Fine temporal structure of Finnish sign language

Daniel Aalto[1], *Stina Ojala*[2]

[1]University of Helsinki, Finland
[2]University of Turku, Finland

daniel.aalto@helsinki.fi, stina.ojala@utu.fi

Abstract

Signs can be divided to syllables and further into transitions and nuclei based on the sign flow of the handshapes. Here, a mixed effects linear regression model is used to describe the variation in the duration of the syllable nuclei in a data set of 341 signs (474 syllables) produced by five native FinSL signers during a map task. The phonetic fixed variables are the duration of the adjacent transitions and syllable nuclei; phonological fixed variables are the syllabic length of the sign, the syllable position within the sign, and the sign type (functional or content bearing). Both preceding and following nucleus had a significant effect on the nucleus duration, while an asymmetric effect was found for the transitions: only the postnuclear transition had a significant effect. The syllable structure had no effect. However, the nuclei were shorter in function signs. These results suggest that signs are produced in two stages where the first stage, preparatory transition, is merged with the production of the previous syllable, and the second stage consists of executing the sign.

Index Terms: Finnish sign language, spontaneous signing, mixed effects model

1. Introduction

Introduction is divided into three subsections, which address different aspects of sign language research respectively. Signs in a sign language correspond to words in a spoken language, and thus they are the basic carriers of meaning in the particular language. The sign flow divides into individual signs and further into syllables. However, the notion of a syllable is under a debate within sign linguistics community. Nevertheless, it is accepted that signs divide temporo-spatially to segments/gestures. In the current work we study the impact of lexical and segmental factors to the duration of signed syllable nuclei.

1.1. Temporal organization of signed languages and prosody

Previous studies have shown that facial expressions can switch the function of a signed utterance from a statement to a question [1]. Facial expressions are markers for linguistic stress as well [2], but as Wilbur et Nolen discovered, it does not increase the duration of the signs as such but manifests in faster movement and larger displacement instead [3].

Temporal structure and prosodic features of signed language have been studied mainly in relation to the linguistic scope [4], [5] while this study concentrates on the articulatory issues related to sign flow *per se*.

Kröger et al. [6] compare the principles of syllable structures in relation to articulatory phonology. This study concentrates on the precise syllable-internal structures and their durations. The current project compiles the approach by Kröger et al. in comparing sign data to articulatory phonology principles, and the approach by Wilbur et Nolen [3] in studying syllable durations.

Signs within a sign language are further divided into articulatory gestures. These alternate temporally within sign stream, so that the stream has salient movements and motionless moments called holds. Holds and movements represent the sequential organisation of a sign flow. It also forms the basis for our definition of a signed syllable.

Recent sign language research has concentrated on rhythm in connection with signed poetry, but some of the phenomena found are not present in ordinary signed discourse [7]. Studies on sign coarticulation have suggested that the index finger has a dominant role in governing the rate and speed of both coarticulatory and interarticulatory phenomena [8]. The intertwining rhythms in sign are found within manual and facial elements. One of the key facial elements in signed discourse are the headnods [9]. Sign language prosody is also expressed in facial expressions, however, they are not in scope in this study. The twofold structure of prosody is similar to spoken language prosody division between segmental and suprasegmental features.

In a previous study [10] we showed that syllable durations change in terms of sentence position and sign type. The data showed a reduction of duration in polysyllabic signs and it would suggest that spatial relations have a special role in timing of signed languages. Now we want to take the study questions one step further and look if there are differences according to the syllable type (nucleus vs. transition), and what type of interaction between sign types (function vs. content sign) can be found. Function signs, such as pronouns and conjunctions, denote grammatical functions whereas content signs account for the meaning in the sentence. The position of the sign in a signed utterance affects its duration but not its linguistic stress.

1.2. Coarticulation, signing rate, and segmental factors

Just like isolated words differ from the words spoken in their context because of coarticulatory patterns and phenomena, the signs produced in isolation, i.e. in their dictionary form, are also different from those produced within their context. In signed languages, the coarticulatory patterns differ in several dimensions, as not only the signs in succession within one hand influence each other, but the hands also interact together to form another layer of coarticulatory patterns. This is sometimes called interarticulation. For example, the hands slow down when they are nearer to each other [8], [11]: p. 23.

