

SIMILARITY OF HANDSHAPE: AN ARTICULATORY MODEL

Paper accepted, but withdrawn from the conference due to a scheduling conflict

Jonathan Keane

University of Chicago
jonkeane@uchicago.edu

ABSTRACT

We propose that the Articulatory Model of Handshape [13] makes predictions about the phonetic and phonological similarity of handshapes. The predictions align with previous work, but are theory-driven instead of data-driven. We propose two methods for calculating this similarity, and then test both with psycholinguistic evidence that shows a clear winner for matching signers' intuitions when asked to judge the similarity of two fingerspelled words.

Keywords: Sign languages, phonetics, phonology, similarity, articulatory phonology

1. INTRODUCTION

Phonetic and phonological similarity has been a topic of exploration for linguists for quite some time (including the seminal Miller and Nicely [21] as well as many subsequent studies on spoken languages). Although it has been well explored for spoken languages, signed languages have seen much less research.

At the beginning of the systematic research into signed languages, there were a number of attempts to quantify handshape similarity within signs [18, 30, 14, 28, 24]. Most of these relied on signer judgements of similarity or confusion between two stimuli. They then produced clusters of handshapes based on this data. In this way, they are using psycholinguistic data to produce a linguistic model of similarity, rather than using psycholinguistic data to confirm the validity of a linguistic model. This was mentioned explicitly by Lane et al. [14] as a necessity because there simply were not appropriate linguistic models to test: "The present study, then, undertakes to see what sort of featural analysis for ASL [(American Sign Language)] results when, using certain specific statistical techniques, we proceed from psychological data to a linguistic model, rather than the reverse".

All of these studies came to the conclusion that there are (at least) two distinct categories of handshapes: open handshapes with the fingers of the hand extended, and closed handshapes with the fingers of the hand flexed. Individually, the studies de-

veloped more finely grained distinctions. For example, Lane et al. [14] found clusters that they separated into distinct features. Moreover, Stungis [28] proposed that this clustering could be turned into a continuous feature space. He found that handshapes could be decomposed along two dimensions: extension (open or closed) and uniform breadth (simplistically this is whether or not all of the fingers have the same configuration). What all of these have in common is that the main distinction is between open and closed.

Since this early period, there has been much more work on phonological models of signed languages broadly [19, 15, 25, 9, 4, 8, 26]. More recently, there has been work on the phonetics of sign languages [29, 12, 11, 16, 17, 31, 20, 13]. Of these, Tyrone et al. [29], Mauk and Tyrone [20] (for location, and contact), and Keane [13] (for handshape) adopted the framework of Articulatory Phonology which explicitly links phonological representations of signs with articulatory gestures that produce those signs.

The models that have been proposed are exactly the kinds of models that Lane et al. [14] observed were missing at the time that of their studies of handshape similarity. Most of these models divide the hand into subcomponents, each of which can take categorical values (via binary features, dependency models, etc.). For example, Brentari's [4] Prosodic Model represents handshapes using a branching feature system. It consists of specifications indicating which fingers are active (selected) and which fingers are inactive (nonselected), as well as what the flexion-extension configuration is of the base (metacarpophalangeal) and the nonbase (proximal interphalangeal and distal interphalangeal) joints.

Keane [13], and his Articulatory Model of Handshape, furthers Brentari's model by developing an explicit connection between the phonological specification for a handshape, and target joint angles for each joint of the (phonetic) hand configuration. His model produces continuous (as opposed to categorical) measures of hand configuration which have been shown in previous studies to better match data on handshape similarity and confusability [28]. Additionally, these continuous measures provide a straightforward way to compare two handshapes.

