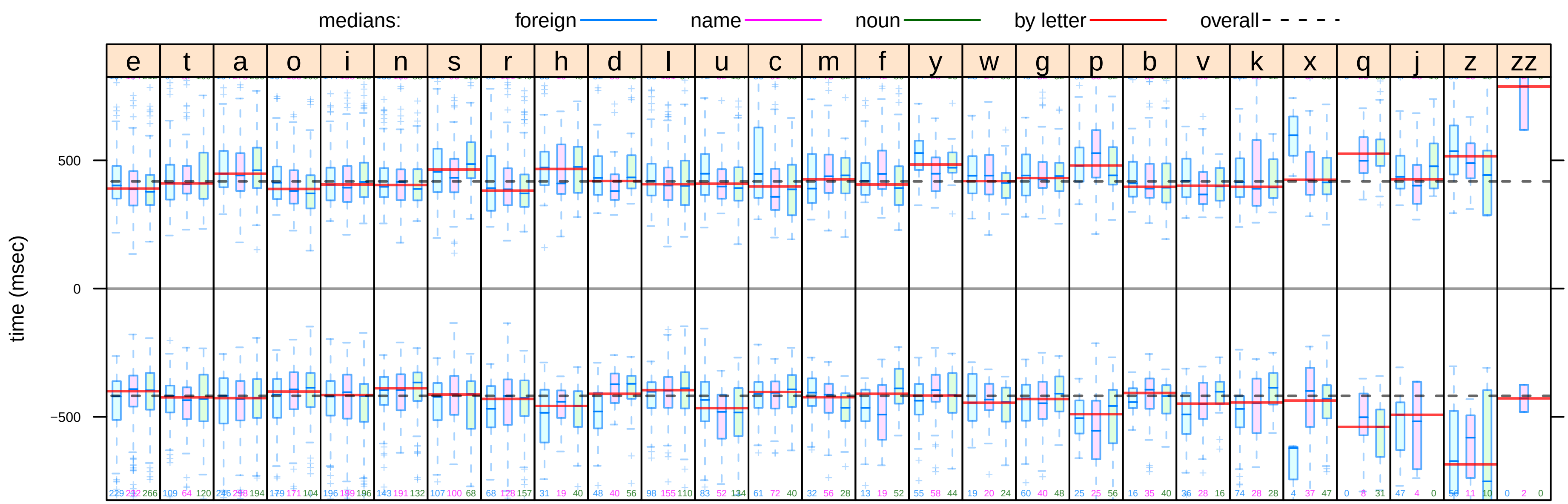
# 5 Conclusion

Fingerspelling is an understudied aspect of ASL. In the course of a larger project working on fingerspelling recognition we have recorded over 3 hours of native signers fingerspelling a variety of words. The apogees for each letter were hand coded, and from that we have extracted transition time information. Transition time relates to segment duration, which has proved fruitful in speech recognition research. We have confirmed a variety of hypotheses about transition time in fingerspelling that conform to our expectations based on previous linguistic research on transition time. and segment duration. When we asked signers to fingerspell carefully, the transition time between letter increased. The two signers fingerspell at different rates, and interpret careful and normal rates differently. Unfamiliar, non-English words are fingerspelled with longer transition times than familiar, English words. We have also found that when fingerspelling longer words there is some evidence that signers chunk letters together in groups of 3–5 letters, shown by longer transition times every 3-5 letters. This effect is mitigated at careful speed, which could suggest that the chunking is due to a memory limitation, and articulatory limitation, or possibly perceiver oriented facilitation. We have also found evidence that the class of letters described as having movement probably needs to be expanded to include -Q-, -X-, and -Y-.

The development of this data will have implications for the continued study of fingerspelling, and the addition of more speakers, and more data will only add to this. Although neglected in the past the phonetics ASL fingerspelling provides a rich source of information that is helpful in automated recognition, as well as to linguistics as a whole.

