Exploring factors that contribute to successful fingerspelling comprehension

Leah C. Geer* & Jonathan Keane†

The University of Texas at Austin*, University of Chicago†
leah.geer@utexas.edu, jonkeane@uchicago.edu

Abstract

Using a novel approach, we examine which cues in a fingerspelling stream, namely holds or transitions, allow for more successful comprehension by students learning American Sign Language (ASL). Sixteen university-level ASL students participated in this study. They were shown video clips of a native signer fingerspelling common English words. Clips were modified in the following ways: all were slowed down to half speed, one-third of the clips were modified to black out the transition portion of the fingerspelling stream, and one-third modified to have holds blacked out. The remaining third of clips were free of blacked out portions, which we used to establish a baseline of comprehension. Research by Wilcox (1992), among others, suggested that transitions provide more rich information, and thus items with the holds blacked out should be easier to comprehend than items with the transitions blacked out. This was not found to be the case here. Students achieved higher comprehension scores when hold information was provided. Data from this project can be used to design training tools to help students become more proficient at fingerspelling comprehension, a skill with which most students struggle.

Keywords: fingerspelling comprehension, ASL corpora, L2 acquisition

1. Introduction

The purpose of this study is to investigate which portions of the fingerspelling stream carry the most rich information for learners of American Sign Language (ASL) trying to comprehend fingerspelled items. Adult learners of ASL learn to produce fingerspelled words by forming the manual representation of each letter of the English alphabet in sequence. This process is rather disjointed, particularly early on. With practice, however, the skilled signer fingerspells fluidly with extensive coarticulation. These productions look very different from the disjointed process of sequencing the citation forms of manual letters in turn. The process by which learners acquire this skill is a long and difficult one. It has been noted – *eg* in a study by Wilcox (1992) – that fingerspelling is one of the hardest aspects of ASL acquisition. What makes acquiring this skill so challenging is not a settled matter. This applies not only to production of fluid fingerspelling, but the perception of fingerspelled items as well. This study begins exploring what parts of the fingerspelling stream benefit new learners most in a word comprehension task.

The past 40–50 years of linguistic research on signed languages have demonstrated many ways in which their features have correlates in spoken languages. Both language modalities have phonetics and phonology, however there are modality specific differences which may impact language processing. Furthermore, there may be different processing requirements for signs as a whole as compared to fingerspelling. Before we describe our experiment and the rationale for this type of research, it is important to understand why we pose a question worthy of investigation.

In an introduction to a volume which examines similarities and differences in signed and spoken languages from around the world, Meier (2002) discusses reasons that these two language modalities necessarily differ. The two that are most relevant for the present discussion are the differing properties of the articulators (oral articulators versus manual articulators) and the different properties of the perceptual systems (aural perception versus visual perception). Compared to the oral articulators, the upper extremities are relatively large. Consequently, the rate of sign production is much slower per unit time as compared to speech, yet propositions can be produced at a comparable rate (Bellugi and Fischer, 1972). To compensate for the slow rate of sign production, it has been suggested that signed languages make use of simultaneous layering of linguistic features, rather than the sequentiality observed in the structure of spoken languages (Klima and Bellugi, 1979; Brentari, 2002).

This organization of linguistic features is consistent with how the perceptual systems are able to process linguistic information. As Bregman (1990) points out, vertical and horizontal processing are both used in vision and audition. The visual system is better equipped to handle vertical processing and peripheral arrangement of information, while the auditory system is better able to cope with horizontal processing, and its associated slower signal transmission, and non-peripheral spatial arrangement of information (Brentari, 2002). This means that, generally speaking, vision better handles vertical processing tasks, while audition is better suited to horizontal processing. The organization of information in the speech and sign modalities clearly reflects the difference in perceptual abilities of the eyes and ears. Crucially, it is notable that in a modality that favors the simultaneous presentation of information, fingerspelling is sequentially organized: fingerspelling consists of the successive production of manual representations of orthographic characters. This poses interesting questions regarding how this part of the language as processed, both by skilled users, and also for students new to the language and likely language modality. The experiment we present here examines the latter group with respect to their ability to comprehend fingerspelled items when varying cues are presented.

Wilcox (1992) proposes that transitions are the most information-rich and salient parts of the fingerspelling signal (p. 59), which he suggests is likely because holds (also known as *targets* as he called them, and *postures* by others) are only briefly achieved. However, Keane (2010) and Keane et al. (2013b) found that signers vary considerably with respect to the proportion of holds to the duration of the word (0.7–0.2). Additionally, the temporal duration of the hold or the transitions is actually independent from the kinds of perceptual information that perceivers of fingerspelling use to identify the letters. Because the holds have hand configurations that are closest to the canonical forms for each letter, we hypothesize that when the proportion of hold duration to the duration of the word is around 0.5 (0.4–0.6), and thus the duration of the holds is approximately equal to the duration of the transitions, the holds provide more perceptual information than the transitions.

