A New N-gram Analytics Tool in ELAN and its Application to Improve Automatic Fingerspelling Generation

Souad Baowidan, Natalie Ningshan Guo, Sarah Johnson, Robyn Moncrief, Larwan Berke
DePaul University
School of Computer Science
Chicago IL
sbaowidan@gmail.com, guo.ningshan@gmail.com, sarahej101@gmail.com,
rkelley5@mail.depaul.edu, larwan.berke@rit.edu

Abstract

A new extension to ELAN offers expanded n-gram analysis tools including improved search capabilities and an extensive library of statistical measures of association for n-grams. This paper presents an overview of the new tools and a case study in American Sign Language synthesis that exploits these capabilities for generating more naturally-appearing fingerspelling for use in self-study applications. The new extension provides a time-saving convenience for language researchers using ELAN.

1 Introduction

In 2004, researchers at The Language Archive introduced ELAN ("Eudico Language Annotator") [19], an annotation tool that features synchronized video, audio, and annotations. Its major applications include gesture studies, documentation of endangered languages, and analysis of sign languages [8]. In subsequent years, researchers added support for synchronized motion capture [11], broadened the scope of video support including enhanced time accuracy [35], and explored automated techniques to reduce the time required to annotate media [13] [4]. In 2008, new extensions allowed users to create references from annotation systems defined in the central ISO Data Category Repository (DCR) [24]. The goal is to continue to foster greater data sharing among language researchers.

Part of ELAN's appeal derives from its powerful and diverse search tools [25]. These provide an immense gamut of search granularity, ranging from finding individual annotations in local files, to accessing web-based corpora. Users can also search for n-grams within a single tier of annotation codes or for phenomena that co-occur on multiple tiers [10]. The two main formats for search results are the concordance and frequency views. In either view, users can elect to "Show hit in transcription", which cues the linked media and annotation to the position where the hit occurs in the ELAN annotation file.

Statistical services available include frequency counts for search queries. The "Statistics for multiple files" search also includes basic descriptive statistics for hits within a tier, including the minimum, maximum, mean, and median for a duration. For analysis techniques not included in the services, ELAN provides the capability to export raw search results for further study.

The rest of this paper is organized as follows: Section 2 focuses on ELAN and presents a new software module that implements several n-gram statistical tools. Sections 3 to 6 look at the applications of a case study result of an n-gram analysis for the improvement of technology for acquisition of fingerspelling recognition skills. Section 3 reviews the properties of fingerspelling that contribute to the challenges for sign language learners; Section 4 discusses the shortcomings of current technologies for self study; Section 5 describes a new corpus of fingerspelling examples; Section 6 describes the use of the new tool to examine certain properties of fingerspelling with the goal of incorporating them into an improved technology, and finally Section 7 concludes with results and future work.

2. A new extension to ELAN: an n-gram analysis tool

The new ELAN extension presented in this section aims to increase the speed and ease of n-gram analysis for users. Similar to [18], combining the built-in tools and the environment of ELAN's interface and data sharing capabilities increases the accessibility of n-gram analysis for users. Previously, performing an n-gram analysis on a corpus required exporting data from ELAN and performing the analysis using a separate process. Reformatting the exported data and/or developing the methods for computing n-gram analyses can be time consuming. In order to address this, the n-gram analysis extension was created for ELAN and is accessible through the "Multiple File Processing" entry in the File menu. What follows is a description of this new capability [7].

The n-gram analysis is located in the "Multiple File Processing" submenu of the File menu (Figure 1).

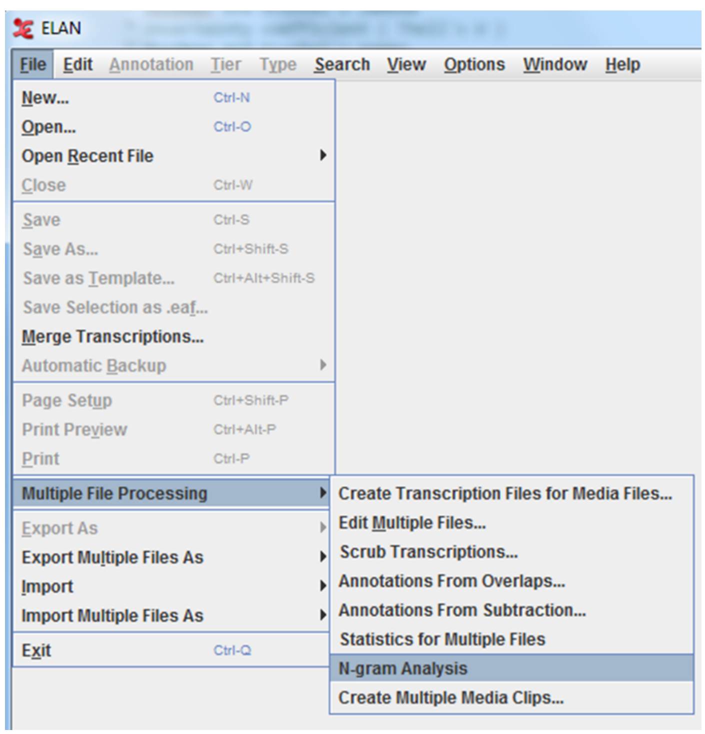


Figure 1: Location of the n-gram analysis in ELAN.

After selecting the n-gram analysis menu item, the user will see a new dialog window pop up that contains the various options for the search (Figure 2).

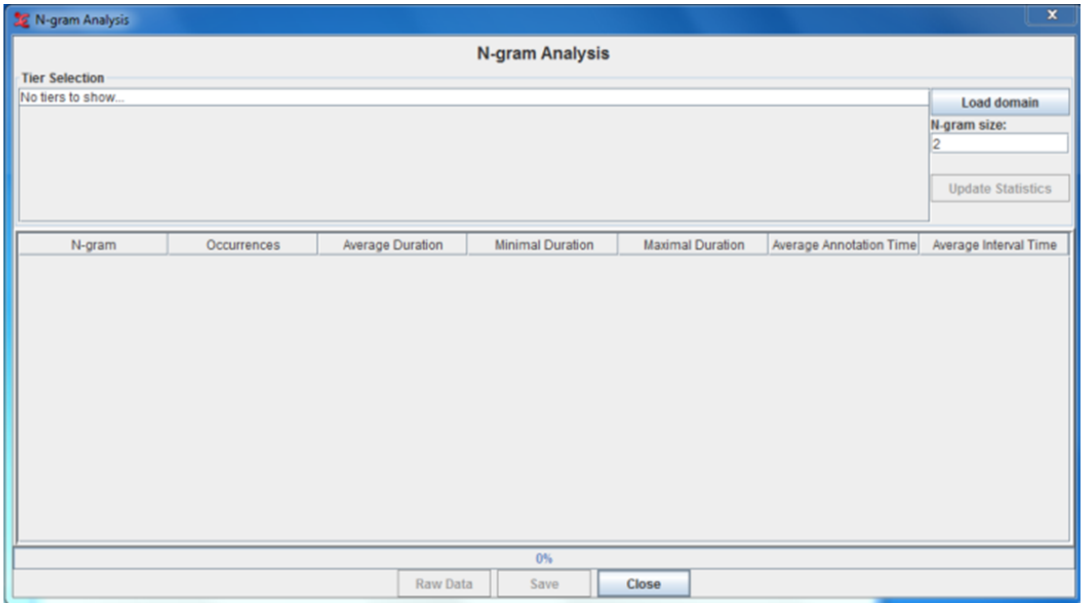


Figure 2: Main n-gram analysis window.

The first step is to select the search domain via the standard ELAN "Load domain" window where the user can specify a list of files or directories. Once the domain is designated, the dialog will display a list of tiers contained in the domain (Figure 3). This functionality assumes that all annotation files in the domain have an identical set of tiers. The software then loads the first file in the domain to extract the tiers and then displays the results in the window.