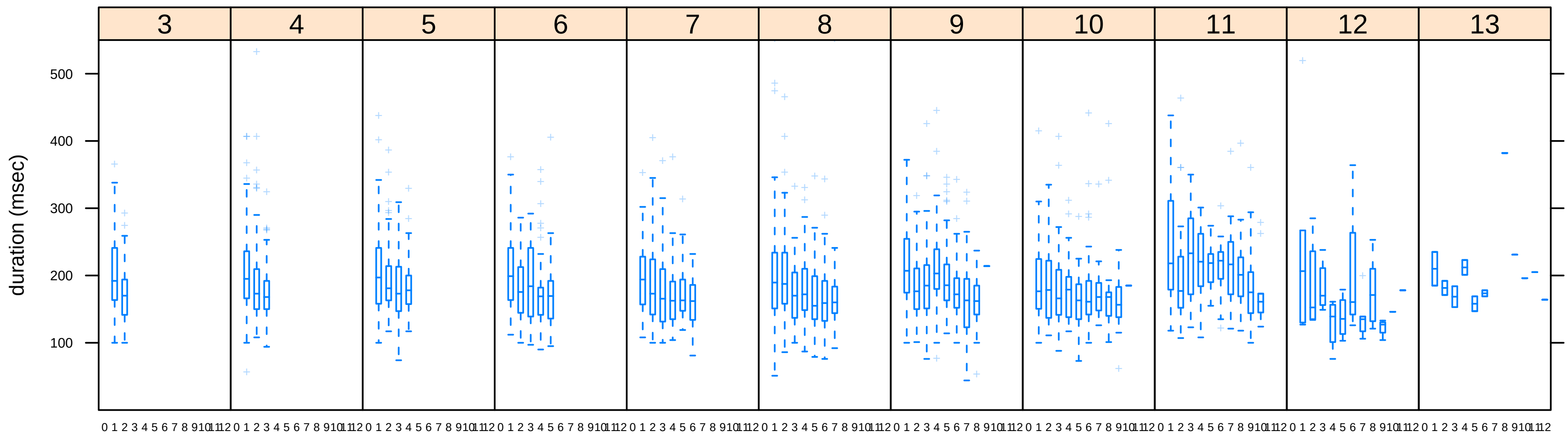



(a) normal speed

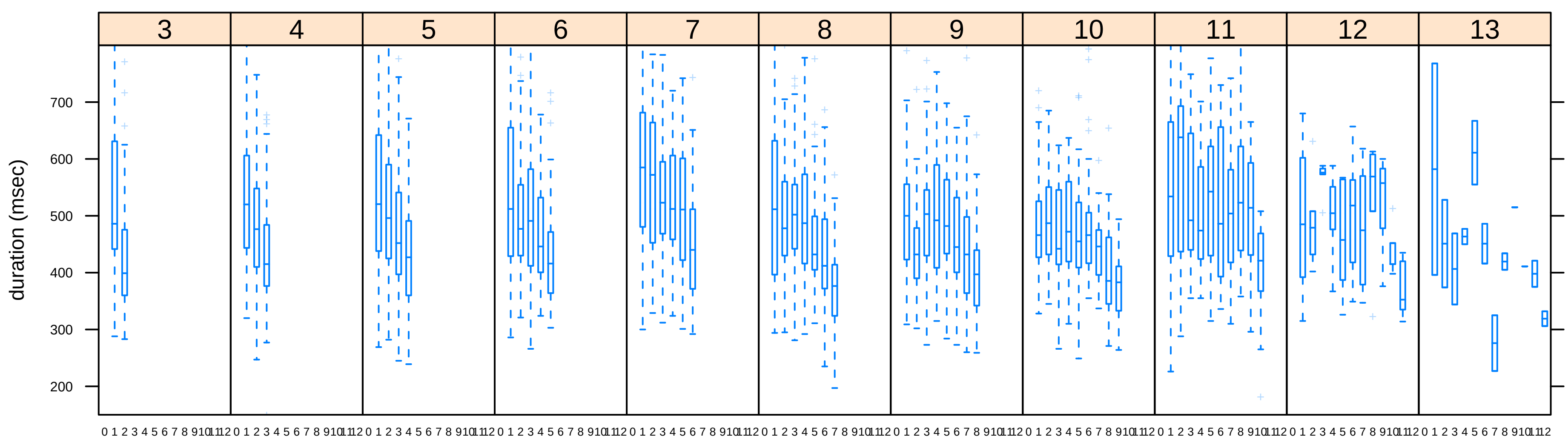



(b) careful speed

Figure 10: individual letter transition times– s1

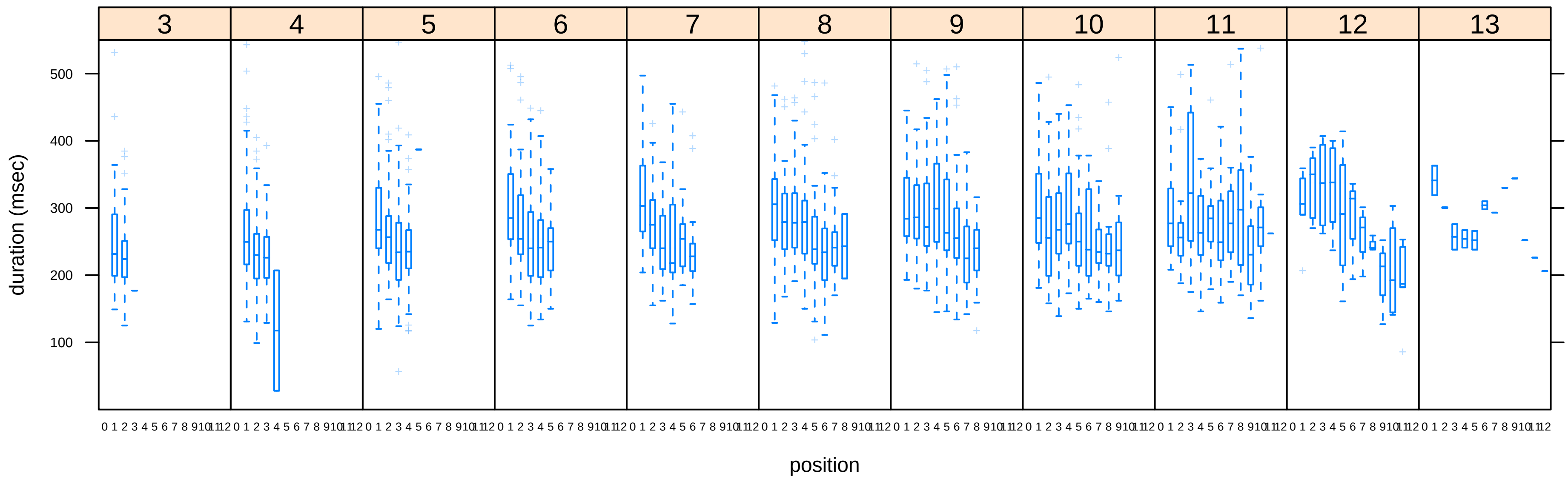


(a) normal speed

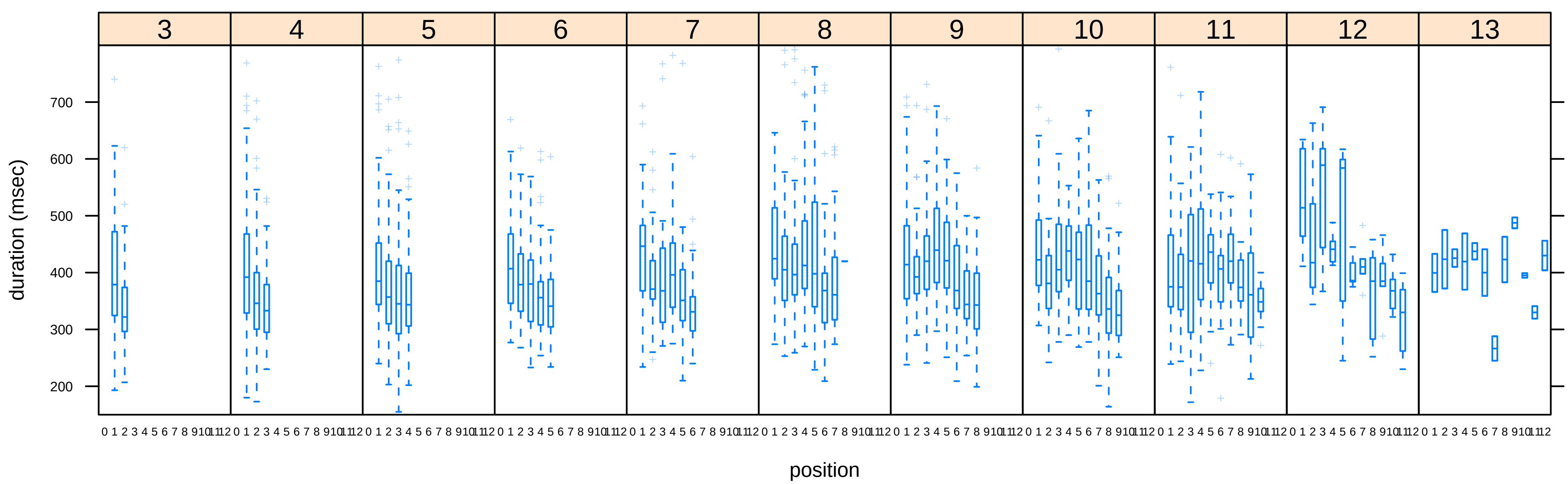


(b) careful speed

Figure 11: durations based on position, grouped by word length (following) – s1



(a) normal speed



(b) careful speed

Figure 12: durations based on position, grouped by word length (following) – s2

# A Wordlist

## A.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. enrique
27. everglades
28. excel
29. exxon
30. felix
31. finn
32. flossmoor
33. francesca
34. franklin
35. fred
36. gary
37. gayle
38. george
39. giordano
40. greg
41. himalaya
42. inglewood
43. izzy
44. jacqueline
45. jason
46. jimmy
47. joe
48. john
49. josh
50. kate
51. kelly
52. leo
53. lexus
54. libya
55. mary
56. matt
57. mauritania
58. mediterranean
59. mexico
60. mia
61. mississippi
62. mongolia
63. moscow
64. naomi
65. naperville
66. nic
67. owen
68. pam
69. paraguay
70. quentin
71. quincy
72. quotation
73. rangerover
74. rita
75. russ
76. sam
77. sara
78. scotland
79. skokie
80. tallahassee
81. tanzania
82. tiffany
83. tobias
84. toby
85. tokyo
86. tom
87. venezuela
88. venice
89. viv
90. will
91. william
92. xavier
93. xerox
94. yellowstone
95. yosemite
96. zack
97. zoe

## A.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

## A.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

## References


Brentari, D. (1998). *A prosodic model of sign language phonology*. The MIT Press.

Brentari, D. and Padden, C. (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, pages 87–119. Lawrence Erlbaum, Mahwah, NJ.

