Author(s): Hernández, Doris; Puupponen, Anna; Keränen, Jarkko; Wainio, Tuija; Pippuri, Outi; Ortega, Gerardo; Jantunen, Tommi

Title: Use of Sign Language Videos in EEG and MEG Studies : Experiences from a Multidisciplinary Project Combining Linguistics and Cognitive Neuroscience

Year: 2022

Version: Published version

Please cite the original version:

Hernández, D., Puupponen, A., Keränen, J., Wainio, T., Pippuri, O., Ortega, G., & Jantunen, T. (2022). Use of Sign Language Videos in EEG and MEG Studies : Experiences from a Multidisciplinary Project Combining Linguistics and Cognitive Neuroscience. In J. H. Jantunen, J. Kalja-Voima, M. Laukkarinen, A. Puupponen, M. Salonen, T. Saresma, J. Tarvainen, & S. Ylönen (Eds.), Diversity of Methods and Materials in Digital Human Sciences : Proceedings of the Digital Research Data and Human Sciences DRDHum Conference 2022, December 1-3, Jyväskylä, Finland (pp. 148-155). Jyväskylän yliopisto. http://urn.fi/URN:ISBN:978-951-39-9450-1

Use of Sign Language Videos in EEG and MEG Studies. Experiences from a Multidisciplinary Project Combining Linguistics and Cognitive Neuroscience

Doris Hernández,[1] Anna Puupponen,[1] Jarkko Keränen,[1] Tuija Wainio,[1] Outi Pippuri,[1]
Gerardo Ortega,[2] Tommi Jantunen[1]
[1]Sign Language Center, Department of Language and Communication, University of Jyväskylä
[2]University of Birmingham
E-mail: doris.m.hernandez-barros@jyu.fi

Abstract

In this paper, we describe our experiences of bringing together methodologies of two disciplines – sign language (SL) linguistics and cognitive neuroscience – in the multidisciplinary ShowTell research project (Academy of Finland 2021–2025). More specifically, we discuss the challenges we encountered when creating and using video materials for the study of SL processing in the brain. Rather than using still images, the study of SL comprehension is better performed by using videos, thus providing more naturalistic stimuli as observed in face-to-face interaction. On the other hand, in neuroimaging (electroencephalography [EEG]/magnetoencephalography [MEG]), it is vital to track the timing of the stimulation exactly and to minimize the noise that could arise from inside and outside the brain. Any brain activity not related to the specific aspect being studied could create artifacts that diminish the signal-to-noise ratio of the measurements, thus compromising the quality of the data. This creates significant challenges when integrating both disciplines into the same study. In the paper, we (i) describe the process of, and requirements for, creating signed video materials that try to mirror naturalistic signing; (ii) discuss the problems in the synchronization of the video stimuli with the brain imaging data; and (iii) introduce the steps we have taken to minimize these challenges in different phases of the process, such as the design, recording, and processing of the video stimuli. Finally, we discuss how, with the use of these steps, we have been able to deal successfully with the resulting data and creating materials that integrate the naturalistic nature of human communication.

Keywords: sign language, multidisciplinary, neurolinguistics, video stimuli

1. Background

In sign languages (SL), linguistic messages are visually captured based on movements of the hands and other parts of the body. This makes studying SLs different from studying spoken (oral-auditory) languages: the difference in the modality requires different methodologies in parts, although ultimately, both languages may be connected to the same biological and cognitive underpinnings. Historically, SLs are not directly related to spoken languages, but in the modern and global world, most SLs are in contact with the surrounding spoken languages of ambient society. Today, SLs are not only used by deaf people, but also by hearing native signers and hearing second language learners who are in contact with deaf societies. Furthermore, many deaf people sometimes use many languages, both other sign languages and/or written languages.

