Decomposing Sign Language Movements: A Multi-Band Visualization Method for Articulatory Analysis

Antonio F. G. Sevilla^{1,2}, **José María Lahoz-Bengoechea**^{3}
^{1}Software Engineering and Artificial Intelligence
^{2}Institute of Knowledge Technology
^{3}Spanish Linguistics and Literary Theory
Universidad Complutense de Madrid
`afgs@ucm.es`, `jmlahoz@ucm.es`

Abstract

Understanding the structure of sign language movements requires methods that can isolate and analyze the hierarchical and simultaneous nature of sign articulation. We present a method for tracking and visualizing sign language movements that progressively isolates dependent movements within the articulatory chain: hand rotation from arm displacement and finger movement from hand movement. Using MediaPipe hand tracking on ordinary 2D video, we decompose motion into separate gestural components and compute velocity and direction for each articulator. We present these movement channels in a time-aligned multi-band visualization that reveals temporal structure, bimanual synchronization patterns, and the coordination of different articulatory components. An interactive web-based viewer synchronizes the visualization with video, enabling researchers to efficiently explore movement patterns and their relationship to signing. We demonstrate the method with examples from isolated signs and continuous signing, showing how it reveals patterns that are difficult to observe in raw video, including bimanual coordination, internal movements, and the distinction between linguistic and non-linguistic segments. This approach provides accessible tools for empirical investigation of rhythmic and prosodic patterns in sign languages.

Keywords: Sign Language Recognition/Generation, Less-Resourced/Endangered Languages, Tools, Systems, Applications

1. Introduction

It is widely recognized that sign languages are not mere codifications of the spoken word in the visuo-gestual modality. They have their own grammar and vocabulary, and beyond that, a particular and very different internal parametric structure (Stokoe, 1960; Brentari et al., 2018). These parameters include hand shape, orientation, location, and one of especial interest to us: movement. The variety and complexity of movement analysis in sign languages has seen an evolution from holistic all-in-one movement parameters to more fine-grained analysis (Liddell and Johnson, 1989; Koech, 2007), to modern machine learning (Börstell, 2023) and statistical modeling (Chassat et al., 2022).

Crucially, sign language movements compose in complex ways. There can be more than one articulatory gesture occurring simultaneously, for example a rotation of the wrist while displacement is happening, or the repetition of some gesture in a linguistically-determined pattern (Aronoff et al., 2005). Understanding this complexity is essential for addressing fundamental questions in sign language phonology. Researchers seek to compare the rhythms and sequencing of movements in sign languages to oral articulatory gestures (Tyrone et al., 2010), and the existence or not of a “syllable” in sign languages, and its exact nature, remain matters of ongoing scientific inquiry (Stack,

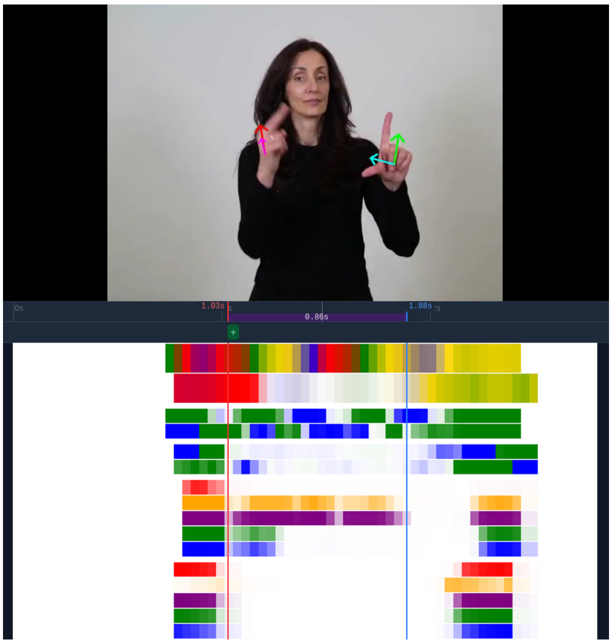


Figure 1: Our viewer showing an analyzed sign language video, with the tracking of hand position and orientation vectors displayed. The generated bandogram shows movement in different channels while synchronized with the video, allowing the user to see each segment in action. A detailed explanation is provided in section 4. The viewer code can be freely accessed online at `https://github.com/agarsev/bandogram`.

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 9559–9568
11–16 May 2026. ©ELRA Language Resources Association (ELRA), 2026

1988; Wilbur and Allen, 1991).