Chung, G. and Seneff, S. (1999). A hierarchical duration model for speech recognition based on the angie framework1. *Speech Communication*, 27(2):113–134.

Cormier, K., Schembri, A., and Tyrone, M. E. (2008). One hand or two? nativisation of fingerspelling in ASL and BANZSL. *Sign Language & Linguistics*, 11(1):3–44.

Emmorey, K., Bassett, A., and Petrich, J. (2010). Processing orthographic structure: Associations between print and fingerspelling. Poster at 10th Theoretical Issues in Sign Language Research conference, West Lafayette, Indiana.

Fujisaki, H., Nakamura, K., and Imoto, T. (1975). Auditory perception of duration of speech and non-speech stimuli. *Fant, Tatham, Auditory analysis and perception of speech*, pages 197–219.

Hanson, V. (1981). When a word is not the sum of its letters: Fingerspelling and spelling. In Caccamise, F., Garretson, M., and Bellugi, U., editors, *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*, pages 176–85.

Huggins, A. (1972). Just noticeable differences for segment duration in natural speech. *The Journal of the Acoustical Society of America*, 51:1270.

Jaeger, T. (2006). *Redundancy and syntactic reduction in spontaneous speech*. PhD thesis, Stanford University.

Klatt, D. (1976). Linguistic uses of segmental duration in english: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, 59:1208.

Klatt, D. and Cooper, W. (1975). Perception of segment duration in sentence contexts. *Structure and process in speech perception*, pages 69–86.

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *The Journal of the Acoustical Society of America*, 51:2018.

Levinson, S. (1986). Continuously variable duration hidden markov models for automatic speech recognition. *Computer Speech & Language*, 1(1):29–45.