The study of languages such as SLs is a complex process that involves diverse perspectives such as linguistic, cognitive, and social viewpoints. Thus, it might be better approached in a multidisciplinary manner by bringing together the diverse approaches of different disciplines. But combining two distinct disciplines is not always easy due to differences in research histories, paradigms, methodologies, concepts, theoretical starting points, etc. This has been experienced in the research project "ShowTell – Showing and telling in Finnish Sign Language" (jyu.fi/showtell), funded by the Academy of Finland 2021–2025. The ShowTell project investigates how showing meaning is connected to telling meaning in Finnish Sign Language (FinSL) by analyzing the relationship between bodily enactment and traditional language use with lexical items. The project addresses this issue by approaching it from the perspectives of language use (FinSL corpus data, Study 1), kinematic movement production (motion capture data, Study 2), and brain-based meaning processing (functional neuroimaging data, Study 3). During the first year of the project, the research material for the Study 3 (signed sentences and individual signs) was built and tested. The content of the current paper is based on these processes.

During the creation of the research material, the combination of two related but distinct disciplines – SL linguistics and cognitive neuroscience – resulted in significant challenges for the completion of this stage (study planning and the creation of stimuli). Although linguistics and cognitive neuroscience might study the same phenomenon, they might also use different traditions and paradigms that can impose the use of conflicting requirements on the stimulus materials to be used. How, then, can the multidisciplinary integration of both methods be approached for a more integrated study of SLs? The aim of this article is to document the experiences of combining SL video materials and cognitive neuroscience methodologies. We would like to highlight that this is not a report of an original research, but rather an article where we document our experiences. The specific challenges we have encountered will be described in the next sections as well as the solutions we have implemented for each of them. Finally, we provide an example of testing the implementations in the pilot results.

2. Doing SL Linguistics from a Cognitive Neuroscience Perspective

At the beginning of the modern study of SLs in the 1960s and 70s, analysis was restricted to observing unrecorded SL use, still images, or textual descriptions of signs. As video technologies have advanced, studies on different aspects of SLs have expanded. The analysis of SL structure and use, for example, has unprecedented possibilities due to the accessibility of high-quality recording equipment and the new corpus infrastructures that have been built during the last decades (Orfanidou, Woll & Morgan 2015; Salonen, Kronqvist & Jantunen 2020). This development is essential for the study of visual - gestural languages. To understand how SLs are produced, what they are, and how they work, analysis needs to be done based on formats, such as video, that capture the characteristics of the three-dimensional and multimodal articulation of signers as effectively as possible. In data-driven work on different fields of SL studies, this is a rather self-evident fact – a starting point, so to speak.

The same also applies to the study of SL processing. To understand the perception and cognitive processing of an SL, the stimuli should be as close as possible to the ways in which SLs are perceived in real life. SL do not have a writing system so psycholinguistic and neurocognitive studies have to resort to videos. While using still images or videos of single lexical signs as stimuli is a good starting point and ensures that the stimulation is as controlled as possible, our understanding from the development of the field (Hernández, Puupponen & Jantunen 2022) is that we should go more towards naturalistic signing in creating the stimuli. This includes, for example, the use of signed sentences recorded on the video.

In cognitive neuroscience, brain signals are usually recorded (with electroencephalography [EEG] or magnetoencephalography [MEG]) simultaneously, while the participants solve a cognitive task or process stimuli that are specifically designed. Cognitive neuroscience paradigms/tasks tend to compare different types of stimuli under different conditions. Later, differences (such as amplitude or latency) in each component or time window might reflect processing differences. This allows researchers to identify functional brain-related markers for processing a task or stimulation (see Figure 1). EEG and MEG rely on synchronized electrical activity (and the magnetic field arising from it) coming from large groups of neurons with a specific orientation with respect to the scalp (Luck 2014). Electricity is conducted swiftly; thus, the signal is particularly sensitive to changes that occur over time.