The two main models of sign language phonology share the goal of segmenting an on-going sign stream, but approach it from different perspectives. The segmental models, includ-

10.21437/SpeechProsody.2014-98

ing HM (hold and movement) model, [3], [12], [4] and the latest contemporary model [13], derive from a more linguistic perspective, while the gesture model takes a more physiology based approach to the problem, and bases the segmentation principles on the general principles of human movement: the alternation of more and less active segments during the signed utterance [14] . This results in a different type of segments within the sign stream. In speech production theories the derivatives of articulatory phonology [15] and motor theory of speech production [16] most resemble the above-mentioned. In this study we reflect on our data in the light of these two types of theories, segmental and gestural theories.

Human action is often a combination of many intertwining rhythms, be it basic human actions, such as walking or eating [17], or something more complex, such as speaking or playing an instrument. The rhythms we use are highly individual. Speech rhythm is one of the key features we observe when trying to identify a speaker [18] – the individual ”figure of speech”.

Lindblom, Mauk et Moon [19] study both sign and speech in equal terms based on the dynamics of the production in a sequence. Here too, the motor equilibrum theory with the human movement principles forms the basis for the analysis. According to Lindblom et al. we should not concentrate on the distinct categories *per se*, but on the interpolation between categories. They ask: if perception likes change, why is phonetic specification built mainly around steady-state attributes?

1.3. Mental lexicon and dictionary forms

The articulatory gestures in spontaneous signing are formed by hands (elbows, wrists, metacarpals and phalanges) in relation to the body (face, chest and shoulders). The particular geometric form taken by bones distal to wrists is called the handshape. Often, the handshapes are static, but fingers might move to create a dynamic handshape. Signs in a sign dictionary are categorised by handshapes, even though handshapes are one of the most difficult parts of the sign, and one of the last elements to be learned by sign acquiring children [18].

As stated earlier, individual signs are the basic building blocks of sign language and as such, they are also the meaning-carriers. In addition, signs are collected into sign language dictionaries, similar to dictionaries of spoken languages. However, unlike spoken languages, signed languages do not have written forms, so a sign language dictionary is based on static pictures or videos. Dictionary forms present normative versions of signs and they also guide the language teachers in their teaching, similarly to how spoken language teachers use dictionaries in their teaching [20].

These dictionary forms are thought to be the basic, prototypical representations in our mental lexicon as well. These mental images or soundscapes of words, signs and sentences are adjusted according to the person we are communicating with. This is done based on so-called perceptual anchors, such as greeting words. In English the word that is often used is ”Hello”. This gives the acoustic guidelines to the speaker’s vocal tract. Similarly in signed language a hand wave is used to greet. This gives many clues to the use of physical space anchors, such as size of hand, positioning of hand in relation to face etc. [21].

For the reasons stated above, the dictionary forms found in Finnish Sign Language online dictionary Suvi (http://suvi.viittomat.net) were used as the main reference guide where applicable to the syllabification processing of the spontaneous signed data. After the syllabification, the signs were labelled as monosyllabic or polysyllabic, and as function or content signs. Syllables were further categorised as nuclei or transitions.

In this study, sign language prosody is in focus, in particular the temporal variation within the constituents of the signs in spontaneous Finnish Sign Language (FinSL). Both phonetic (duration of adjacent constituent) and linguistic factors will be considered. We concentrate on the fine temporal structure of Finnish Sign Language based on the movement patterns of the dominant hand. The purpose of the article is to investigate if lexical factors can affect the duration of the syllables.