2. Methods

2.1. Materials

Video clips of fingerspelled items produced in isolation and used in related experimental studies – (Keane et al., 2013a; Keane, 2010; Keane et al., forthcoming) – were used as stimuli for this experiment. The fingerspelling of four deaf signers was recorded, including three native ASL signers, and one early learner. The corpus consists of approximately six hours of video, which includes 5,700 words (11,400 tokens) and approximately 71,250 letters. A number of wordlists were used, but the one used to create stimuli for the present experiment was the 300 most common nouns in the CELEX corpus (Baayan et al., 1995).

The data were collected across different sessions that consisted of all of the words on one wordlist. During each session signers were presented with a word on a computer screen. They were told to fingerspell the word, and then press a green button to advance if they felt that they fingerspelled it accurately, and a red button if they had made a mistake. If the green button was pressed, the word would be repeated and the signer would produce it a second time, then advance to the next word. If the red button was pressed the sequence was not advanced, and the signer repeated the word. Most sessions were collected at a normal speed, which was supposed to be fluid and conversational. Signers were instructed to fingerspell naturally, as if they were talking to another native signer.[1] Each session lasted between 25-40 minutes and there was a self-timed break in the middle of each session for the signer to stretch and rest.

Video was recorded using at least two cameras, both at 45 degree angles from straight on. Video from the camera to the signer's left was used to create these stimuli. Each of these cameras recorded video that was 1920×1080 pixels, 60 fields per second, interlaced, and using the AVCHD format. These files were then processed using FFMPEG to deinterlace, crop, resize, and reencode the video files so that they were compatible with the ELAN annotation software (Crasborn and Sloetjes, 2008). Figure 1 shows an example from this data.

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

Figure 1: A still image depicting how participants saw the stimuli presented on screen.

In order to quantify timing properties of the fingerspelled words (FS-words), the time where the articulators matched the target for each FS-letter in the word needed to be identified. In other words, the fingerspelling stream had to be segmented. The period of hand configuration and orientation stability for each FS-letter is the *hold* (*ie* where the instantaneous velocity of the articulators approached zero). This point was the period where the hand most closely resembled the canonical handshape, although in normal speed the hand configuration was often very different from the canonical handshape (Keane, 2010). To our knowledge, there is no other corpus of comparable size annotated with this type of attention to the timing properties of fingerspelling, making this database uniquely suited for this type of experimentation.

Tokens from one signer who happened to fingerspell words, in general, with about half of the duration of each word consisting of holds (a proportion of holds to word duration of 0.4–0.6) were selected for this experiment.

Clips were modified in three ways. First, all 94 clips (90 experimental items and four practice items) were reduced to half speed in order to facilitate student learners' perception. An early pilot with these materials but with full-speed videos showed many participants exhibited a floor effect across conditions. Thus, in order to examine the relevant cues within the fingerspelling stream with student participants, reduced-speed videos were used. Crucially, this had no effect on the phenomenon at issue here: the relative proportion of holds and transitions in the signing stream (see §1.). One-third of the experimental clips were not modified further. This subset of items was used for the control round to establish a baseline of fingerspelling comprehension for each participant. The remaining two-thirds of clips were further modified to create two experimental conditions. The first condition, *holds only* (transitions black), presented clips modified such that the transitions were blacked out, leaving only the hold portion of the fingerspelling stream; see figure 3. This is the portion in which the manual representation of letters is held, and there is little movement of the handshape. The second condition, *transitions only* (holds black), included clips in which the hold portion of the fingerspelling stream was blacked out, leaving only the portion of the production in which the signer transitioned from one letter to the next. This is pictured in figure 4. This

(a) -S- (b) S-O (c) -O- (d) O-R (e) -R- (f) R-T (g) -T-

Figure 2: Control stimulus for the word S-O-R-T

(a) -S- (b) S-O (c) -O- (d) O-R (e) -R- (f) R-T (g) -T-

Figure 3: Holds only stimulus for the word S-O-R-T

(a) -S- (b) S-O (c) -O- (d) O-R (e) -R- (f) R-T (g) -T-

Figure 4: Transitions only stimulus for the word S-O-R-T

order was chosen based on our initial hypothesis: we expected each condition to be harder than the last. With this setup, as participants become more familiar and comfortable with the task, we expect that their performance should be higher on later blocks. This possibility and its significance will be explored more in the results section (§3.).