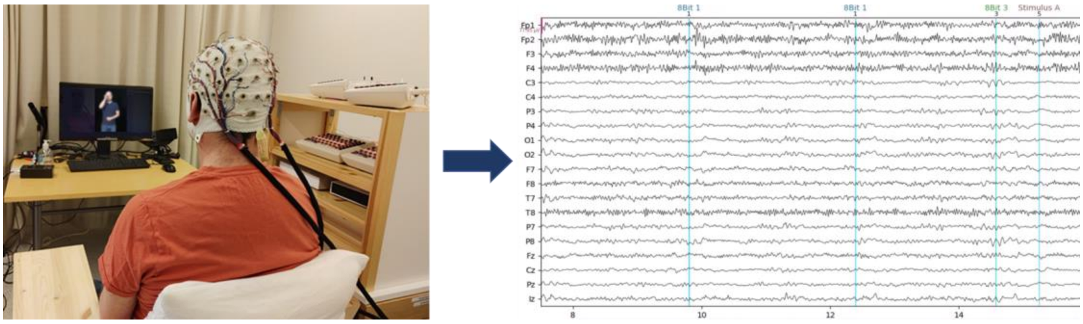

Figure 1. Schematic representation of the EEG recording process.

In EEG and MEG recordings, the temporal characteristics of the stimulus require particular attention. Every change in the stimulus that can influence neurocognitive processing should be properly and accurately marked (in the order of milliseconds [ms]) to allow its link with brain data. Different stimuli elicit different responses (components), whose amplitudes might reflect the degree of neural engagement in different cognitive processes. Furthermore, the stimuli to be used need to be carefully created. Frequently, simplified stimulation (such as geometric figures for visual or simple tones for auditory stimulation) is used to ensure that their physical properties can be controlled as much as possible.

Video stimuli have been used in EEG/MEG studies involving sign languages (Baus, Gutiérrez & Carreiras 2014; Hosemann et al. 2013), but their adaptation to the techniques' requirements is still challenging. The complexity of the video material (compared to still images) adds extra issues to be controlled. The complexity of videos can, for instance, increase the noise that needs to be avoided to record reliable brain data. As the size of the activity arising from the brain is miniscule (in the order of microvolts [µV in EEG]/femtotesla [fT in MEG]), competing electrical activity coming from the body (such as eye movements or heartbeat) or externally (such as electrical activity from other devices needed in the room during the measurements, see Figure 1) is considered noise. In videos, variations such as changes in luminance or color contrasts, can involve additional brain areas and processes to the ones that are intended to be studied. Brain activity is also considered internal noise. Preferably, any kind of noise should be reduced as much as possible to obtain high-quality data.

More problems arise with videos showing human actions, as in the case of signing. Specifically, for SL research, it is challenging to mark very specifically the phases in the production of signs, such as the offset of the preparation phase, the recognition point, or the onset of meaningful parts of signs (see Table 1). This occurs because the borders of these phases are gradient (i.e. not categorical) and may happen at different times across the videos (Jantunen 2015). ERPs can be locked (i.e. onset/zero-point of ERPs) based on these sign production phases (see Table 1) to identify their related brain processing (Emmorey, Midgley & Holcomb 2022). In addition to the difficulties in identifying signs' parts, the exact same duration in all the videos (and signs) cannot be ensured, which can also elicit changes in brain processing (Skukies & Ehinger 2021) and internal noise.

All signs include semantic information during the preparation phase (for a comprehensive review, see Jantunen 2015). This is because the production of the formational components of the sign must begin before the onset of the stroke, the most meaningful phase in sign production. In particular, the parameters of handshape, orientation, and nonmanual elements provide information that reveals the meaning of the whole sign before the stroke. This happens both with isolated signs as well as in connected signing.

Specifically, when planning/creating the stimuli, we identified two main challenges that needed to be solved before the actual measurements could start. First, the higher complexity of the video material might increase the physiological noise in the EEG signals. Second, the differences in duration between the videos could create large variability within and across conditions beyond the ones that are intended to be studied. If these problems remain unresolved, they can induce effects that can be camouflaged within the studied signal and lead to misguided results and conclusions.

Home-Excursion	**Movement phase**	
Home position	-	
Excursion	Liberation	
	Preparation	
	The most expressive phase of the sign	Pre-stroke hold
		Stroke
		Post-stroke hold
	Recovery	
Home position	Settling	
	-	

Table 1. Motion (excursion) and movement phases of an individual, isolated sign/gesture. All phases are optional, but signs/gestures typically have Preparation, Stroke and Recovery. The table has been adapted from Arendsen, van Doorn & de Ridder (2007: 317). For visualizations of the phases, please see Jantunen (2015).