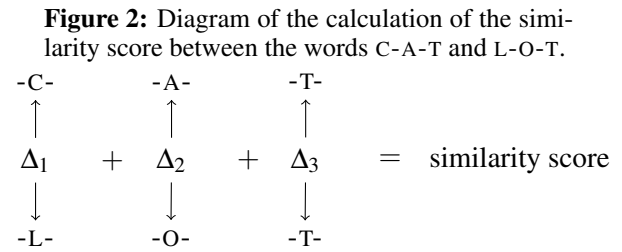
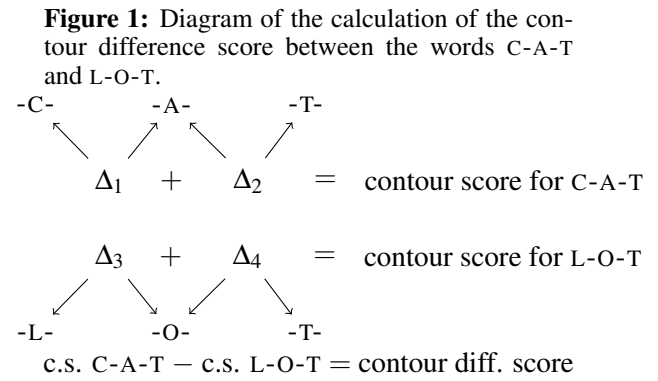
Other phonological models could, in principle, be used, although each would require the development of a translation from categorical phonological features to continuous joint angles or an independent method of comparing the categorical features directly to each other. For these reasons, we will use Keane’s model as our theory-driven measure of phonetic similarity. This similarity will be described in detail in the next section, and then tested (and supported) with psycholinguistic evidence in section 3.

2. WHAT IS SIMILARITY FOR ASL FINGERSPELLING?

Handshapes in sign languages do not occur in a vacuum: they are just one component that makes up lexical signs, along with the other major parameters: location, movement, orientation, and non-manual markers [27, 3]. In ASL, fingerspelling is a loanword system used to borrow (written) English words into the language. In the fingerspelling system, each orthographic letter is mapped onto a unique handshape (and in a limited number of cases, palm orientation). These handshapes are executed in quick succession, in the sequence of the letters of the written word. Fingerspelling has been found to conform to much of the same phonological system as ASL broadly speaking [23, 4, 5, 7]. Because the main contrastive letters in fingerspelling are, for the most part, only handshape contrasts, fingerspelling represents a perfect place to test theories of the representation of handshape independent of the possible confounds of movement or location that would be inherent in using lexical signs or nonce signs that conform to the phonological structure of lexical signs.

Keane’s model [13] provides joint angle targets for each handshape used in ASL fingerspelling. This allows for a straightforward comparison of individual handshapes by taking the difference between the two sets of joint angle targets. This difference can be thought of as the similarity between any given pair of handshapes. This difference is further refined by weighting each joint based on how proximal (or how close to the center of the body) it is. This weighting is supported by work that shows that movement of more proximal joints generates larger visual differences, which has been linked to visual sonority for signed languages [4].

This proposal for individual handshape similarity must be extended to account for fingerspelled words that are composed of sequences of multiple handshapes. Based on the previous research, we have developed two methods of calculating the similarity between fingerspelled words.



The first method we call *contour difference score*; it is based on the general finding that there are (at least) two classes of handshapes (open and closed). In this method, each handshape in the word is compared to the one that follows it, that is, the differences between each sequential pair of letters is calculated and then summed together. Under this metric a word that has a sequence of all open or all closed handshapes will have a low score, and a word that has a sequence of open-close handshapes will have a high score. Using this method, to compare two words, the contour score for each word will be calculated, and then the difference between them will be calculated to determine their overall similarity. See figure 1 for a diagram of an example pair of words.

The second method we call *similarity score*; it is inspired by the finding that a multidimensional, continuous measure of similarity is a superior fit to the data than a purely categorical one [28]. In this method, each pair of letters in the same position within the two words are compared to each other and their difference is calculated. The differences for each position in the word are then summed together. Under this metric words that are similar will have a low score, and words that are dissimilar will have a high score. See figure 2 for a diagram of an example pair of words.

Although the contour difference score can easily compare two words of different lengths, the similarity score as described above, is limited to words that have the same number of letters in them. In the experimental data described below, words with either 3 or 4 letters were compared in pairs that were

the same length, as well as in pairs that differed in length. In order to attain a similarity score, a composite metric was developed: The shorter word was held constant, but then compared to all possible strings of the longer word where one of the letters was deleted. The mean of this score resulted in the final similarity score for mismatched lengths. For example, to generate an overall similarity score for the pair of words L-O-T and L-E-A-N, a score was calculated for each of the following pairs: (L-O-T ; E-A-N), (L-O-T ; L-A-N), (L-O-T ; L-E-N), and L-O-T v. L-E-A. The mean of these four individual scores was taken as the similarity score for $\delta(L-O-T ; L-E-A-N)$. Though there are other methods that could be used to compare words with mismatched lengths, this method is a first step in that direction, which deserves further research.