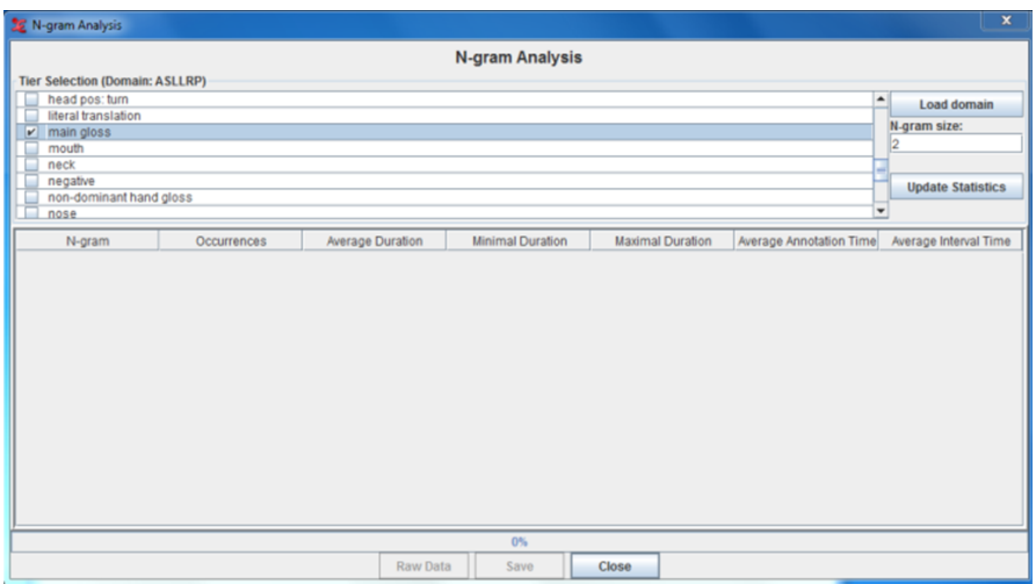


Figure 3: N-gram analysis window displaying possible tiers for searching.

The user then enters the n-gram size in the textbox. The software can handle a n-gram of any size, however the contingency table analysis can only be performed on bigrams. With a n-gram size entered, the user can click the "Update Statistics" button to start the search. The annotations are extracted from the files, the n-grams are created from them, and the results are finally grouped by n-gram for statistical analysis.

When the search is done, a report window will pop up displaying some search-related statistics. This can be used to check the integrity of the search. A sample is seen in Figure 4.

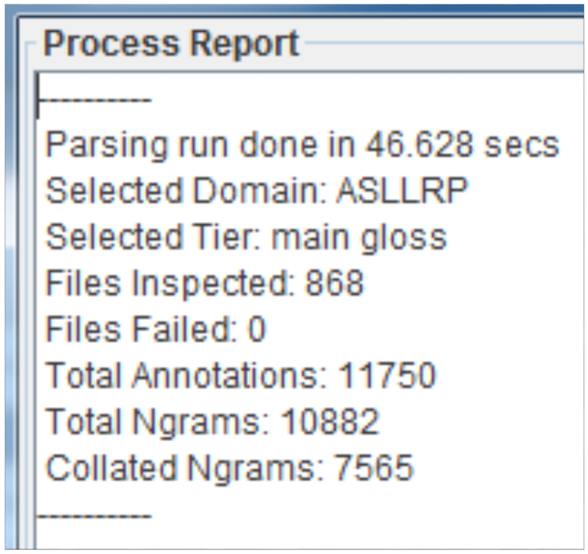
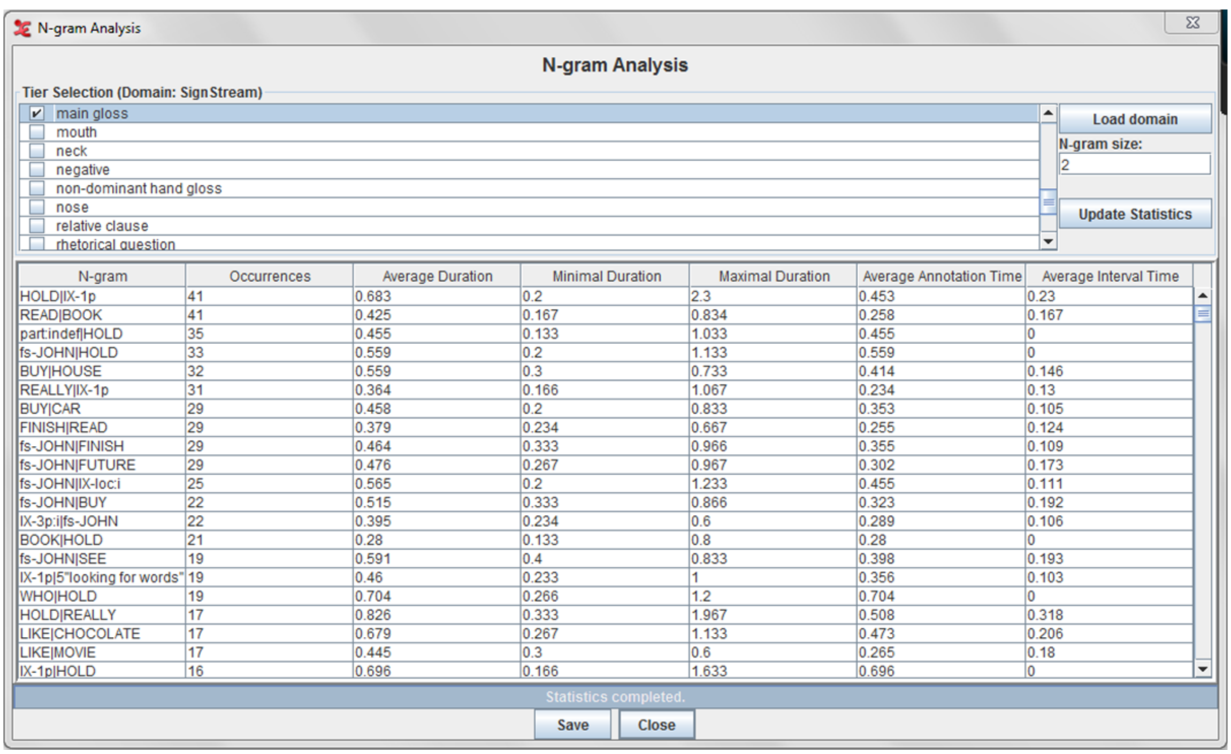


Figure 4: N-gram analysis report window.

Figure 5 is a screenshot of the user interface. To make the new extension easy to use, we designed the interface to be as consistent as possible with other ELAN search dialogs.