2.2. Participants & Procedure

Sixteen ASL students from The University of Texas at Austin were involved in this study. Each had completed at least two semesters of formal ASL instruction and 15 were currently enrolled in the third semester of coursework. Of these, 12 were native speakers of English. We briefly discuss the potential confound of including non-native English speakers in the results section (§3.).

Participants completed the experiment on a laptop computer in a quiet room with no distractions. Each stimulus item was presented twice via the program PsychoPy (Peirce, 2007; Peirce, 2009), a psycholinguistics software program designed for delivery of experimental stimuli and collection of responses. Figure 1 shows an example of what participants saw; in this particular frame, the signer is producing a hold in the form of an 'L' handshape. Figures 2-3 show select frames (one for each hold and transition) for the word S-O-R-T.

Participants were presented with three rounds of 30 experimental tokens in the following order: *control, holds only, transitions only*. Using the visible parts of the token, and after viewing each twice, participants were asked to type whatever word they saw in the response box in PsychoPy. Only exact matches for the fingerspelled word were accepted as correct.

3. Results

Using a hierarchical mixed effects regression model, we found that participants were significantly better at perceiving the fingerspelled words in the *holds only* (transitions black) condition than in the *transitions only* (holds black) condition. This result suggests that transitions themselves, are not privileged with more perceptual information than holds, c.f. Wilcox (1992), but rather the opposite: the holds in fingerspelling convey the most perceptual information when the relative durations of holds and transitions is held constant. In more detail, we found that for the *transitions only* condition, the learners in this experiment had 37.75% accuracy in response to the fingerspelled stimuli (log odds of -0.50 of a correct response; this is the intercept in the model in table 1). For the *holds only* condition accuracy increased 64.57% (log odds of 0.6). Finally, for the control condition, there was 53.25% accuracy (log odds of 0.13). In this model the intercepts and slopes of the effects were allowed to vary by participant, and intercepts were allowed to vary by word. Full output from the statistical model is presented in table 1 and is visualized in figure 5.

As we mentioned in §2.2., of the 16 participants, 12 were native speakers of English. The four non-native speakers' first languages varied: first languages for this subset of participants included Cambodian, Urdu, Russian, and Spanish. Although each of the non-native English speakers had a high degree of English proficiency, it is possible their non-native English could reduce their ability to successfully perceive fingerspelled English words. Although we

	coef.	conf. interval	p-value
(Intercept)	-0.50	[-1.22; 0.20]	0.153
control	0.63	[-0.08; 1.36]	0.078·
holds only	1.10	[0.42; 1.80]	0.001**
AIC	1707.69		
BIC	1760.38		
Log Likelihood	-843.85		
Deviance	1687.69		
Num. obs.	1434		
Num. groups: word	90		
Num. groups: participant	16		
Variance: word.(Intercept)	1.33		
Variance: participant.(Intercept)	1.05		
Variance: participant.control	0.26		
Variance: participant.holds only	0.10		
Variance: Residual	1.00		

Table 1: Mixed effects logistic regression coefficient estimates and standard errors. The reference level (here, the intercept) of the conditions variable is *transitions only* (holds black) condition. The effects of the *holds only* (transitions black) and *control* conditions are compared to this level. A positive coefficient means the participants have higher accuracy with a given condition when compared with the reference level. A negative coefficient means the participants have lower accuracy. Intercepts and slopes were allowed to vary by participant, and intercepts were allowed to vary by word.

do not have enough data to robustly test this hypothesis, we fit the same model described above with only the native English speakers to confirm that these four non-native English speakers did not alter the outcome of the experiment. Although the coefficients vary slightly, the overall results and significant effects remain the same (see table 2 in the appendix for the model fit with only native English speakers).

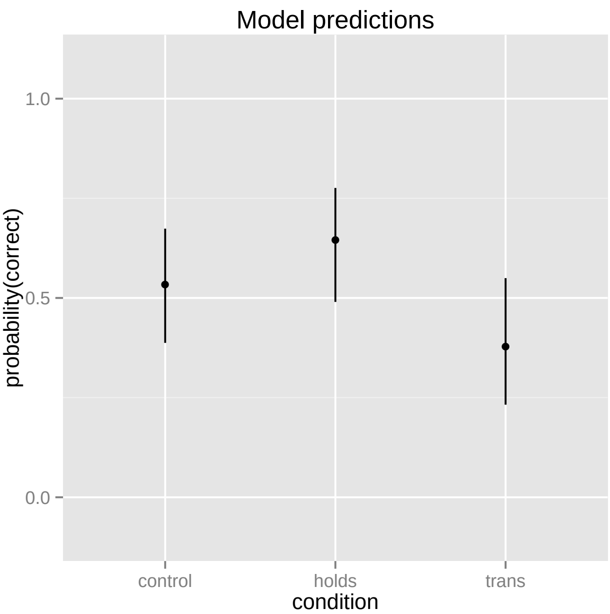

Figure 5: Model predictions: Dots represents model predictions, lines represent 95% confidence intervals. *Control* is the control condition with both holds and transitions, *trans* is the transitions only condition (holds black), and *holds* is the holds only condition (transitions black).