3.1. Describing the Procedure

To test how SL is processed at the neural level, EEG measurements need to be taken while participants watch signing situations at a natural pace. As mentioned before, videos need to be carefully created to fulfil several conditions arising from the fields of linguistics and cognitive neuroscience (e.g., visual complexity, luminance, length), while still ensuring the naturalistic nature of the communication process. In the next subsections, we explain the procedure in each step (design, recording, and processing) of preparing the video material that could be used as stimuli in a study on SL processing.

3.1. Stimulus Design, Recording, and Processing

As mentioned before, some physical properties of the videos can create significant processing differences at the brain level that are unrelated to the conditions studied. For this reason, luminance conditions and colors should be kept constant across the videos. Additionally, at a psychological level, other characteristics of the videos could influence the EEG results. To avoid this, we recommend including only one signer, wearing the same clothes, and signing in front of the same background across the videos. The signer should stand in a fixed location so that, at the beginning and end of each stimulus, the location of the signer on the screen is the same. In addition, we do not recommend using a black background because it may cause reflection on the viewer's monitor while doing a measurement. This last recommendation is based on the participants' feedback as a black background was used in our study. We cannot say that the black background would affect the EEG measurements, but it might affect the participant's well-being during the measurements. In our opinion, a light grey background color would work better for this purpose.

Other SL-related characteristics included in the video should be carefully planned. For example, the signing rate (signs per second) should be natural and approximately similar among different stimuli. Articulation should be controlled when it comes to the use of, for example, different facial expressions, mouth actions, and body movements. In addition, the signer's hands and body should be brought back to an identical neutral position after the signing of each stimulus and left there for at least three to five seconds before continuing the signing. In this way, later in the editing phase, this part can be shortened according to the requirements of the study. Not many more characteristics of the stimuli need to be controlled if they comprise single signs, as no other signs are included in the conditions of the study.

However, if the video stimuli include signed sentences that are, for example, repeated under different conditions, the video materials should be edited so that they are as similar as possible. For example, signers can be asked to watch sentences that are either correct (Condition 1) or semantically incorrect (Condition 2). Then, the brain responses to each kind of sentence are compared. For this purpose, the only thing that should vary between the conditions (kinds of sentences) is the object of the study. In other words, the sentences should be edited so that the same material is used for the unchanging parts of the sentences (i.e., the *sentence frame*), while only the target varies. In addition, when recording varying targets, the signer should produce them in the whole sentence. In this way, the target sign's articulation will be natural in its preparation and offset phases so that it can be edited (if needed) to the sentence frame with a naturalistic result.

Editing should also be considered when choosing the signs for the sentences. For example, large differences in the vertical locations of the signs (hands) before and after the editing point should be avoided. In addition, signs produced before the target should not include a highly repetitive or otherwise varying movement trajectory or a less fixed vertical location in the signing space (e.g., signs CAR, SEARCH, PEN in FinSL; Finnish Signbank, the University of Jyväskylä, Sign Language Center 2018). Furthermore, the aperture of the eyes should be controlled so that it does not vary drastically before and after the editing point. These characteristics can cause a stronger glitch at the editing point of the stimulus. Instead, signs that end with a clear hold (with or without contact with the signer's body) and/or signs that are produced in a relatively fixed location on the signer's body or the signing space are suitable options to be used prior to the editing point (e.g., signs GRANDMA, WOMAN, SHOES in FinSL). Using these types of signs ensures that the position of the signer's hand(s) is as similar as possible between the sentence frame and the utterance from which the target sign is added to the frame. This ensures that the editing of the video results in as natural a result as possible. Finally, if the stimuli also include sentences that do not need editing, an artificial glitch (a point that visually seems like an editing point) should be added to those to avoid variation between the conditions that can affect the results.