2. Methods

2.1. Subjects and data recording

The recording was done in a well-lit office. No spotlights were used but lighting was kept constant by shutting the blinds and curtains to prevent possible glare from outside. The camera was on a tripod 2 metres from the subject facing the subject directly. The subject was facing the camera and was sitting on a sofa to prevent the obstruction of signing space by armrests on one hand and to make the situation more comfortable for the subject at the same time. The equal positioning of the camera and the subjects was ensured by using tape markings on the floor for the camera tripod and the subject’s feet position. The background was a light brick wall with solid colouring. Five native signers of FinSL were recorded with 24 Hz frame rate during a map task. The task elicited free signing from a map with a starting point, a path, and an end-point. A more detailed description of the gathering procedure and the data can be found in [18]: pp. 77-78, 85-86. Here, a subset of the data is further analysed.

2.2. Sign flow segmentation

The sign flow of the dominant hand was segmented and labelled. The hand a signer uses to sign one-handed signs is called the dominant hand. Most people use their right hand but there are left-handed signers as well. A labelled sign was compared to the dictionary form. Thereafter, the sign was classified either as a function or a content sign. The data contained a total of 335 signs out of which 95 were function signs (28%).

The signs were divided into smaller constituents. We used a phonetic definition of a syllable although there is no generally accepted definition for a signed syllable. The signs were divided into syllables according to the following rules:

1. Each sign consists of at least one syllable.
2. If the sign can be decomposed to temporal segments each constituting an independent sign, then the sign is called a compound sign with each component being composed of at least one syllable; otherwise the sign is simple.
3. If the movement in the handshape flow is cyclic inside a simple sign or compound sign component then each cycle contains at least one syllable.
4. Every signed segment of an utterance belongs to at least one syllable.

The syllable boundaries were defined following the definition of a syllable. The boundaries of the signs were always syllable boundaries and the beginning of the new cycle (even if it remained incomplete) defined a syllable boundary. Only when two consecutive signs overlapped, the syllables could overlap. This could be a result of the two hands articulating different signs simultaneously or when the hand dominance switched temporarily to the other hand.

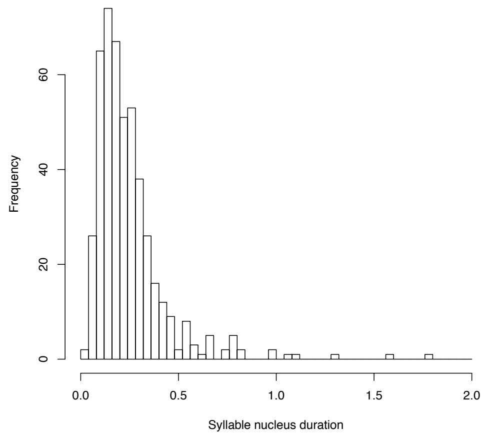

Figure 1: *A histogram of the syllable nucleus durations. The distribution is unimodal with mean 0.24 s, standard deviation 0.20 s, and median 0.21 s.*

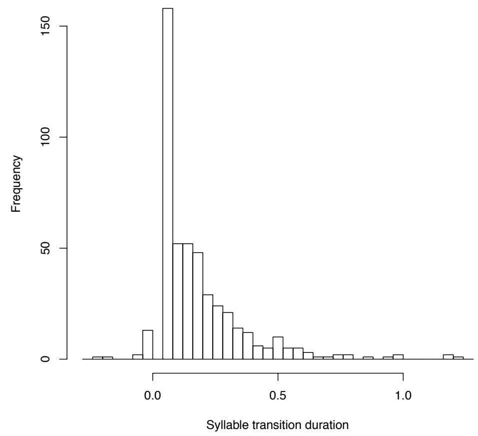

Figure 2: *A histogram of the syllable transition durations. The distribution is unimodal with mean 0.17 s, standard deviation 0.19 s, and median 0.13 s.*

The monosyllabic signs were most common (235 signs; 70% of the signs). In addition, there were 70 disyllabic signs (21%), 22 signs with three syllables (7%), 7 signs with four syllables (2%), and 1 sign with five syllables (<1%) The mean duration of a syllable was 0.41 s with standard deviation 0.30 s.