Addressing these questions requires empirical methods that can reveal the hierarchical and simultaneous nature of sign language articulation. Recent advances in automatic video tracking enable recovery of detailed movement information from recorded signing (Rodríguez et al., 2020), but such methods should incorporate further processing in linguistically-motivated ways (Yin et al., 2021). Specifically, extracted parameters must align with articulatory structure, for example, separating hand rotation from arm displacement, and finger movement from hand movement. Additionally, researchers need tools to explore movement data interactively, examining the relationship between quantitative profiles and the visual signing they represent. Such tools can help address a variety of empirical questions; for example, which movements constitute the lexical content of a sign as opposed to preparation and relaxation gestures, how sign syllables are constructed from simultaneous articulatory layers, or whether and how prosody affects signing speed and rhythm.

In this article, we present such a method. We use MediaPipe (Lugaresi et al., 2019) to track hand movement in sign language video, and then progressively isolate dependent movements within the articulatory chain: hand rotation from arm movement and finger movement from hand movement. We present velocity and direction of movements for each articulator in separate, time-aligned tracks in a multi-band visualization that we call a *bandogram*, which makes structure, synchronization, and rhythmic patterns visually apparent, and we further provide an interactive viewer that allows researchers to click and interact with the velocity plots to see the video segments these patterns represent. An example screen capture of the viewer can be seen in Figure 1. We believe that this combination of linguistically-motivated decomposition and interactive visualization can become a useful tool for the empirical investigation of sign language structure.

The remainder of this paper is organized as follows: section 2 reviews related work in sign language movement analysis and automatic tracking methods; section 3 describes our methodology for tracking, decomposing, and visualizing sign language movements; section 4 presents examples demonstrating the capabilities of our method and tools; and section 5 discusses the implications and future directions of this work.

2. Related Work

Understanding the structure of sign language movements has long been a goal of sign language linguistics and computational modeling. Various approaches have been developed to analyze and decompose the complex, multi-dimensional movements involved in signing. For an example of such movements, an example can be seen in Figure 2[1].

Figure 2: A diagram showing the complex articulation of a sign in Spanish Sign Language (Cabeza Cuadrada, “Block-head”). In this sign, there are both external movements of the hand (vertical and horizontal displacements) and internal ones, with a change of finger configuration in the second “syllable” and rotation of the dominant hand wrist.

2.1. Sign Language Phonology and Movement Structure

The displacement and velocity profiles of hand movements during signing have been identified as key metrics for analyzing prosodic structure at both lexical and post-lexical levels (Wilbur and Martinez, 2002). Changes in velocity and direction can demarcate segment transitions in phonological structure, reflecting the hierarchical organization that is essential for conveying meaning in sign languages, analogous to prosody in oral languages (Fenlon and Brentari, 2021). These kinematic features (articulator displacement and its associated attributes of duration and velocity) are posited to sufficiently encapsulate the dynamic structure of sign language (Wilbur and Martinez, 2002).

Understanding the internal structure of signs requires methods that can reveal the coordination and timing of different articulatory components. Liddell and Johnson (1989) proposed a framework structuring sign language movements into segments characterized by periods of movement and hold. However, movement in signing is multi-dimensional and hierarchically organized. A single sign may involve simultaneous arm displacement, wrist rotation, and finger configuration changes, each potentially following independent temporal patterns. This complexity motivates approaches that can isolate and

---

[1] Fundación DILSE: `https://fundacioncnse-dilse.org/?buscar=cabeza`

separately analyze these dependent movements within the articulatory chain.

2.2. Automatic Tracking and Pose Estimation for Sign Language

Various methodologies have been employed to extract kinematic information from sign language videos. Early approaches utilized specialized hardware. For example, Wilbur and Zelaznik (1997) employed a 3D motion analyzer system, Koech (2007) used electromagnetic sensors and wearable gloves, and Abdullahi and Chamnongthai (2022) sourced data from a Leap Motion Controller. While effective, these techniques entail complex experimental setups and specialized datasets, limiting their accessibility for broad research applications. Alternative computational approaches include optical flow analysis (Borneman et al., 2018), though such methods provide only global measurements without discriminating between different articulators or movement directions.

More recently, pose estimation from 2D video has become feasible through machine learning advances. These methods can extract hand or body pose from ordinary video recordings, making existing video corpora accessible for kinematic analysis (Börstell, 2023). Such approaches estimate 3D hand pose from 2D videos, though they may involve complexities of annotation and dataset specificity (Ohkawa et al., 2023). Point-tracking methods offer an alternative by tracking arbitrary points within videos without requiring prior object information (Karaev et al., 2024). However, these yield point traces requiring further analysis to identify articulators.