3. PSYCHOLINGUISTIC EXPERIMENT

Data was collected from 24 Deaf signers in two groups via a computer. The subjects were presented with a pair of fingerspelled words, and then asked to rate the pair as similar or dissimilar on a 5 point scale (where higher ratings are more similar). In the first group they were presented with 107 word pairs, and in the second group they were presented with 132 word pairs.

In order to test which of the two methods of scoring (contour difference scores or similarity scores) predicts signers' ratings, multiple hierarchical linear regressions were fit and then compared. All models were fit with the `lme4` package [2] in R. All scores were divided by the length of the words in order to not unfairly penalize long words. In cases with mismatched word lengths, the scores were divided by 3, since the similarity score for mismatched pairs is the mean of the set of comparisons across the three letter word and all combinations of the four letter word minus one letter. For both scores, a higher score is less similar, and a lower score is more similar (i.e. perfect similarity is 0). If the scores are predictive of the signers' similarity ratings, we expect a negative correlation. All scores were scaled to z-scores for comparison of effect sizes. The models were:

1. *Null model* with no predictor variables, which had varying intercepts (AKA mixed effects) for subject group, subject, first word, and second word.

2. *Contour difference score model* with predictor variables of the contour difference score for the word pair, the length of the words (3 letters, 4 letters, or mismatched), and the two way interaction of these. There were varying intercepts and slopes for subject group, subject, first word, and second word.

3. *Similarity score model* with predictor variables of the similarity score for the word pair, the length of the words (3 letters, 4 letters, or mismatched), and the two way interaction of these with the same varying intercepts and slopes as the previous model.

4. *Full model* which included predictor variables of the similarity score, contour difference score, the length of the words (3 letters, 4 letters, or mismatched), and all possible two and three way interactions with the same varying intercepts and slopes as the previous model.

The null model serves as a baseline of comparison to see if the complexity associated with adding predictors to the model is justified given the data.

In the contour difference score only model, the contour difference score alone does not significantly predict the signers' ratings. There is a significant effect of length, where four letter words are more similar than mismatched words. The interaction of length and contour difference score is significant, when the words are both three letters, then the smaller the contour difference score, the more similar the signers rated the pair. No other predictors had significant effects.

In the similarity score only model, the similarity score significantly predicts the signers' ratings, in the expected direction (the lower the similarity score, the higher the signers' ratings). Additionally word pairs that had the same lengths (either both 3 letters or both 4 letters long) were rated significantly more similar than word pairs that were mismatched. No other predictors had significant effects.

Finally, in the full model, the similarity score significantly predicts the signers' ratings, in the expected direction (the lower the similarity score, the higher the signers' ratings). Word pairs that were both four letters long were rated significantly more similar than word pairs that were mismatched. The effect of the similarity score was (marginally) magnified when both words were four letters. No other predictors had significant effects, which includes predictors for contour difference score, as well as all interactions with the contour difference score.

It stands out that in no model does the contour difference score alone significantly predict signers' ratings of the similarity of fingerspelled words. In contrast, the similarity score, does significantly predict signers' similarity ratings and in the predicted direction. Appendix figure 3 shows predictor coefficients for all models except the null model.

Although there is not a single method for model comparison, especially for hierarchical models like those used here, a number of methods have been proposed and have seen some acceptance.

Table 1: Model comparison using AIC, BIC, and marginal R^2 .

model	AIC	BIC	R^2
null	9166.8	9202.6	0.000
contour diff. score	8868.5	8981.8	0.040
similarity score	8600.7	8714.0	0.173
full	8540.9	8761.6	0.161