N-gram	Occurrences	Average Duration	Minimal Duration	Maximal Duration	Average Annotation Time	Average Interval Time
HOLD|IX-1p	41	0.683	0.2	2.3	0.453	0.23
READ|BOOK	41	0.425	0.167	0.834	0.258	0.167
part:indef|HOLD	35	0.455	0.133	1.033	0.455	0
fs-JOHN|HOLD	33	0.559	0.2	1.133	0.559	0
BUY|HOUSE	32	0.559	0.3	0.733	0.414	0.146
REALLY|IX-1p	31	0.364	0.166	1.067	0.234	0.13
BUY|CAR	29	0.458	0.2	0.833	0.353	0.105
FINISH|READ	29	0.379	0.234	0.667	0.255	0.124
fs-JOHN|FINISH	29	0.464	0.333	0.966	0.355	0.109
fs-JOHN|FUTURE	29	0.476	0.267	0.967	0.302	0.173
fs-JOHN|IX-loc:i	25	0.565	0.2	1.233	0.455	0.111
fs-JOHN|BUY	22	0.515	0.333	0.866	0.323	0.192
IX-3p:i|fs-JOHN	22	0.395	0.234	0.6	0.289	0.106
BOOK|HOLD	21	0.28	0.133	0.8	0.28	0
fs-JOHN|SEE	19	0.591	0.4	0.833	0.398	0.193
IX-1p|5"looking for words"	19	0.46	0.233	1	0.356	0.103
WHO|HOLD	19	0.704	0.266	1.2	0.704	0
HOLD|REALLY	17	0.826	0.333	1.967	0.508	0.318
LIKE|CHOCOLATE	17	0.679	0.267	1.133	0.473	0.206
LIKE|MOVIE	17	0.445	0.3	0.6	0.265	0.18
IX-1p|HOLD	16	0.696	0.166	1.633	0.696	0

Figure 5: Search results from new n-gram analysis. Not all columns are visible.

Similar to the existing n-gram search, users can designate a search domain. However, the new extension augments the n-gram search results by adding the n-gram measures of association as described in [5]. These are listed in Table 1. The extension allows users to choose the metrics most appropriate to their work while still offering the option to export the results as a tab-delimited file for further study.

Chi-squared	Dice coefficient	Jaccard coefficient
Fisher exact two tailed	Fisher exact left sided	Fisher exact right sided
Phi coefficient	Pointwise mutual information	Log-likelihood
True mutual information	Odds ratio	Poisson-Stirling measure
	T-score	

Table 1: N-gram metrics of association

The new n-gram tools have several promising applications for improving sign language synthesis such as fingerspelling generation. The following sections discuss the motivation for generating fingerspelling and present an analysis that uses the new tools to inform a strategy for improving the current state of the art.

3 Uses and properties of fingerspelling

Fingerspelling is the process of spelling words using signs that represent individual letters of a language alphabet [28]. Such signs are known collectively as a *manual alphabet*. Fingerspelling is a tertiary form of communication because it is a representation of writing (a secondary form), which is itself a representation of speech (a primary form).

Fingerspelling is an integral part of American Sign Language (ASL) and is a necessary skill for complete communication in sign [6]. It is used to spell proper nouns, technical terms that lack a universally-accepted sign, places without a sign name and loan words [31]. It can also convey words from a spoken language that a signer does not know in the signed language, and can clarify signs unknown to others. Signed communication systems based on English such as Signed English (SE), Signing Exact English (SEE) and Pidgin Signed English (PSE) also rely on fingerspelling as an essential component of communication [26].

Padden found that on average, fingerspelling makes up approximately 6 percent of the signs produced in everyday ASL conversations, but that in certain contexts, fingerspelling can comprise as much as 12 percent of the signs present in a discourse [21].

Sign Language learners find fingerspelling receptive skills to be one of the most difficult aspects of sign language to master. In interpreter training programs, it is the first skill taught, but the last skill mastered [14]. [23] notes that "hearing people who are learning ASL as adults tend to have great difficulty in correctly recognizing fingerspelled words" (p. 19). There are two key contributing factors. The first is the marked difference between the handshapes that actually appear in fingerspelling production and the idealized manual letters shown in textbooks. The second is the lack of practice opportunities which will be discussed in the following section.

Properties of fingerspelling production that contribute to the challenge of fingerspelling recognition include 1) speed of production, 2) the motion of the transitions between letters and 3) the precision in forming individual letters in a word.

The speed of production can vary depending on the context of the fingerspelling. In *careful fingerspelling,* fluent signers produce fingerspelled letters at a rate of four per second. In *rapid fingerspelling,* production speed can rise to a rate of six letters per second [22]. This is in contrast to a sign in ASL, which has at most two hand shapes on the dominant hand [6]. Thus, a person observing fingerspelling needs to comprehend a larger number of handshapes being produced at a faster rate.

The transitions between fingerspelled letters also pose challenges to fingerspelling reception. Fingerspelling is more than the production of a sequence of static hand configurations. Studies as early as 1971 [37] suggest that people fluent in ASL do not read individual letters, but rather the total pattern of the motion. Particularly in rapid fingerspelling, it is a smoothly-flowing motion that does not come to rest until the last letter. In [2] Akamatsu called this the "motion envelope' [2]. In [34] Wilcox posits that learning to fingerspell involves learning both the static hand configurations and the set of possible transitions. He created a model of *targets* and *transitions* suggesting that fingerspelling can be seen as a series of *movements.*

Lastly, discerning the individual letters in a fingerspelled word also depends on the degree to which any instantaneous hand configuration in a fingerspelled word will match one of the 26 canonical fingerspelled letters. In *careful fingerspelling* the signs representing individual letters are "produced fully and completely" [22]. Careful fingerspelling typically occurs when a word first appears in a discourse. It also occurs in response to such questions as "What is the English word for _______" or "What is your name?". When a word appears in subsequent occurrences, a signer will spell it more rapidly. *Rapid fingerspelling* also occurs in informal settings, and appears more frequently than its careful counterpart. In rapid fingerspelling, the hand movement is a smoothly-flowing organic whole.

While forming the movement comprising a fingerspelled word, the individual handshapes can influence each other in a manner similar to how spoken words can [34]. The effects of *coarticulation* can cause a blending of one fingerspelled letter into the next so that the forms in the fingerspelled word differ from the canonical forms of the fingerspelled letters [3].

The study of coarticulation and compression processes that occur in rapid fingerspelling have been an area of active research. In 1978, Battison examined the process of how a fingerspelled word becomes a loan word [7]. From interviews with nineteen prelingually deaf informants, he identified a total of 40 fingerspelled words which generally became accepted as loan words in ASL. The words ranged in length from two to five letters. He found nine separate categories of potential change. These included:

1) deletion of a fingerspelled letter
2) changes when comparing the produced fingerspelled letter to the idealized fingerspelled letter in:
- location,
- handshape
- movement
- orientation

3) reduplication of a movement
4) addition of a second hand when producing the loan sign
5) morphological involvement such as inflection of the loan sign to show the addition of grammatical information
6) change in the semantics where the loan sign now has a meaning substantially different from the original fingerspelled word

One of the first changes that typically occurs is the deletion of a letter or letters in the fingerspelled word. An example is the deletion of both medial letters from the fingerspelled B-A-N-K as compared to the loan word #BK (bank).

Jerde [16] studied coarticulation as a question of assimilation or dissimilation between hand configurations in varied series of fingerspelled letters [16]. He recruited four participants fluent in ASL from an interpreter service. Each participant donned a Cyberglove before producing a series of 40 fingerspelled sequences. Each sequence was either 1) a English word, 2) a pronounceable non-word or 3) a nonpronounceable, non-word. All of the fingerspelled sequences contained either the letter string I-S-C or the letter string N-T-R. After the three-letter string was a vowel. He then examined the velocity profiles and movement times of individual joints while producing these two letter strings. He found that in the index and middle fingers, the proximal interphalangeal joint showed dissimilation. He posits that the dissimilation may serve to enhance visual discrimination among handshapes of fingerspelled letters.

To search for patterns of anticipation and perseveration in fingerspelling, Channer [9] chose ten words based on their likelihood to exhibit these behaviors, then recorded five hearing ASL signers who each fingerspelled the words. She found that anticipation occurred more often than perseveration and that coarticulation occurred more often in medial letters of a word. Additionally, she discovered that coarticulation occurred very frequently, being exhibited in some form in 53 percent of all fingerspelled letters in the study. Deletion was also most prevalent in medial locations, occurring in five percent of all medially-located fingerspelled letters.