Every (phonetic) syllable was then further divided into a nucleus and a transition. The onset of the syllable nucleus was defined as the beginning of the syllable. The end point of the syllable nucleus, i.e. the beginning of the transition, was defined as the best match with the end handshape or secondarily the end position of the hand compared to the dictionary form of the syllable, and thirdly, as the first moment that clearly belonged to the movement towards the next syllable, i.e. the transition. The syllable nucleus durations are shown in Figure 1. The mean duration was 0.24 s and the standard deviation was 0.20 s.

The segments between consecutive nuclei were transitions (with one transition occurring between the nuclei – possibly with zero duration). Sometimes no transition took place (13 cases in the dataset). This was the case in polysyllabic signs consisting of a reduplicated simple sign like *sauna*. The transition could be even negative (4 cases in the dataset) when the end of a sign was reached (e.g. a hand-body contact ending a sign) after the handshape already had transformed to the beginning of the following sign. A histogram of the transition durations is shown in Figure 2. The mean duration of a transition was 0.17 s with standard deviation 0.19 s. The transition were in average somewhat shorter than the nuclei.

2.3. Statistical analyses

A mixed effects linear regression model was used to describe the variation in the duration of the syllable nuclei in the data set. The phonetic fixed variables were the duration of the adjacent transients and syllable nuclei; phonological fixed variables were the syllabic length of the sign, the syllable position within the sign (1 for the first syllable, 2 for the second, etc.), and the sign type (function sign or content bearing). The subjects were treated as random variables in the model.

3. Results

The minimal mixed effects model used to explain the syllable nucleus durations, which did not significantly differ from the full model with all the main effects (ANOVA), consisted of the adjacent syllable nuclei durations, the duration of the post-nuclear transition, and the type of the sign (function vs. content sign). The minimal model is described in Table 1. Syllable nuclei within a function sign were produced in average 56 ms faster in comparison to the syllable nuclei within a content sign. The overall signing rate was reflected by the strong impact of the previous and following syllable nucleus durations. Somewhat surprisingly then, only the post-nuclear transition duration had an impact on the nucleus duration (and not the pre-nuclear transition).

Although the syllable structure was taken into account by including the number of the syllables and position within a sign as fixed factors, this did not have any significant effect in explaining the nucleus durations. The phonological factors, syllable position and the syllabic length of the sign, did not explain the variation in the data, and were dropped from the minimal model.

Table 1: *The minimal mixed effects model fitted to the data to explain the variation in syllable nucleus durations.*

Fixed effect	Estimate	t value
Intercept (s)	0.12	7.1
Type (function sign)	-0.056	-2.9
Duration of the post-nuclear transition	0.21	4.6
Duration of the following nucleus	0.19	4.3
Duration of the previous nucleus	0.20	4.7

4. Discussion

Lexical factors affected the duration of the syllable nuclei of the signs even when the duration of the transitions were controlled for. The function signs were shorter possibly because of faster retrieval from the mental lexicon or more fluent production. This could be either because of the special syntactic role function words have in signing or it could be a consequence of function signs being more frequent in average than the content signs. As sign frequency is also a property of the sign, the results suggest that lexical properties influence the sign durations. In the current data, no attempt was made to control the sign frequency. Most likely the function signs in the current data set were more frequency than the content signs, which could have led to a confound. It is fully possible that the sign frequency *alone* would be sufficient to explain the sign type effect were it so that more frequent signs are produced faster *and* the function signs in the data set were more frequent. Unfortunately, the sign frequencies are not readily available for FinSL, yet.

The impact of adjacent constituents to the syllable nucleus durations is highly expected and it reflects to the local signing rate. In the results, there was an asymmetry in the influence of the transitions: only the post-nuclear transition had an significant effect on the nucleus duration while the effect of prenuclear transition was only half in size and did not reach significance (ANOVA for dropping the term, $p > 0.1$). The statistical model fitted to the current data set suggests the following asymmetry: for transitions that last longer, the preceding nucleus would be lengthened but not the nucleus following the transition. This could be a consequence of a trade-off between available time and necessary articulatory effort: for a longer distance between two consecutive nuclei more time could be allocated which could be compensated for by reduced articulation. Hence, the salience of the sign would be intact. Moreover, if a long distance within a transition is performed faster this would result in hyperarticulation. This trade-off then would move the boundary between the sign nucleus and the post-nuclear transition earlier during hyperarticulation. Importantly, the boundary is defined through the handshape which can be controlled rather independently compared to the position of the hand [22], [23].