The first kind of comparison is the use of information theoretic measures to determine if the extra complexity of adding predictors is justified by the data. In other words, does adding a given parameter give us enough predictive power to justify the added complexity it introduces to the model. There are two mainstream information theoretic measures: Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Both methods can be fit to different non-nested models applied to the same underlying data set (how we are using them here) [6, 1]. For both AIC and BIC lower numbers indicate a better fit of the model to the data. In the most conservative recommendations, a difference of 10 or more indicates that the models differ significantly and the model with the lower score should be preferred (all differences between the AIC and BIC for our models were larger than this threshold). Using the AIC the simplest model that is justified given the data is the full model (AIC: 8540.9) that includes both the similarity score and the contour difference score (however, it should be reiterated that the contour difference score does not have a significant effect in this model). Additionally, the similarity score only model (AIC: 8600.7) is significantly more well supported than the contour score only model (AIC: 8868.5). Using the BIC the simplest model that is justified given the data is the similarity score only model (BIC: 8714.0). Additionally, the full model (BIC: 8761.6) is significantly more well supported than the contour score only model (BIC: 8981.8). See table 1 for AICs and BICs.

The second kind of comparison is to use a new method for calculating R^2 , or the variance of the data explained by the model. Traditionally, calculating R^2 for hierarchical models has not been straightforward. However recent work [22, 10] has developed a method that gives a marginal R^2 , which corresponds to the R^2 of the predictors alone, and a conditional R^2 , which corresponds to the R^2 of the predictors along with the varying intercepts and slopes. With both traditional R^2 and with this new calculation, R^2 ranges from 0 (no variance of the data is explained by the model) to 1 (all of the variance of the data is

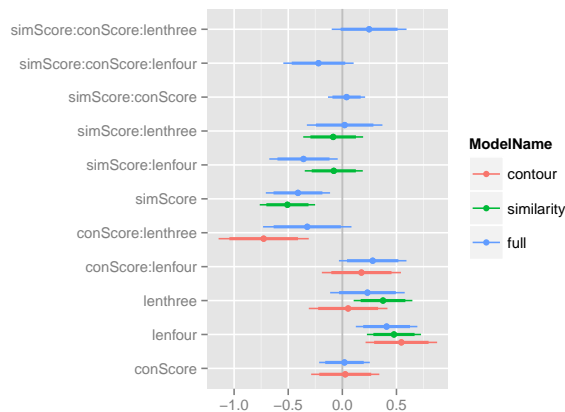
explained by the model). We will only discuss the marginal R^2 here, because we are concerned with the variance explained by the predictors, and not the varying intercepts or slopes. Under this metric, the model that explains the most variance of the data is the similarity score only model ($R^2 = 0.173$) and in a close second, is the full model ($R^2 = 0.161$). Both the contour difference score only model and the null model explain very little variance of the data ($R^2 = 0.040$ and $R^2 = 0.000$, respectively).

Although the model comparisons do not all agree on one specific model, it is clear that two stand out: the similarity score only model (the simplest model justified given the data using BIC, and marginal R^2 , and the second simplest model justified given the data using AIC) and the full model that includes both the similarity score and the contour difference score (the simplest model justified given the data using AIC, and the second simplest model justified given the data using BIC and marginal R^2). Additionally, even when the contour difference score is included in the full model, it does not significantly predict signers' similarity ratings.

4. CONCLUSION

It is clear that similarity score is the theory-driven description of similarity that best matches signers' intuitions when asked to rate the similarity of fingerspelled words. The similarity metric proposed here is exactly the kind of theory-driven similarity that was recognized as missing from the similarity research in the 1970s and 80s, which has also been independently confirmed with signers' intuitions of similarity.

Appendix Figure 3: Coefficient plot for contour difference score, similarity score, and full models. Thick lines are 95% CI, thin lines: 99% CI, and dots: estimates of the predictor coefficients. conScore: contour diff. score; len: length of word with levels four, three, and mismatched (reference level); simScore: similarity score.