Wager [33] found different rates of occurrence of coarticulation and deletion in a recorded address by a native Deaf signer. While searching for occurrences of careful fingerspelling and rapid fingerspelling, she identified 45 fingerspelled words. Although the discourse was a public address in a formal register, nearly half of the fingerspelled words displayed characteristics of rapid fingerspelling. Among her measurements was a *coarticulation index* metric, which consisted of the average number of coarticulatory processes identified per fingerspelled letter. Of the fingerspelled words, 44 percent exhibited deletion, and approximately 40 percent of all fingerspelled letters exhibited coarticulatory processes.

Thumann [29] also explored a comparison of careful versus rapid fingerspelling by analysing a recording of a conversation between two native ASL users discussing the city of Mobile, Alabama. In the conversation, the women fingerspelled the word "Mobile" 23 times. Thumann found occurrences of both deletion and coarticulation, which resulted in shorter durations.

Keane [17] reported on a first step to analyze a newly-established fingerspelling corpus for coarticulation by considering the spread of pinky extension across multiple fingerspelled letters. They found that the spread was more prevalent in rapidly-fingerspelled words.

As a result of letter deletion and coarticulation, the letters in a fingerspelled word may appear differently from the idealized form shown in a textbook illustration, and it can be difficult for novice signers to recognize it in the context of a word.

4 Technologies for practicing fingerspelling recognition

As mentioned in the previous section, one of the challenges to acquiring fingerspelling recognition skills is the lack of opportunities for self-study. When learning a spoken language such as English, students have access to a rich and varied supply of materials for self-study including newspapers, video recordings, learning software and entire libraries of written material. There are far fewer opportunities for a student wanting to practice fingerspelling recognition.

Previous technologies used for self-study include 1) video recordings of fluent signers, 2) flash card technology, and 3) 3D avatar technology. The 1980s marked the appearance of videotaped recordings of fingerspelling produced by fluent ASL signers. In the 1990s, CDs and DVDs designed for fingerspelling practice became available [15]. These media showed skilled signers demonstrating words in careful and rapid styles of fingerspelling production. Because these are fluent signers, the fingerspelling naturally exhibits both coarticulation and deletion. However, in media of this type, the vocabulary words are fixed at the time of recording. Adding new vocabulary required more recording sessions at an additional cost. Since the videos were recorded at low frame rates, motion blur was a problem, as was a lack of variation in the presentation order. As a student viewed and reviewed the same vocabulary presented in the same order, it was not clear if the student was improving their recognition skills or merely memorizing the recording.

The advent of Internet-based technologies paved the way for several browser-based applications such as [32] that offer fingerspelling practice. When using one of these applications, a student can view a word as a succession of static snapshots or flash cards, each showing a single letter. Once the spelling is complete, students can guess the word and receive feedback. The advantage of these sites is their flexibility. A site can spell any word by simply shuffling the flash cards and can introduce new vocabulary without incurring costs for additional recordings. However, there is a drawback due to the static nature of the snapshots. There is no connective movement between the static images in these practice tools. Linguistic research has revealed that the transitions between fingerspelled letters are not only important, but vital to fingerspelling recognition. Students need to view the movement envelope that is intrinsic to fingerspelling.

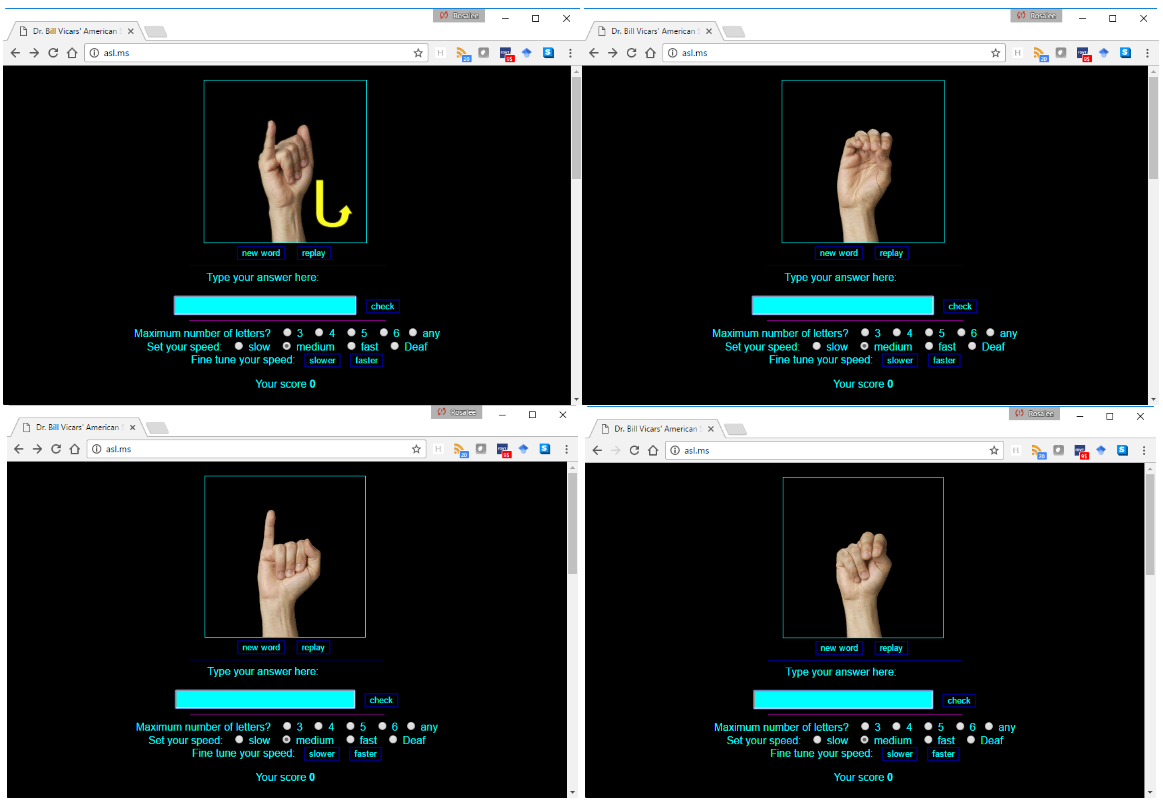


Figure 6: "Flash-card" style of fingerspelling presentation [32]

A third alternative is 3D avatar technology, which promises the extensibility for new word formation while producing smoothly flowing motion, but it poses some challenges of its own. Fingerspelling puts greater demands on avatar technology than those required for other applications such as conventional video game play. Using a 3D avatar for fingerspelling requires careful attention to simulating the flexible webbing between the thumb and index finger and mimicking the complex behavior of the base of the thumb [21].

Avatars suffer from a lack of physicality. Unless prevented, the thumb and fingers will pass through each other when transitioning between closed handshapes such as the ASL manual letters A, M, N, S, and T manual letters in ASL. Figure 7 demonstrates the differences between a naive interpolation of the transition from N to A and a human production of the same transition. In the naive interpolation, the index and middle fingers descend and the thumb cuts through the flesh of the two fingers on its way to the radial side of the hand. In contrast, a human signer will straighten the metacarpophalangeal joints of the index and middle fingers, lifting them upwards momentarily to allow the thumb to pass underneath the fingers. Such motions are part of the movement envelope described by Akamatsu [2] and are essential to a realistic display of fingerspelling. To simulate this physically-based transition via an avatar requires a system to prevent finger collisions in order to faithfully replicate the motion envelope produced by fluent signers.