Compared to an earlier analysis of the *syllable durations* by the current authors [10], the syllable nucleus durations show somewhat different behaviour. The syllable durations alone were shorter in polysyllabic signs than in monosyllables but here the syllable nuclei did not significantly differ as a function of the number of syllables. Also, function sign syllables were not significantly shorter than content signs although here the nuclei durations varied according to the sign type.

These opposing results can be reconciled. The bisyllabic signs often consist of reduplicating a monosyllabic sign leading to short (or non-existant) transitions. Indeed, in the current data, the sign internal transition are much shorter (mean 0.054 s) than the transitions between the signs (mean 0.22 s; Wilcoxon signed rank test, $p < 10^{-15}$). The transition durations themselves probably depend primarily on the spatial distance between the articulatory locations of the syllables, which is smaller sign internally. This remarkably shortens the duration of the syllables in polysyllabic signs. In addition to the evidence from the analysis of the last section, a direct comparison of the first syllable nucleus durations of mono vs. polysyllabic signs reveals no significant difference (Wilcoxon, $p > 0.1$).

In the current analysis of syllable nucleus durations, a relatively small effect of the sign type emerged although such an effect was not visible in the earlier syllable duration analyses. That the effect was not earlier visible could be simply due to larger variances in the syllable durations as opposed to the nucleus durations. In addition, it could be that only the nucleus durations are affected by the sign type and the duration of the transitions is solely determined by the spatial relations of the adjacent syllables.

The distributions of the durations for syllables, syllable nuclei, and syllable transitions were all non-normal. The logarithms of the respective durations show less non-normal shapes but are still clearly (QQ-plots, normality tests) different from normal. To diminish the possibility that the results would be an artifact of the non-normality of the data, we built a minimal mixed effect model for the logarithmically transformed durations. The resulting minimal model has exactly the same terms and the statistical significance of the individual factors is improved. This hints that the non-normality is not the main source of the results.

During spoken word production, frequent words tend to be produced faster when controlled for phonological length and the frequency of the word [24]. This is true for signs as well (at least what comes to the phonological length) as the syllables in the polysyllabic signs are faster [10]. The current work suggests that this phenomenon might be completely explained in the sign language context by the spatial relations of the sign components (syllables). An alternative explanation that the phonologically longer signs are produced faster because of repetition induced optimization of the syllable sequences within a sign as opposed to across sign boundaries, is not supported by the current sign language data.

5. Conclusion

In the current work we demonstrated using Finnish Sign Language that even in a sign language lexical factors affect the duration of signed segments. In addition, the syllable nucleus durations are affected on the phonetic level by the durations of adjacent nuclei and the postnuclear transition but not by the prenuclear transition. Hence, the current data analysis suggests that the transitions are more strongly attached to the signs (syllables) preceding the transitions. In addition, function signs tended to have faster syllable nuclei in the data set similar to the spoken language results. This points to a sign production model where the spatial relation of the signs (and sign syllables) determines the transition durations which in turn affect the nucleus durations alongside with the lexical properties of the sign.

6. Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement n° 287678. The authors would like to thank the participants of the study.

7. References

[1] M. Nespor and W. Sandler, "Prosody in Israeli sign language," *Language and Speech*, vol. 42, no. 2-3, pp. 143–176, 1999.

[2] R. B. Wilbur, "Stress in ASL: Empirical Evidence and Linguistic Issues," *Language and Speech*, vol. 42, no. 2-3, pp. 229–250, 1999.

[3] R. B. Wilbur and S. B. Nolen, "The duration of syllables in American Sign Language," *Language and speech*, vol. 29, no. 3, pp. 263–280, 1986.

[4] D. Brentari, *A prosodic model of sign language phonology*. The MIT Press, 1998.