Modern hand landmark detection models, such as the MediaPipe Hand Landmarker (Börstell, 2023), provide detailed tracking of hand structure by identifying multiple keypoints per hand, including the wrist and all finger joints. These models offer a balance between detail and practicality, working directly with 2D video while providing sufficiently rich information to support detailed analysis of hand movements and configurations.

2.3. Movement Decomposition Approaches

Several computational methods have been proposed to decompose sign language movements into meaningful components. Bigand et al. (2021) applied Principal Component Analysis (PCA) to upper-body motion capture data from French Sign Language, decomposing spontaneous signing into elementary movements called “principal movements” (PMs). Their analysis showed that the first eight common PMs explained 94.6% of the variance in movements across six signers producing spontaneous discourses. Importantly, they found that PM subspaces were highly similar across signers, suggesting that complex sign language movements can be described through a reduced set of elementary movements shared across individuals. This work demonstrates the value of dimensionality reduction for understanding motor control strategies in signing and has implications for automatic sign language processing.

Chassat et al. (2022) developed a generative model for sign language hand trajectories using the Frenet-Serret representation. Their approach separates temporal variations (speed) from spatial variations (geometry and shape) by representing trajectories through curvature, torsion, and arclength functions. Using functional principal component analysis on these geometric parameters, they modeled the variability in hand movements to generate realistic signing movements for avatar animation. This work highlights the importance of separating different sources of variation in movement data.

While these approaches successfully decompose sign language movements for analysis and generation, they differ from our work in their goals and methods. Bigand et al.’s holistic PCA decomposition identifies principal movement patterns across the entire upper body, while Chassat et al. focus on modeling trajectory variability for generative purposes. In contrast, our method specifically targets the isolation of dependent movements in an articulatory chain. We separate hand rotation from arm displacement and finger movement from hand movement to reveal linguistically-relevant structure and enable empirical observation of rhythmic and prosodic patterns. This linguistically-motivated decomposition, combined with accessible tools for ordinary 2D video and interactive visualization, addresses the need for methods that make articulatory structure directly observable for empirical linguistic research.

3. Methodology

The main goal of our methodology is to track and analyze sign language movements in a way that reveals their internal articulatory structure. Rather than treating the hand as a single moving point, or directly exploring the landmarks as output by the tracking algorithm, we decompose motion into separate gestural components: the hand’s path through space, the rotation of the wrist, and the movements of individual fingers. This decomposition reflects the coordinated but separable actions that signers control during articulation according to many sign language grammars (Stokoe, 1960; Muñoz Baell, 1999; Blanco, 2009; Brentari et al., 2018).

3.1. Hand Tracking with MediaPipe

The MediaPipe Hand Landmarker model (Vakunov et al., 2020) provides the foundation for hand tracking. This model analyzes each video frame and identifies 21 three-dimensional landmarks per hand, including the wrist and all finger joints. The landmarks include the metacarpophalangeal (MCP) joints where the fingers meet the palm. This spatial information captures both hand position and the detailed shape and orientation of the hand.

From these 21 landmarks, we extract a more succinct and representative model. First, we capture hand orientation by computing two vectors from the four knuckle (MCP, metacarpophalangeal joint) landmarks. A distal vector is computed from the wrist to the MCP center, representing the longitudinal axis of the hand. Subtracting the pinky MCP joint from the index MCP joint yields a lateral vector pointing in the direction of the thumb (when not flexed). These vectors form a coordinate system for points in the hand, independent of larger movements of the arm and centered on the palm for better interpretability. Our method outputs the original video with these vectors drawn for each hand, so that accuracy or potential errors can be observed.

Since the coordinate center moves with the wrist, examining the angle of the distal and lateral vectors provides the hand rotation, excluding any arm movement that might confound the method. This technique also serves to isolate finger movement. Landmarks of the fingers are represented as points in the space determined by the distal and lateral vectors, meaning that rotations of the hand while the finger is static do not induce a change in finger landmark positions. This allows computation of finger movements regardless of larger articulatory movements such as arm displacements or hand rotations.

3.2. Velocity computation

A key aspect of our approach is that it focuses not on landmark positions themselves, but rather on their movement along the signing interval. Therefore, we transform landmark and vector positions into measurements of movement.