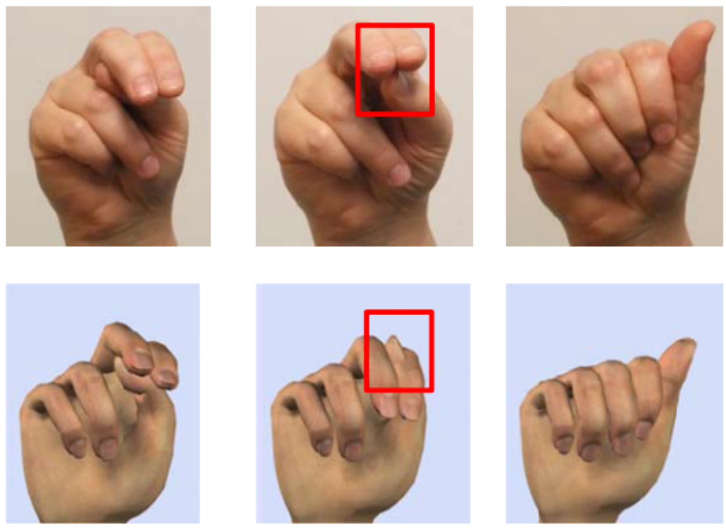


Figure 7: The contrast between human collision avoidance and naive interpolation in an avatar [30]

Accurately portraying this high level of realism in an avatar entails large computational requirements. For this reason, some previous efforts sacrificed realism to gain real-time speeds by using a simplified 3D model that did not accurately portray a human hand and/or did not prevent collisions [27] [12]. Another early approach to collision avoidance was to move the hand into a neutral position between each letter[1]. However, the resulting motion of this approach does not follow the shape of the motion envelope and introduces handshapes not present in fingerspelling produced by fluent signers.

Other efforts [36] [30] created a real-time fingerspelling avatar that required only modest computing resources and addressed the collision problem to produce a smoothly-flowing motion envelope. The approach involved a pre-rendering step that carefully organized the animation from a series of small video clips that each contained a single letter-to-letter transition. These transition clips are very short: if the avatar is spelling at two letters a second, then there are 15 frames in a transition; if the avatar is spelling at three letters a second, then there are 10 frames in a transition. Since each clip had a transition between only two letters, the problem of collisions became more tractable. As part of the pre-rendering step, animation artists reviewed each video clip, and manually added animation keys to remove any collisions. For example, in the N to A transition mentioned earlier, an animator added keys to the index and middle metacarpophalangeal joints to cause the index and middle fingers to rotate out of the way of the thumb's path. Figure 8 shows selected frames from the video clip.

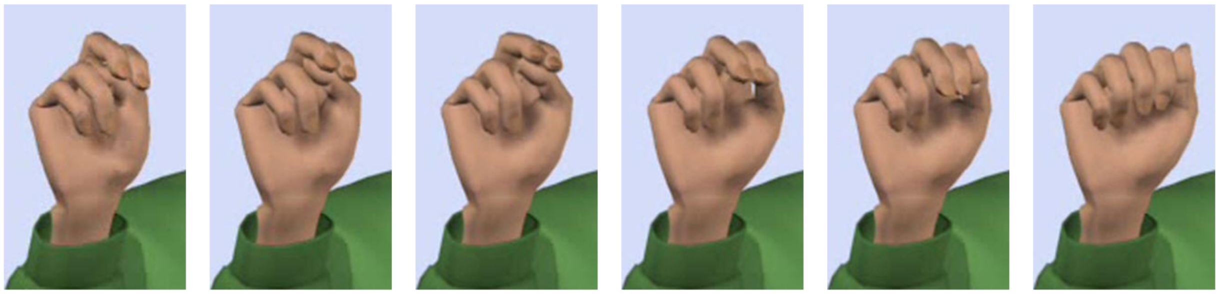


Figure 8: Frames from a video clip depicting the transition from the letter N to the letter A

This technique creates any word by combining transition clips to display the fingerspelled word in real time. As shown in Figure 9, the motion produced closely resembles careful fingerspelling. Currently, it is the best tradeoff between extensibility of vocabulary and naturalness of motion. For a neophyte, this technology provides a good starting place for practicing fingerspelling recognition.

Figure 9: An animation of the fingerspelled word T-U-N-A, created from three clips depicting letter transitions [36]

However, in its present form, it is only capable of displaying perfectly-produced fingerspelling. This may be a good option for a beginner, but for a novice wishing to study rapid fingerspelling, it has limited usefulness. Adding a speed control capability to display individual frames at a faster rate is certainly possible, but the resulting animation would still portray all letters in their canonical form. The processes of coarticulation and deletion, which occur in rapid fingerspelling, are not displayable with the current technology. We wanted to know if it would be possible to identify characteristics of rapid fingerspelling that could be incorporated into the current technology. Gaining insights into this question required a corpus study.

5 Developing a corpus

To build a corpus that would satisfy the requirements of such a study, we first had to choose the medium for the recording. Two primary methods exist for capturing fingerspelling: video and motion capture. While motion capture can give far more detailed data than video, it is also quite invasive, requiring a glove or series of sensors applied to the fingers. This has the potential to radically change or slow a signer's fingerspelling. Because we were most concerned with capturing natural rapid fingerspelling, we chose to work with video for this corpus.

A primary challenge in building a video corpus for fingerspelling is the incredible rate at which letters are produced, particularly for fast fingerspelling. For example, if fingerspelling is occurring at a rate of 5 letters per second, and video is recorded at 30 frames per second (fps), then at most 6 video frames will be dedicated to each transition. This can result in extremely blurry frames, especially when the fingers are moving quickly. Thus to capture transitions and coarticulations with high fidelity, we used a high-frame-rate video camera capable of 240 frames per second at 640x480 standard definition resolution. This allowed us to record clear frames even for the most rapid fingerspelling.

The corpus was designed to capture a range of different fingerspelling phenomena, both in the context of a larger signed discourse and in isolated examples. To accomplish this, we used two separate stimuli:

1. A script of someone detailing a list of people to invite to a wedding reception. This script made it natural to chain together lists of names with connecting phrases and thoughts. The names were chosen to exercise a range of letter combinations.
2. A list of isolated words designed to include fingerspelled letter combinations where open handshapes were followed by closed handshapes and vice-versa. Thus this list contains some "worst-case" situations for extreme finger movement.

Certified ASL interpreters were hired for the recordings. The interpreter signed each script in the following styles:

a. A "teacher" style in which the intended audience had little fingerspelling recognition experience and who needed fingerspelling that was as crisp and clear as possible. This corresponded closely to careful fingerspelling.
b. A "fluent" style in the manner that they would sign to a native signer. This corresponded to rapid fingerspelling.

Each style was recorded with the signer being asked to sign at an appropriate speed and then recorded again signing at a faster speed. Each script was projected as text directly in front of the signer, and recordings were taken from a front camera only.

The captured videos were cut into individual clips, each containing a single fingerspelled word. In the case of the names from the wedding invitation discourse, transitions into the first letter were included in the clip to give context to the position and orientation of the first fingerspelled letter in the name. The result was a corpus of 524 fingerspellings of 80 unique words recorded in standard definition and at a high frame rate, allowing us to clearly see the shape of the hand through the entire fingerspelling action.

Each of the individual video clips were annotated in ELAN by a student familiar with ASL fingerspelling and were then checked by a faculty member similarly familiar with the handshapes. The following ELAN tiers were generated:

a. Word: containing a single annotation for the fingerspelled word spanning the entire motion.
b. Letter: containing annotations for fully formed letters. The annotations span the length of time that the full handshape is held. This may include some movement in the orientation of the wrist that did not significantly affect the shape of the hand.
c. Coarticulation: containing annotations of handshapes that have significant modifications, but in which some aspect of the handshape was still recognizable. The annotations span

the recognizable elements of the handshape. In addition, this tier includes all instances of coarticulation wherein two letters are signed within the same motion. Where there was ambiguity in which tier a letter should be placed, the annotator favored including it in this "Coarticulation" tier.

d. Deletion: This tier marks letters that are deleted. The annotation length is not significant here as the letter is completely unrecognizable anywhere in the sequence of frames. The annotation is placed between the annotations of surrounding present or coarticulated letters.