The ordering of the blocks could affect the results. We chose the order based on our initial hypothesis: we expected each condition to be harder than the last. If the participants become more comfortable with the task as they progressed, we expect that their performance should be higher on later blocks. That is, if we factor out the effect of the masking condition itself: we expect that performance on the transitions only block (the last block) to be higher than the holds only block (the middle block), and the performance on both the transitions only and the holds only blocks (the third and second block respectively) to be higher than the control block (the first block). For the first comparison, the results show the opposite pattern: participants are significantly better at the holds only block than the transitions only block. This shows that the effect of the masking condition is not only extant, but also stronger than the effect of participants becoming more comfortable with the task. The second comparison (control versus holds only and transitions only) is a bit more complicated. The control block was not significantly different than the holds only, but participants were significantly better at the control condition than they were the transitions only condition. In §5., we discuss briefly how issues of condition ordering are being addressed in ongoing work.

4. Discussion

We hypothesized that when the proportion of holds represents about half of the overall fingerspelling duration – 0.4–0.6, rather than 0.2–0.7 – holds would provide more rich information to language learners, and thus result in better fingerspelling comprehension when only this cue is made available. Our results show that ASL students do not seem to get more perceptual information from the transitions of fingerspelling when the duration of holds and transitions has been controlled. We propose that perceptual information is multicomponential, and cannot be adequately described with an appeal to duration alone. This is consistent with Bregman (1990) and his suggestion that while one type of processing, vertical or horizontal, might be more active given a particular type of input needing to be processed, both are always active. Some parts of the fingerspelling signal (the holds) provide better visual cues for letter identity than others (the transitions), which is critical to understanding the fingerspelled word. The information provided by these cues is, in principle, independent of the duration of their realization. Our experiment was designed to factor out the durations as much as possible, in order to determine whether the holds or transitions provided more reliable cues for successful fingerspelling comprehension. Although the relationship between the overall duration of holds to the overall duration of transitions in fingerspelling in general is a complicated one, past work has shown that for some signers, as well as for many short words of all signers, the ratio of holds to transitions is much closer to one to one than was previously assumed (Keane et al., 2013c). This result is counter to some conventional wisdom in the ASL linguistics and ASL pedagogy fields.

5. Future work

The fingerspelling corpus used here is unique. Future work will vary the hold-transition ratio along with the blacking out of either to determine the effect sizes resulting from having more or less holds versus transitions duration, as well as the relative effect size when compared to the presence or absence of holds and transitions.

Additionally, native signers and proficient second language learners (e.g., ASL interpreters) need to be tested to determine the degree to which their patterns are similar or different to those reported here by student learners. In the course of piloting this study we had a small number of native signers take the experiment as described here. They all had near ceiling accuracy rates for all stimuli. We do not expect that more proficient signers (especially deaf, native signers) will show precisely the same results as the later learners studied here. A similar paradigm (albeit without the stimuli slowed down, and possibly only presented once in order to remove the ceiling effect we observed in the pilot) will be able to test both of these populations to see if the overall pattern is the same.

We mentioned in §3. that the ordering of the blocks in the present experiment did not vary across participants. In currently ongoing work, we are collecting more subjects with a design that includes randomized blocks. In this paradigm, participants are presented with one of two variants of four blocks. They either receive *control, transition only, hold only, control*; or *control, hold only, transition only, control*. Not only does this allow us to counterbalance the experimental conditions, it also affords the opportunity to examine a potential learning effect.

The broader impacts of this work are great: having quantitative norms of fingerspelling perception and comprehension could help with the development of metrics and tests for what types of comprehension fall outside of the range of typical signers. Norms for typical signers are needed before analyzing how people from different language backgrounds (early learners, late second language learners, etc.) differ in their fingerspelling comprehension. This has further potential impacts on diagnosing language disorders, which has been particularly understudied in ASL signers. There has been research showing a correlation between fingerspelling ability and literacy (Padden, 2006; Emmorey and Petrich, 2011) thus understanding the basic facts about the comprehension of fingerspelling will allow for more detailed future work on perception of fingerspelling. Furthermore, understanding how fingerspelling is perceived will enable the study of this correlation in more detail.