All in all, the signer needs to pay attention to many aspects of the contents and articulation of signing while producing the materials, so visual aids – such as a prompter or a large screen – should be used in the recording situation to maximize the control of the variation in the signing. As expressed before, when SL video material is going to be used for research in the linguistic field, the video start and end points need to show how the hands rise from or return to the initial position. Typically, to get a naturalistic result, videos are edited so that they include at least 15–20 frames before and after the production of the sign/sentence. For cognitive neuroscience, that time is significantly longer, based on the high temporal resolution of the functional techniques (EEG and MEG). When creating SL-related videos to be used in cognitive neuroscience studies, our recommendation is to reduce that time as much as possible while keeping the initial position of the signer at the beginning and the end of the videos. In this way, the beginning of the video closely represents the beginning of the sign. We have found that the production of the sign can start as early as three frames after the start of the video while not having a negative effect on the comprehension of the signs.

3.2. Temporal Alignment of Video and EEG Data

After the video stimuli are recorded and edited, they should be synchronized with the EEG/MEG data. The timing properties of the videos are crucial to successfully performing this task. Depending on the research questions and analysis methods, not only could the beginning of the videos be important, but also other components of the signs, such as signing onset and offset times, and duration.

However, what happens if the phenomenon studied includes duration differences? As this was our case while studying semantic processing at the brain level, we tried to ensure that while the videos showed differences in duration across conditions, the stroke (the expressive part of the sign, according to Kita, Gijn & Hulst 1998) onset across conditions was not significantly different. Thus, the beginning of the stroke marks the “definite beginning of meaning.” However, in general, the meaning is already comprehensible during the preparation phase preceding the stroke (Jantunen 2015). This has also been confirmed by some cognitive neuroscience studies looking at prediction in signed sentences (Hosemann et al. 2013). If the beginning of meaning does not differ between conditions, even though they have differing durations, their semantic processing at the brain level can be studied with stimulus-locked event-related potentials (ERPs; for more information, see Hernández, Puupponen & Jantunen 2022). This is because the semantic processing at the brain level should be completed around 500–550 ms after the beginning of meaning (around 700–800 ms after the video onset) at the sign level. This time is fairly before the whole sign is performed (around a couple of seconds after the video onset), according to previous studies performed in SL research (Emmorey, Midgley & Holcomb 2022; Ortega, Özyürek & Peeters 2020).

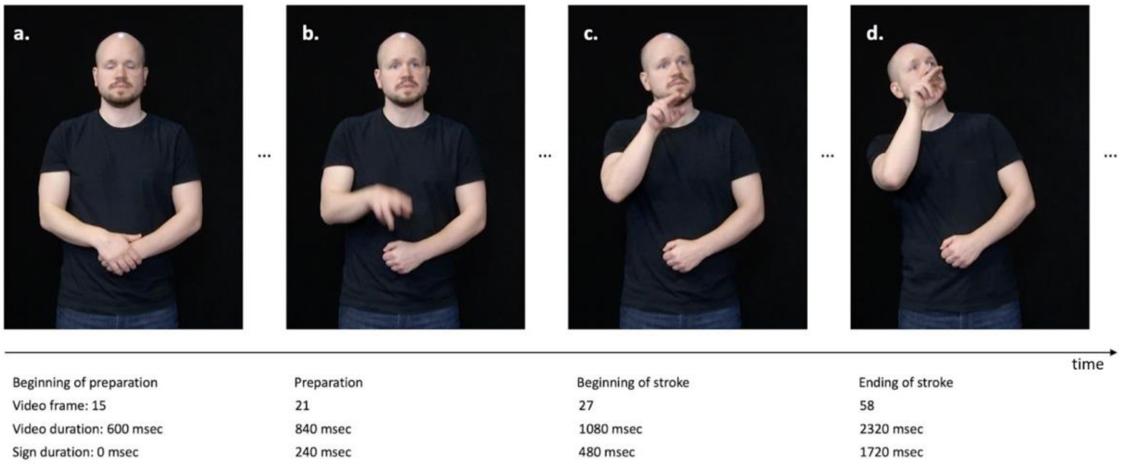

Figure 2. Sign phases (see Table 1) demonstrated with frames (25 fps) extracted from the stimulus video sign LOOK produced with simultaneous enactment