An example of the fingerspelled word V-E-R-O-N-I-C-A with annotationS on all four tiers is displayed in Figure 10. In this example, the letters R, N, C and A were fully formed but the O was deleted between the R and N. In addition, the E was altered so that it only involved the index and middle fingers, as they were subsequently used to make the R. Interestingly, this example also contained a leading deletion because the V was subsumed by an initial enumeration sign involving the index finger.

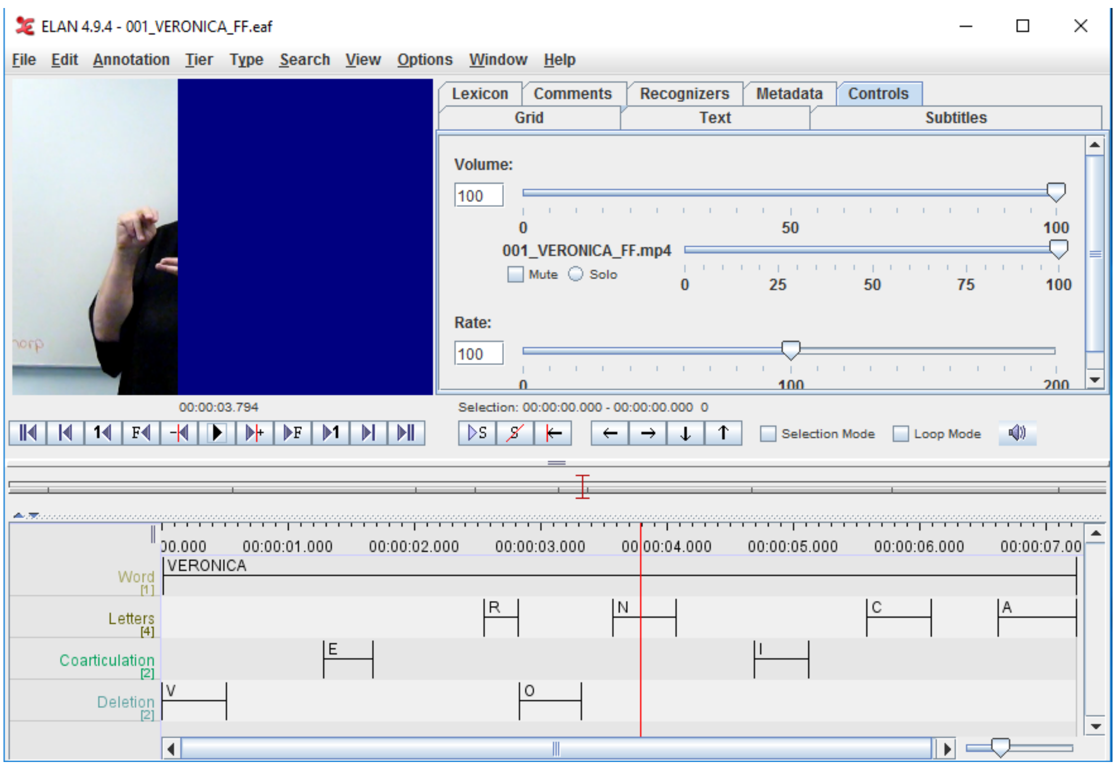


Figure 10: Example of annotations for the fingerspelled name "Veronica"

6 Using the n-gram analytic tool for fingerspelling analysis

In looking to improve the fingerspelling display technology, we wished to determine the nature of the relationship between the speed of fingerspelling and the occurrences of coarticulation and deletion. In particular, we wished to determine how coarticulation and deletion could be incorporated into the tool's fingerspelling as the rate of fingerspelling increased. This would allow the tool to more faithfully reproduce rapid fingerspelling for more advanced students. To study this, we used the n-gram analytic tool to analyze the following statistical patterns:

- Fingerspellings of words that involve any coarticulated or deleted letters
- Duration of 3-grams that have either a coarticulation or a deletion as the second letter of the 3-gram
- Letters which were most often coarticulated or deleted in the 3-gram

For brevity, we will call any coarticulation or deletion generally a *modification* of a letter. Our first analysis was to get a bird's eye view of letter modification by looking at the overall frequency of occurrence in words. The results were consistent with past studies of fingerspelling coarticulation. Among the 524 fingerspellings of words, coarticulation or deletion happened in 65% of the cases. Breaking this out between coarticulation and deletion yields the following rates for all speeds of fingerspelling.

- Coarticulation occurred in 55% of fingerspellings
- Deletion occurred in 25% of fingerspellings

To gauge the relationship between speed and coarticulation, we analyzed the full set 1390 individual 3-grams occurring in the corpus' 524 fingerspellings. For example, in a fingerspelling of V-E-R-O-N-I-C-A, there are the following six 3-grams:

VER, ERO, RON, ONI, NIC, ICA

To avoid duplications, we chose to look only at coarticulation or deletion that occurred on the middle letter of each 3-gram. This necessarily excluded the initial and final letters of each word from the analysis, where coarticulation and deletion were both expected to be relatively rare. We will call such letters *interior* letters. Overall coarticulation and deletion of interior letters occurred in 38% of all the 3-grams.

As a measure of the fingerspelling speed, we chose to look at the overall duration of each 3-gram, which allows for varying speed during a fingerspelling production. While certain letters do take a little more time to produce, J and Z for example, this difference averages out somewhat with the presence of the two other letters in the 3-gram. A histogram of interior modification to 3-grams indexed by duration is shown in Figure 11 and reveals that modifications happen far more often with more rapid 3-grams.

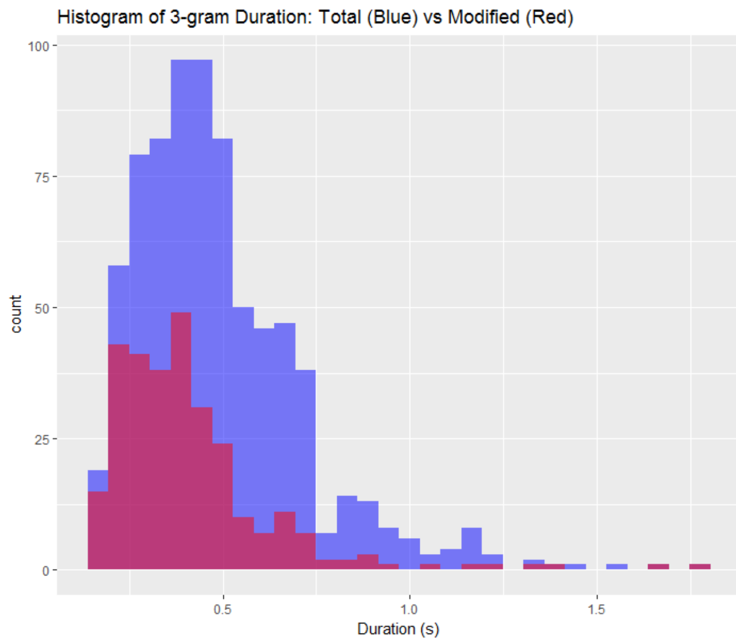


Figure 11: Histogram of 3-gram durations and modifications

To further quantify this, Figure 12 graphs the percentage of 3-grams containing interior letter modifications by duration. The graph shows a clear decreasing linear relationship between the log-percentage of letters modified vs 3-gram duration.