Before computing velocities, automatic tracking errors that disrupt temporal continuity must be corrected. While we find the MediaPipe Hand Landmarker model remarkably accurate, sometimes a hand is lost in a few frames or left and right hands are mistaken for each other. Since our method has the full context available, it can reliably fix these errors. If a supposedly left hand appears suddenly when before (and after) there was a right hand, it is probably a mistaken identification. This does not happen as frequently when hands cross, where MediaPipe is quite reliable, but rather when one of the hands is not present or temporarily lost to the algorithm. A simple context-based switch fixes this, and any missed frames are then filled by interpolation.

To compute hand movement velocity, the series of hand positions and vector angles are differentiated using the Savitzky-Golay filter (Savitzky and Golay, 1964), which fits a local polynomial to position data and analytically differentiates it, providing a good balance of smoothness and precision. We then transform the velocities from x,y pixel-space vectors to polar space, yielding speed of movement and heading (direction).

For hand rotation, since we are primarily interested in the angle, we invert the order. We first transform the vectors to polar coordinates and then differentiate the angle function, again using Savitzky-Golay. To avoid the problem of discontinuity at $\pm\pi$ boundaries, which creates artificial discontinuities where 179° and -179° are separated by only 2° but appear to differ by 358°, angles are first unwrapped before differentiation by adding or subtracting 2π where necessary to maintain continuity. The resulting signal has a magnitude (speed of rotation) and a sign (clockwise or counter-clockwise rotation).

With movements decomposed into magnitude and direction, patterns and structuring of signs become more discernible. Speed primarily serves to separate segments and mark rhythm, while heading or rotation direction carries the lexical distinctions. These two channels of information combined enable reliable identification of when each sequential component of the sign starts or ends in the video.

3.3. Finger Movement Analysis

MediaPipe tracks all finger joints, providing four landmarks per finger and five fingers per hand for a total of 20 finger landmarks in addition to the wrist. These finger landmarks are projected onto the hand-relative coordinate system (the distal/lateral basis), yielding finger positions relative to the palm rather than to the camera frame. This provides very detailed information about finger configuration.

However, extracting meaningful velocity signals from finger landmarks presents unique challenges. Finger articulators are substantially smaller than hand-level movements. A finger displacement of a few millimeters constitutes a linguistically significant gesture such as handshape change, but such small movements are difficult to distinguish reliably from tracking noise, especially during fast movements where the hand can become quite blurred. Computing finger velocities using derivatives (as applied to hand position) results in noise overwhelming the signal.

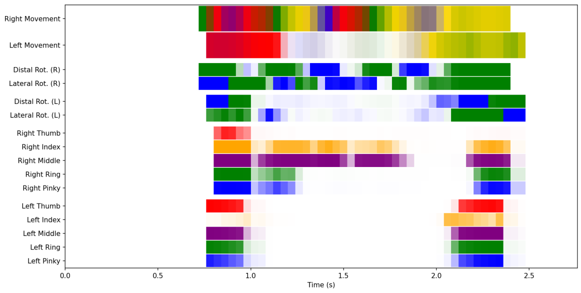


Figure 3: Bandogram analysis of the sign PINTOR (painter), showcasing many of the observations that can be visually inspected with our method. This is a bimanual sign with passive non-dominant hand and some internal movement of the selected (index and middle) dominant hand fingers.

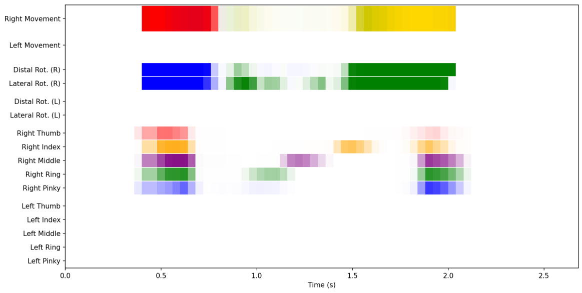


Figure 4: Bandogram analysis of the sign PREFERIR, which features a staggered closing of the fingers while the hand remains static. Interestingly, the hand starts descending for the relaxation while the index is still folding.

A more robust measure addresses this problem. Finger movement is quantified as dispersion within a sliding window rather than instantaneous velocity. For each finger landmark, the mean root square (MRS) of deviations from mean position within an 8-frame window is computed. This measures positional variation without requiring explicit derivatives, providing greater robustness to noise. MRS dispersion is computed for each finger landmark's distal and lateral coordinates, then averaged to obtain a single movement measure per landmark. Averaging across all finger landmarks then produces an overall movement measure for each finger, capturing whether it is moving or static without tracking individual trajectories in detail.