We identified the beginnings of meaning with the help of strokes from the finished video stimulus material. In practice, the beginning of the stroke corresponds to the video frame in which the hands’ movement first changes its direction toward the phonemic location of the sign (e.g., Kita, Gijn & Hulst 1998; Jantunen 2015). In Figure 2, this is frame c. From the stroke onset frame, we then moved backward frame by frame toward the moment when the phonemic handshape of the sign was first identifiable. This moment typically occurred with a shift in the eye gaze away from the camera. In Figure 2, this moment of meaning beginning is identified in frame b.

When identifying the sign phases, it is essential to be able to play the video frame by frame and mark the frame number from the video onset. After the frame one is looking for has been identified (e.g., frame number 27), it can be converted to ms by multiplying the frame number with the 1,000 ms per used frame per second (fps) ratio (e.g., 27 frames x 1,000 ms/25 fps = 1,080 ms, assuming that EEG equipment and camera clocks are synchronized).

4. Combining SL Linguistics and Cognitive Neuroscience in a Pilot Measurement

Previously described measures were implemented in an oddball paradigm (for more details about this kind of paradigm, see Hernández, Puupponen & Jantunen 2022) and tested with an L1 signer pilot participant. In the task, videos of signs (ranging in duration from 1.5 to 2 seconds) were serially presented in random order. Thirty videos showing lexical signs were frequently (82% of probability) presented, while the other 30 videos showing lexical signs with enactment were rarely (18%) presented, with an interval of 500 ms between them. For signs without enactment, the meaning onset started at 320 ms, while for signs with enactment, the meaning onset started at 360 ms. No significant differences were found between the meaning onset times of either condition (frequent and infrequent signs). The participants were asked to press a button as soon as they noticed that "something else" was added to the signs. The resulting EEG responses are shown in Figure 3. The expected ERP (P3b) was visible in the waveforms for the infrequent signs peaking around 710 ms from the video onset and around 350 ms from the beginning of the meaning onset, thus showing its typical characteristics (Hernández, Puupponen & Jantunen 2022).