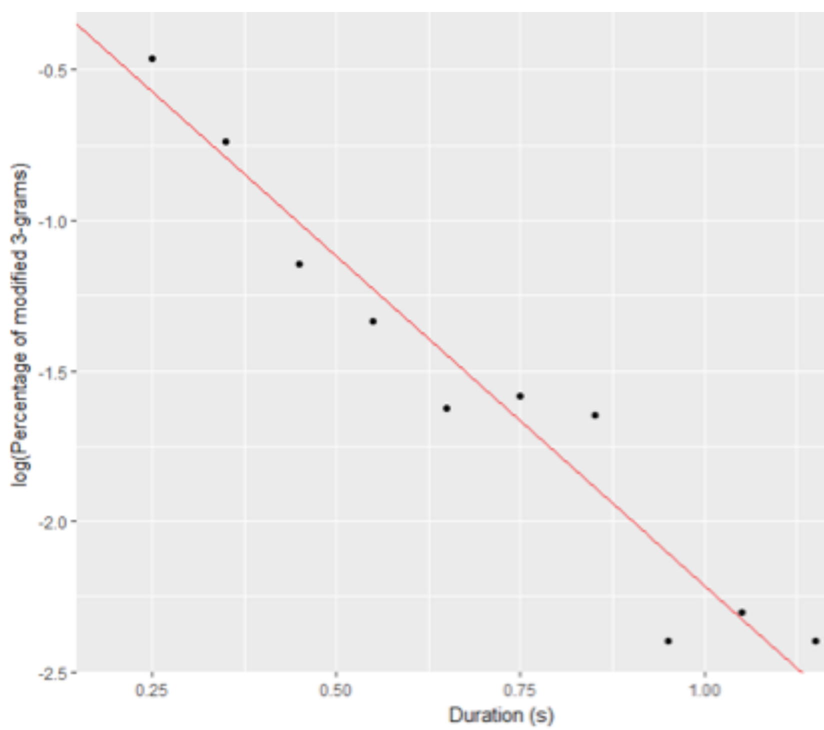


Figure 12: Relationship between percentage of modified letters and 3-gram duration

Running a regression analysis on this relationship yielded the following results:

Intercept	-.026
Slope	-2.19
Residual Standard Error	.166 (9 d.o.f)
R^2	.95

This yields the following approximation for the percentage of letter modifications

$$p = .974e^{-2.19s}$$

This tells us that at a duration of zero seconds, the percentage of modifications is essentially 100%, and each 0.1 second increase in duration multiplies the percentage of modifications by a factor of 0.8.

We can further analyze this by looking at the percentages of modifications at each duration that are deletion vs coarticulation. The graph in Figure 13 displays the percentage of modifications that are deletions at each 3-gram duration, and shows that the deletions form an increasing percentage of the modifications at shorter durations.

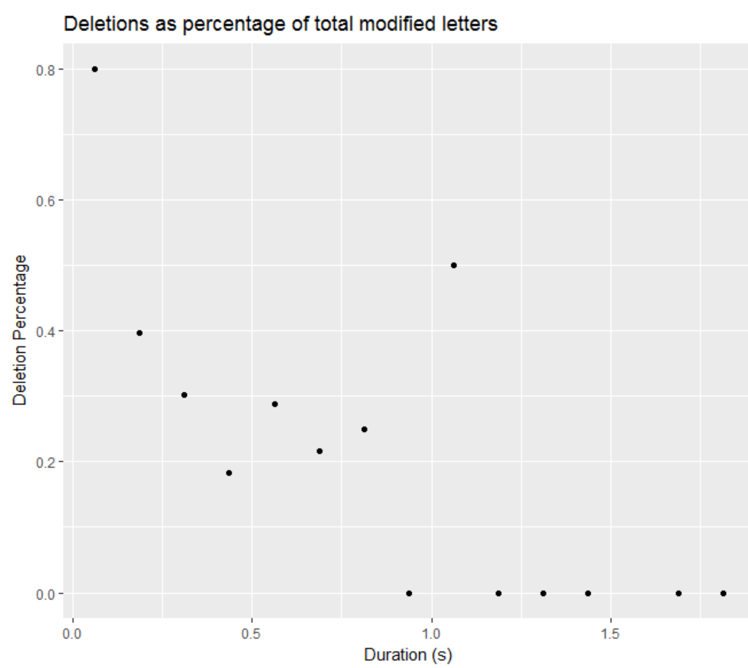


Figure 13: Deletion as percentage of total modifications vs duration

Since these recommendations are only a percentage or a probability that letters will be deleted or coarticulated, we can sharpen this analysis by looking at the overall frequencies of letter modifications to see which letters are more likely to be affected. Figure 14 shows the letters that are most commonly modified. The red bars delineate the vowels from the other letters and we can see that for both coarticulation and deletion, the vowels E, O, and A are commonly modified, while I is more likely to be deleted than coarticulated. For non-vowel letters, N is the most likely to be modified. For coarticulation, 67.7% of the letters were vowels and for deletion, 59.8% of the letters were vowels.

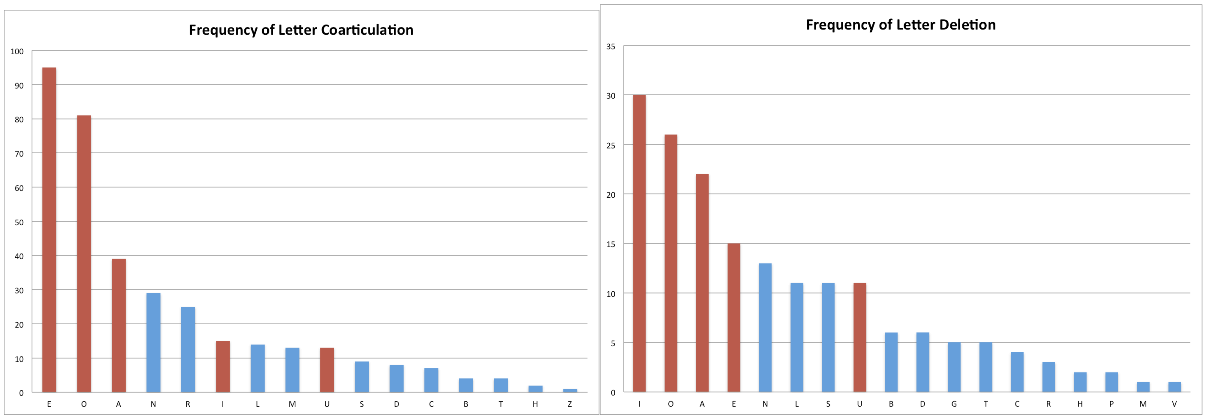


Figure 14: Frequency of letter coarticulation and deletion

7 Results and future work

From the last section, we can draw the following conclusions for modifying the current fingerspelling technology:

- Rather than simply increasing the rate of playback, as the speed of fingerspelling increases, we should modify or delete letters. Further, as the rate of fingerspelling increases, coarticulation will give way to deletions.
- The first letters that should be modified are the vowels, followed by handshapes such as those in the letters.

These findings suggest two modifications to the current fingerspelling technology to produce a more realistic rapid fingerspelling style than would be possible by simply increasing the frame rate. The first is to introduce deletions in medial letter positions. The second is to introduce coarticulation by careful editing of the transition clips before assembling and displaying the fingerspelled word.

Introducing deletions is simply implemented via a preprocessing step where the word to be spelled is edited to remove the letters affected by deletion. The first choices for deletion are medial vowels, followed by a preference for the medial letters L, M, N, and R. For example, a deletion applied to the word M-A-R-Y would result in M-R-Y, which would utilize only the transitions M to R, and R to Y.

Introducing coarticulation is more involved and requires editing of the individual transition clips. As demonstrated in Figure 9, the initial frame of a transition clip depicts a letter in its canonical form. Instead of using the entire transition clip, a new software module shortens the clip to start with frame 2 or frame 3, depending on the speed of the fingerspelling. The letter undergoes modifications to generate coarticulatory effects, and the resulting animation is smoother and presents a more accurate simulation of the overall motion envelope as described by [2]. The link http://asl.cs.depaul.edu/video/LinguisticsAndLiteratureStudies/comparison.mp4 shows two versions of the fingerspelled word V-E-R-O-N-I-C-A. The version on the right demonstrates the effect of deletion and coarticulation, and more accurately simulates the motion envelope.