5. REFERENCES

- [1] Anderson, D., Burnham, K. 2006. AIC myths and misunderstandings.
- [2] Bates, D. 2010. *Linear mixed model implementation in lme4*. Department of Statistics, University of Wisconsin-Madison.
- [3] Battison, R. 1978. *Lexical Borrowing in American Sign Language*. Silver Spring, Maryland: Linstock Press.
- [4] Brentari, D. 1998. *A prosodic model of sign language phonology*. The MIT Press.
- [5] Brentari, D., 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, 87–119. Mahwah, NJ: Lawrence Erlbaum.
- [6] Burnham, K. P., Anderson, D. R. 2004. Multimodel inference understanding AIC and BIC in model selection. *Sociological methods & research* 33(2), 261–304.
- [7] Cormier, K., Schembri, A., Tyrone, M. E. 2008. One hand or two? Nativisation of fingerspelling in ASL and BANZSL. *Sign Language & Linguistics* 11(1), 3–44.
- [8] Eccarius, P. 2002. Finding common ground: A comparison of handshape across multiple sign languages. Master's thesis Purdue University.
- [9] van der Hulst, H. 1995. The composition of handshapes. *Working Papers in Linguistics* (23), 1–17.
- [10] Johnson, P. C. 2014. Extension of nakagawa & schielzeth's r2glmm to random slopes models. *Methods in Ecology and Evolution* 5(9), 944–946.
- [11] Johnson, R. E., Liddell, S. K. 2011. Toward a phonetic representation of hand configuration: The thumb. *Sign Language Studies* 12(2), 316–333.
- [12] Johnson, R. E., Liddell, S. K. 2011. Toward a phonetic representation of signs: Sequentiality and contrast. *Sign Language Studies* 11(2), 241–274.
- [13] Keane, J. December 2014. *Towards an articulatory model of handshape: What fingerspelling tells us about the phonetics and phonology of handshape in American Sign Language*. PhD thesis University of Chicago. Doctoral dissertation, defended 22 August 2014 Advisors: Diane Brentari, Jason Riggle, and Karen Livescu.
- [14] Lane, H., Boyes-Braem, P., Bellugi, U. 1976. Preliminaries to a distinctive feature analysis of handshapes in american sign language. *Cognitive Psychology* 8(2), 263–289.
- [15] Liddell, S., Johnson, R. E. 1989. American Sign Language: A phonological base. *Sign Language Studies* 64, 195–277.
- [16] Liddell, S. K., Johnson, R. E. 2011. A segmental framework for representing signs phonetically. *Sign Language Studies* 11(3), 408–463.
- [17] Liddell, S. K., Johnson, R. E. 2011. Toward a phonetic representation of hand configuration: The fingers. *Sign Language Studies* 12(1), 5–45.
- [18] Locke, J. L. 1970. Short-term memory encoding strategies of the deaf. *Psychonomic Science* 18(4), 233–234.
- [19] Mandel, M. 1981. *Phonotactics and morphology in American Sign Language*. PhD thesis University of California, Berkeley.
- [20] Mauk, C. E., Tyrone, M. E. 2012. Location in asl: Insights from phonetic variation. *Sign Language & Linguistics* 15(1), 128–146.
- [21] Miller, G. A., Nicely, P. E. 1955. An analysis of perceptual confusions among some english consonants. *The Journal of the Acoustical Society of America* 27(2), 338–352.
- [22] Nakagawa, S., Schielzeth, H. 2013. A general and simple method for obtaining r2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4(2), 133–142.
- [23] Padden, C. 1998. The ASL lexicon. *Sign Language and Linguistics* 1(1), 39–60.
- [24] Richards, J. T., Hanson, V. L. 1985. Visual and production similarity of the handshapes of the american manual alphabet. *Perception & psychophysics* 38(4), 311–319.
- [25] Sandler, W. 1989. *Phonological Representation of the Sign: Linearity and Nonlinearity in American Sign Language*. Foris Pubs USA.
- [26] Sandler, W., Lillo-Martin, D. 2006. *Sign language and linguistic universals*. Cambridge Univ Press.
- [27] Stokoe, W. C., Casterline, D. C., Croneberg, C. G. 1965. *A Dictionary of American Sign Language on Linguistic Principles*. Washington DC: Gallaudet College Press.
- [28] Stungis, J. 1981. Identification and discrimination of handshape in American Sign Language. *Perception & Psychophysics* 29(3), 261–276.
- [29] Tyrone, M. E., Nam, H., Saltzman, E. L., Mathur, G., Goldstein, L. 2010. Prosody and movement in American Sign Language: A task-dynamics approach. *Speech Prosody 2010* 1–4.
- [30] Weyer, S. A. August 1973. Fingerspelling by computer. Technical Report 212 Institute for Mathematical Studies in the Social Sciences, Sanford University Stanford, CA.
- [31] Whitworth, C. December 2011. *Features, Clusters, and Configurations: Units of Contrast in American Sign Language Handshapes*. PhD thesis Gallaudet University.