3.4. Plotting of results

To visualize the decomposed movement components, we employ the multi-band visualization introduced above (the bandogram) which presents all articulatory channels simultaneously while preserving their temporal alignment, as can be seen in Figure 3 (PINTOR[2], painter, same as Figure 1). The resulting plot resembles a spectrogram, with horizontal bands stacked vertically, each representing a different gestural component over time. While this visualization may appear complex at first glance, it allows researchers to see the full articulatory picture at once and identify the precise temporal relationships between different movement components. This is critical for understanding how sign languages coordinate multiple actuators to perform each sign or sentence.

The visualization uses a two-channel encoding strategy, building on our previous work (Sevilla et al., 2024): color intensity represents movement magnitude, while hue encodes different information depending on the articulatory channel.

For hand movement bands (the top two bands), hue represents movement heading in two-dimensional pixel space. The colormap assigns distinct colors to the four cardinal directions: gold for downward movement, green for rightward, red for upward, and blue for leftward. Circular hand motions produce smooth color transitions as the heading changes continuously, while abrupt changes of direction appear as sharp color discontinuities that clearly delimit sign segments. Such abrupt changes are common at sign boundaries. Movement speed controls color intensity: faster movements appear more saturated, while slower movements fade toward transparency. A sigmoid transformation maps speed to intensity, saturating at high velocities to emphasize regime changes rather than absolute magnitude.

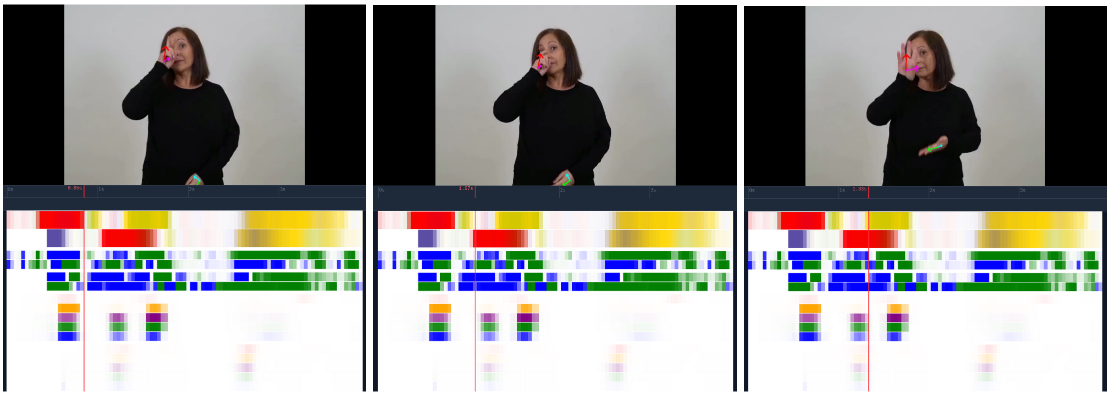
For rotation bands (four bands showing distal and lateral rotation for each hand), hue represents the sign of rotational velocity. Green indicates counterclockwise rotation while blue indicates clockwise rotation, with intensity again representing rotation speed. This encoding makes alternating rotational patterns, common in many signs, immediately visible as alternating green and blue segments.

For finger movement bands (ten bands, one per finger), color serves a different purpose: it distinguishes finger identity. Each finger receives a fixed color (thumb=red, index=orange, middle=purple, ring=green, pinky=blue) to enable reliable identification across the plot. Here, only intensity varies with movement magnitude, allowing observers to quickly identify which fingers are active during different phases of a sign, or even their relative sequentiation, as in Figure 4.

This visualization approach provides a comprehensive view of sign articulation that reveals both the sequential structure of signs (visible as temporal segments) and the coordination patterns between articulators (visible as temporal alignment of activity across bands).

---

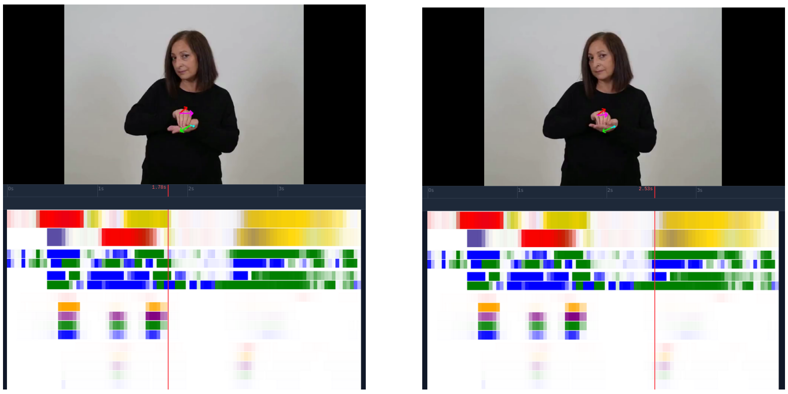
[2] `https://griffos.filol.ucm.es/signario/signo/11933`



(a) 0.85s: right hand near nose, index extended.