[5] S. K. Liddell and R. E. Johnson, "American Sign Language: The Phonological Base." *Sign language studies*, vol. 64, pp. 195–278, 1989.

[6] B. J. Kröger, P. Birkholz, J. Kannampuzha, E. Kaufmann, and I. Mittelberg, "Movements and holds in fluent sentence production of Aerican Sign Language: The action-based approach," *Cognitive computation*, vol. 3, no. 3, pp. 449–465, 2011.

[7] M. Blondel and C. Miller, "Movement and rhythm in nursery rhymes in LSF," *Sign Language Studies*, vol. 2, no. 1, pp. 24–61, 2001.

[8] S. Ojala, "Rytmin ja koartikulaation vaikutus viittomiin ja puheeseen. The effect of rhythm and coarticulation on sign and speech." *Jantunen, Tommi (ed.) Näkökulmia viittomaan ja viittomistoon. Jyväskylä: Jyväskylän yliopisto*, pp. 29–43, 2010.

[9] R. Sutton-Spence and B. Woll, *The linguistics of British Sign Language: an introduction*. Cambridge University Press, 1999.

[10] D. Aalto and S. Ojala, "The duration of syllables in spontaneous Finnish Sign Language," in *Nordic Prosody, Proceedings of the XIth Conference, Tartu, 2012*, E.-L. Asu and P. Lippus, Eds. Peter Lang, 2012, pp. 39–47.

[11] S. Ojala, T. Salakoski, and O. Aaltonen, "Coarticulation in sign and speech," in *actes de NODALIDA 2009 workshop Multimodal Communication, from Human Behaviour to Computational Models, Costanza Navarretta, Patrizia Paggio, Jens Allwood, Elisabeth Alsén and Yasuhiro Katagiri (Eds). NEALT Proceedings Series*, vol. 6, 2009, pp. 21–24.

[12] S. K. Liddell, *American sign language syntax*. Mouton The Hague, 1980.

[13] R. E. Johnson and S. K. Liddell, "A segmental framework for representing signs phonetically," *Sign Language Studies*, vol. 11, no. 3, pp. 408–463, 2011.

[14] S. Kita, I. Van Gijn, and H. Van der Hulst, "Movement phases in signs and co-speech gestures, and their transcription by human coders," in *Gesture and sign language in human-computer interaction*. Springer, 1998, pp. 23–35.

[15] C. P. Browman and L. Goldstein, "Articulatory phonology: An overview," *Phonetica*, vol. 49, no. 3-4, pp. 155–180, 1992.

[16] A. M. Liberman and I. G. Mattingly, "The motor theory of speech perception revised," *Cognition*, vol. 21, no. 1, pp. 1–36, 1985.

[17] E. R. Kandel, J. H. Schwartz, T. M. Jessell *et al.*, *Principles of neural science*. McGraw-Hill New York, 2000, vol. 4.

[18] S. Ojala, "Towards an Integrative Information Society: Studies on Individuality in Speech and Sign," Ph.D. dissertation, University of Turku, 2011.

[19] B. Lindblom, C. Mauk, and S.-J. Moon, "Dynamic specification in the production of speech and sign," *Dynamics of Speech Production and Perception*, vol. 374, p. 7, 2006.

[20] B. Fuchs, "Phonetische Aspekte einer Didaktik der Finnischen Gebärdensprache als Fremdsprache," Ph.D. dissertation, University of Jyväskylä, 2004.

[21] S. Ojala, "Rytmi puheessa ja viittomisessa," *XXVI Fonetiikan päivät 2010*, pp. 47–49, 2010.

[22] B. Lindblom, "Explaining phonetic variation: A sketch of the H&H theory," in *Speech production and speech modelling*. Springer, 1990, pp. 403–439.

[23] ——, *Economy of speech gestures*. Springer, 1983.

[24] A. Bell, J. M. Brenier, M. Gregory, C. Girand, and D. Jurafsky, "Predictability effects on durations of content and function words in conversational English," *Journal of Memory and Language*, vol. 60, no. 1, pp. 92–111, 2009.