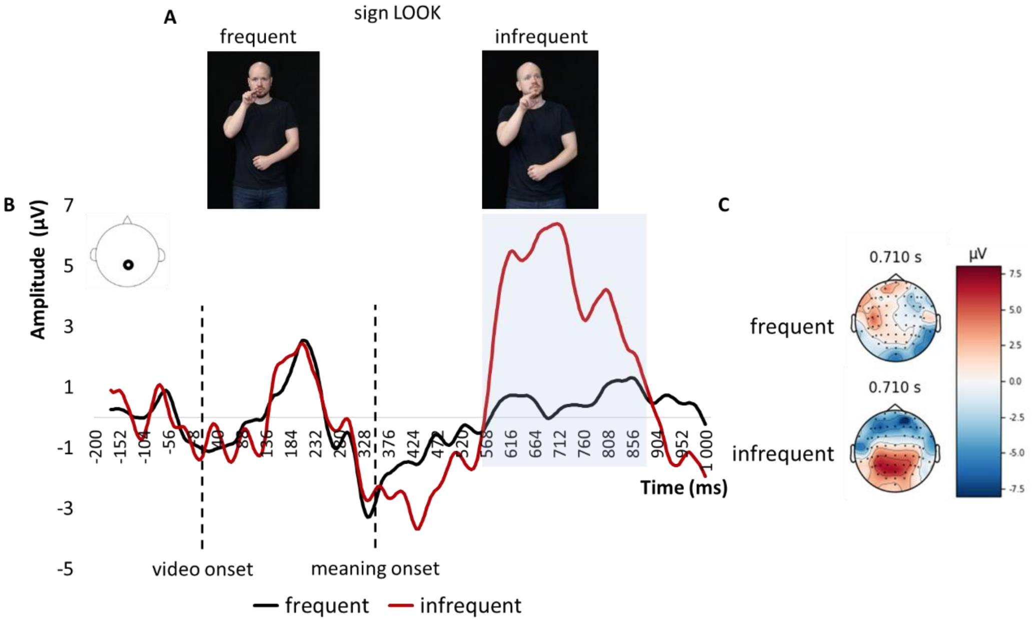

Figure 3. Brain responses from a pilot participant to the described task. A) Example of the kinds of signs used as frequent and infrequent. B) Average waveforms for frequent (black) and infrequent (red) signs in the centroparietal electrode (Pz). The amplitude of the waveforms (in µV) is shown on the y axis, while the latency (in ms) is shown on the x axis. P3b is highlighted in grey. The dashed lines represent the video onset at 0 ms and the meaning onset at 360 ms from the video onset. C) Topographic distribution in the scalp of P3b around 710 ms for the frequent and infrequent signs.

While applying the suggested procedure described in the current paper, we were able to effectively record the intended ERP (P3b). We interpret this as proof that the procedure is successful.

5. Conclusion

This paper has shown our experiences while preparing the stimulus material for the EEG substudy of the ShowTell project. The key challenges (video complexity and duration variability) that we encountered during this research stage were described according to each stage (planning, recording, and processing) of the video stimulus creation process. The solutions given for each of these problems have been described and proved to be successful for our study. Therefore, we are not claiming that our solutions would be the only or the best ones. We hope these will be reviewed and developed in future studies.

6. Acknowledgements

Funding from the Academy of Finland under Project 339268 (ShowTell) is gratefully acknowledged. The authors wish to thank Josh Seligman for the English Language checking of this manuscript.

References

Arendsen, J., van Doorn, A. J., & de Ridder, H. (2007). When and how well do people see the onset of gestures? *Gesture* 7(3), 305-342.

Baus, C., Gutiérrez, E., & Carreiras, M. (2014). The role of syllables in sign language production. *Frontiers in Psychology* 5, 1254.

Emmorey, K., Midgley, K. J., & Holcomb, P. J. (2022). Tracking the time course of sign recognition using ERP repetition priming. *Psychophysiology* 59 (3) e13975.

Hernández, D., Puupponen, A., & Jantunen, T. (2022). The contribution of event-related potentials to the understanding of sign language processing and production in the brain: Experimental evidence and future directions. *Frontiers in Communication* 40.

Hosemann, J., Herrmann, A., Steinbach, M., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Lexical prediction via forward models: N400 evidence from German Sign Language. *Neuropsychologia* 51 (11), 2224–2237.

Jantunen, T. (2015). How long is the sign? *Linguistics* 53 (1), 93–124.

Kita, S., Gijn, I. V., & Hulst, H. V. D. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In International Gesture Workshop. Springer, Berlin, Heidelberg, 23–35.

Luck, S. J. (2014). An introduction to the event-related potential technique. MIT Press.

Orfanidou, E., Woll, B., & Morgan, G. (2015). *Research Methods in Sign Language Studies*. West Sussex: John Wiley & Sons, Inc.

Ortega, G., Özyürek, A., & Peeters, D. (2020). Iconic gestures serve as manual cognates in hearing second language learners of a sign language: An ERP study. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 46 (3), 403.

Salonen, J., Kronqvist, A., & Jantunen, T. (2020). The corpus of Finnish Sign Language. In Efthimiou, E., F. Stavroula-Evita, T. Hanke, J. Hochgesang, J. Kristoffersen, & M. Mesch (eds.), *Proceedings of the 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*. Paris: European Language Resources Association (ELRA), 197–202.

Skukies, R., & Ehinger, B. (2021). Modelling event duration and overlap during EEG analysis. *Journal of Vision* 21 (9), 2037–2037.

The University of Jyväskylä, Sign Language Centre (2018). Finnish Signbank. Available in the Language Bank of Finland (Kielipankki). https://signbank.csc.fi. [Last accessed 5.3.2022].