Levy, R. and Jaeger, T. (2007). Speakers optimize information density through syntactic reduction. *Advances in neural information processing systems*, 19:849.

Livescu, K. and Glass, J. (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.

Mitchell, R., Young, T., Bachleda, B., and Karchmer, M. (2006). How many people use ASL in the united states? why estimates need updating. *Sign Language Studies*, 6(3):306–356.

Oller, D. (1973). The effect of position in utterance on speech segment duration in english. *The Journal of the Acoustical Society of America*, 54:1235.

Padden, C. (1991). *Theoretical issues in sign language research*, chapter The Acquisition of Fingerspelling by Deaf Children, pages 191–210. The University of Chicago press.

Padden, C. and Gunsauls, D. C. (2003). How the alphabet came to be used in a sign language. *Sign Language Studies*, 4:10–33.

Padden, C. and Le Master, B. (1985). An alphabet on hand: The acquisition of fingerspelling in deaf children. *Sign Language Studies*, 47:161–172.

Peterson, G. and Lehiste, I. (1960). Duration of syllable nuclei in english. *The Journal of the Acoustical Society of America*, 32:693.

Port, R. (1981). Linguistic timing factors in combination. *The Journal of the Acoustical Society of America*, 69:262.

Quinto-Pozos, D. (2010). Rates of fingerspelling in american sign language. Poster at 10th Theoretical Issues in Sign Language Research conference, West Lafayette, Indiana.

Rizzo, S. (2010). Performance errors in the manual modality: the case of ASL fingerspelling. Ms.

Tyrone, M., Kegl, J., and Poizner, H. (1999). Interarticulator co-ordination in deaf signers with parkinson's disease. *Neuropsychologia*, 37(11):1271–1283.

Wilcox, S. (1992). *The phonetics of fingerspelling*. John Benjamins Publishing Company.