6. Acknowledgements

This study would not have been possible without the generous students who volunteered their time to participate in this experiment and the signers who participated in the initial collection of fingerspelling data.Additionally, the Rella I Cohn fund supported the collection of a subset of the data that was used to generate the stimuli.

A References

Harold R. Baayan, R. Piepenbrock, and L. Gulikers. 1995. *The CELEX Lexical Database. Release 2 (CD-ROM)*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, Pennsylvania.

Ursula Bellugi and Susan D. Fischer. 1972. A comparison of sign language and spoken language. *Cognition*, 1:173–200.

Albert S. Bregman. 1990. *The Auditory Scene Analysis*. MIT Press, Cambridge, MA.

Diane Brentari. 2002. Modality differences in sign language phonology and morphophonemics. In Richard P. Meier, Kearsy Cormier, and David Quinto-Pozos, editors, *Modality and Structure in Signed and Spoken Languages*, pages 35–64. Oxford University Press, New York.

Onno Alex Crasborn and Han Sloetjes. 2008. Enhanced elan functionality for sign language corpora. In *Proceedings of LREC 2008, Sixth International Conference on Language Resources and Evaluation*. Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. http://tla.mpi.nl/tools/tla-tools/elan/.

Karen Emmorey and Jennifer Petrich. 2011. Processing orthographic structure: Associations between print and fingerspelling. *Journal of Deaf Studies and Deaf Education*, 17(2):194–204.

Jonathan Keane, Diane Brentari, and Jason Riggle. 2013a. Coarticulation in ASL fingerspelling. In *Proceedings of the North East Linguistic Society*, number 42. North East Linguistic Society.

Jonathan Keane, Jason Riggle, and Diane Brentari. 2013b. The timing of ASL fingerspelling. conference presentation, November. Texas Linguistics Society 14.

Jonathan Keane, Jason Riggle, and Diane Brentari. 2013c. The timing of ASL fingerspelling. In *Presented at the 14th meeting of the Texas Linguistics Society*, Austin, TX.

Jonathan Keane, Diane Brentari, and Jason Riggle, forthcoming. *The Segment in Phonology and Phonetics*,

	coef.	conf. interval	p-value
(Intercept)	-0.32	[-1.07; 0.42]	0.383
control	0.57	[-0.18; 1.32]	0.126
holds only	1.04	[0.32; 1.76]	0.004**
AIC	1313.20		
BIC	1363.01		
Log Likelihood	-646.60		
Deviance	1293.20		
Num. obs.	1076		
Num. groups: word	90		
Num. groups: participant	12		
Variance: word.(Intercept)	1.40		
Variance: participant.(Intercept)	0.92		
Variance: participant.control	0.19		
Variance: participant.holds only	0.07		
Variance: Residual	1.00		

Table 2: Mixed effects logistic regression coefficient estimates and standard errors for native English speakers only. All of the details here are the same as in table 1 The reference level (here, the intercept) of the conditions variable is *transitions only* (holds black) condition. The effects of the *holds only* (transitions black) and *control* conditions are compared to this level. A positive coefficient means the participants have higher accuracy with a given condition when compared with the reference level. A negative coefficient means the participants have lower accuracy. Intercepts and slopes were allowed to vary by participant, and intercepts were allowed to vary by word.

chapter Segmentation and pinky extension in ASL fingerspelling. Wiley.

Jonathan Keane. 2010. Segment duration and the phonetics of fingerspelling ASL. Master's thesis, University of Chicago, October.

Edward Klima and Ursula Bellugi. 1979. *The Sign of Language*. Harvard University Press, Cambridge.

Richard P. Meier. 2002. Why different, why the same? Explaining effects and non-effects of modality upon linguistic structure in sign and speech. In Richard P. Meier, Kearsy Cormier, and David Quinto-Pozos, editors, *Modality and Structure in Signed and Spoken Languages*, pages 1–26. Oxford University Press, New York.

Carol Padden. 2006. Learning to fingerspell twice: Young signing childrens acquisition of fingerspelling. In Brenda Schick, Mark Marschark, and Patricia Elizabeth Spencer, editors, *Advances in the Sign Language Development of Deaf Children*, pages 189–201. Oxford University Press, New York.

Jonathan W. Peirce. 2007. PsychoPy – psychophysics software in python. *Journal of Neuroscience Methods*, 162:8–13.

Jonathan W. Peirce. 2009. Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, 2:1–8.

Sherman Wilcox. 1992. *The Phonetics of Fingerspelling*. John Benjamins, Philadelphia.