Future plans include additional analysis of the coarticulatory processes in rapid fingerspelling to make further modifications to the portrayal of automatic fingerspelling generation by adding annotations for orientation change and having someone with more linguistic training annotate these instances with further detail of the types of modifications that the handshapes undergo.

Bibliography

[1] Adamo-Villani, N., & Beni, G. (2004). Automated finger spelling by highly realistic 3D animation. *British Journal of Educational Technology, 35*(3), 345-362.

[2] Akamatsu, C. T. (1982). *The acquisition of fingerspelling in pre-school children.* Ann Arbor: University Microfilms.

[3] Armstrong, D. F., Stokoe, W. C., & Wilcox, S. E. (1995). *Gesture and the nature of language.* Cambridge: Cambridge University Press.

[4] Auer, E., Russel, A., Sloetjes, H., Wittenburg, P., Schreer, O., & Masnieri, S. (2010). ELAN as Flexible Annotation Framework for Sound and Image Processing Detectors. *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)* (pp. 890-893). Paris: ELRA.

[5] Banerjee, S., & Pedersen, T. (2003). The design, implementation, and use of the Ngram statistics package. *Computational Linguistics and Intelligent Text Processing*, 370-381.

[6] Battison, R. M. (1978). *Lexical Borrowing in American Sign Language.* Silver Spring, MD: Linstok Press.

[7] Berke, L. (2013). *Empowering ELAN with N-gram Analytics for Corpora.* DePaul University, School of Computing, Chicago. Retrieved from http://asl.cs.depaul.edu/software/nGramELANtechReport.pdf

[8] Brugman, H., & Russel, A. (2004). Annotating Multi-media/Multi-modal Resources with ELAN. *Proceedings of the 2nd Workshop on the Representation and Processing of Sign languages: Lexicographic matters and didactic scenarios* (pp. 82-87). Paris: ELRA.

[9] Channer, C. S. (2013). *Coarticulation in American Sign Language Fingerspelling.* (Master's Thesis) University of Minnesota.

[10] Crasborn, O., Hulsbocsh, M., Lampen, L., & Sloetjes, H. (2013). New multilayer concordance functions in ELAN and TROVA. *Proceedings of the Tilburg Gesture Research Meeting [TiGeR 2013].*

[11] Crasborn, O., Sloetjes, H., Auer, E., & Wittenburg, P. (2006). Combining video and numeric data in the analysis of sign languages within the ELAN annotation software. *Proc. LREC 2006 workshop on representation & processing of sign language* (pp. 82-87). Paris: ELRA.

[12] Dickson, S. (2013). Advanced Animation in Mathematica. The Mathematica Journal, 15(2). doi:dx.doi.org/doi:10.3888/tmj.15-2

[13] Dreuw, P., & Ney, H. (2008). Towards automatic sign language annotation for the ELAN

tool. *LREC Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora.* Paris: ELRA.
[14] Grushkin, D. A. (1998). Lexidactylophobia: The irrational fear of fingerspelling. *American Annals of the Deaf*, 404-415.
[15] Jaklic, A., Vodopivec, D., & Komac, V. (1995). Learning sign language through multimedia. *International Conference on Multimedia Computing and Systems* (pp. 282-285). Washington, DC: IEEE.
[16] Jerde, T., Soechting, J., & Flanders, M. (2003, March 15). Coarticulation in fluent fingerspelling. *The Journal of Neuroscience, 23*(6), 2383-2393. Retrieved February 24, 2013, from http://www.jneurosci.org/content/23/6/2383.full.pdf+html
[17] Keane, J. (2014). *Towards an articulatory model of handshape What fingerspelling tells us about the phonetics and phonology of handshape in American Sign Language.* The University of Chicago.
[18] Kipp, M. (2010). Multimedia annotation, querying and analysis in ANVIL. *Multimedia information extraction, 19*.
[19] Max Planck Institute for Psycholinguistics. (2013, September 19). *ELAN.* Retrieved from The Language Archive: http://tla.mpi.nl/tools/tla-tools/elan/
[20] McDonald, J., Alkoby, K., Carter, R., Christopher, J., Davidson, M., Ethridge, D., . . . Wolfe, R. (2001, May). An improved articulated model of the human hand. *The Visual Computer, 17*(3), 158-166.
[21] Padden, C. (1991). The acquisition of fingerspelling by Deaf children. In P. Siple, & S. Fischer (Eds.), *Theoretical Issues in Sign Language Research* (pp. 191-210). Chicago: University of Chicago.
[22] Patrie, C., & Johnson, R. (2011). *Fingerspelled Word Recognition through Rapid Serial Visual Presentation.* San Diego, CA: DawnSignPress.
[23] Patrie, C., & Taylor, M. (2007). *Outcomes for graduates of baccalaureate interpreter preparation programs specializing in interpreting in K-12th grade settings.* Retrieved from http://scholarworks.rit.edu/cgi/viewcontent.cgi?article=1067&context=books
[24] Sloetjes, H., & Wittenburg, P. (2008). Annotation by category - Elan and ISO DCR. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008).* Paris: ELRA.
[25] Stehouwer, H., & Auer, E. (2011). Unlocking Language Archives Using Search. *Proceedings of the Workshop on Language Technologies for Digital Humanities and Cultural Heritage*, (pp. 19-26).
[26] Strong, M. (Ed.). (1988). *Language Learning and Deafness.* Cambridge: Cambridge University Press.
[27] Su, A. (1998). VRML-based Representations of ASL fingerspelling on the World-Wide Web. *The Third International ACM SIGCAPH Conference on Assistive Technologies.* Marina del Rey, CA: ACM. Retrieved January 28, 2014, from http://www.csdl.tamu.edu/~su/publications/assets98_.html
[28] Sutton-Spence, R., & Woll, B. (1998). Fingerspelling as a resource for lexical innovation in British sign language. (M. Verma, Ed.) *Sociolinguistics, language and society. Language and development, 5*, 10-143.
[29] Thumann, M. (2009). Fingerspelling in a word. *Journal of Interpretation*, 104-126.
[30] Toro, J. A., McDonald, J. C., & Wolfe, R. (2014). Fostering better Deaf/hearing communication through a novel mobile app for fingerspelling. (pp. 559-564). Springer

International.
[31] Valli, C., & Lucas, C. (1995). *Linguistics of American Sign Language: An Introduction.* Washington, DC: Gallaudet University Press.
[32] Vicars, B. (2005). *Dr. Bill Vicars' American Sign Language (ASL) Fingerspelling Practice Site*. Retrieved January 28, 2014, from http://asl.ms
[33] Wager, D. S. (2012). *Fingerspelling in American Sign Language: A case study of styles and reduction.* The University of Utah.
[34] Wilcox, S. (1992). *The Phonetics of Fingerspelling.* Amsterdam: John Benjamins Publishing.
[35] Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). Elan: a professional framework for multimodality research. *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)* (pp. 1556-1559). Paris: ELRA.
[36] Wolfe, R., Alba, N., Billups, S., Davidson, M., Dwyer, C., Gorman Jamrozik, D., . . . Young,
J. (2006). An Improved Tool for Fingerspelling Recognition. *Technology and Persons with Disabilities Conference.* Los Angeles.
[37] Zakia, R. D., & Haber, R. N. (1971). Sequential letter and word recognition in deaf and hearing subjects. *Attention, Perception & Psychophysics, 9*(1), 110-114.