(b) 1.07s: right hand down somewhat, index flexed.

(c) 1.33s: right hand opens in preparation, left hand rises.



(d) 1.78s: right hand closed, touching top of left hand.

(e) 2.53s: end of the sign, after a significant "hold".

Figure 5: Five moments of the sign ANÉCDOTA selected by clicking on articulatory boundaries visible in the bandogram. The red line marks the selected time in both the timeline and the plot. The dominant hand velocity bands (top rows of the plot) are the primary indicator for identifying transitions between sign components, while the finger activity bands reveal the timing of the opening and closing of the hand.

3.5. Interactive Viewer

To facilitate detailed analysis of specific sign segments, we developed a basic interactive web-based viewer that synchronizes the original tracked video with the bandogram. The plot can be generated in borderless mode, removing axes and margins to create a clean visualization that aligns precisely with the video timeline. This viewer enables researchers to rapidly navigate to moments of interest and examine them through repeated playback.

The viewer interface presents the video and plot vertically stacked with a timeline bar between them. Researchers interact primarily through the plot: clicking seeks the video to that time point, while clicking and dragging sets loop markers. A red marker indicates current position or loop start, and a blue marker indicates loop end, if any. Both markers are visible across the plot and the timeline, and labeled with precise timestamps. When loop markers are set, the video automatically begins playback of the selected interval and repeats continuously, allowing close examination of specific articulatory events. The timeline displays second markers for reference and shows the loop interval duration. When no loop is active, the selected frame is shown.

This simple interaction model has proven remarkably effective for sign structure analysis. Researchers can quickly isolate individual signs or sub-sign units, compare the temporal alignment of different articulatory components by observing when bands show activity, and identify precise boundaries between sequential segments. The tight coupling between video and plot eliminates the friction of manually coordinating separate playback and visualization tools, enabling fluid exploration of the data. Figure 5 illustrates this workflow applied to the compound sign ANÉCDOTA[3] (anecdote), discussed further in subsection 4.3.

---

[3] `https://griffos.filol.ucm.es/signario/signo/11031`

4. Results and analysis

Our method and tools enable systematic and empirical study of sign language movement composition and structure. While a full corpus study remains future work, we present examples of the analyses enabled by this contribution.

First, some notes on interpreting the bandograms beyond the information given above. Since the intensity of color in the bands represents magnitude of movement, white stretches naturally represent holds, where the hands are kept static, or in single-handed signs, the absence of the non-dominant hand. One might expect that high intensity patches represent the most relevant segments of signing. However, non-linguistic segments of movement, such as preparations, relaxations and transitions, tend to have the highest speed, not because they are linguistically salient, but because perceptibility is not a concern while reaching the target position quickly is. Therefore, most plots show very prominent movement at the beginning (usually red, indicating upward motion) for the preparation and gold (downward motion) at the end for the relaxation. During these phases, internal movements are also less relevant, since the articulators are getting into position or may not be properly detected due to the high speed of arm motion.

Linguistic rhythm and patterning emerge in between, in the regular and medium speed segments between preparation and relaxation. We saturate intensity in our plots using a sigmoid function to better capture this region, but this distinction remains important to bear in mind.

4.1. Movement synchronization

The bandogram seen in Figure 1 and Figure 3 provides a good example of the types of information our methodology makes apparent. In the top two bands, we can see the initial preparation and closing relaxation. Both hands are involved, but in the actual lexical segment, they act differently. The right hand performs a circular motion, evidenced by the continuous change of color, while the left hand remains static. This is a bimanual sign with a passive non-dominant hand (H2). This is also evidenced in the rotation and finger bands, which repeat this pattern. In the right hand fingers, furthermore, we can see that this is a sign involving the index and middle fingers, while the rest remain unselected. A short hold can be seen before the relaxation, though it is easy to miss in the video itself.

4.2. Bimanual Symmetry

By plotting the right and left hands consecutively, the synchronization of hands during signing can also be observed. In Figure 6 and Figure 7, two signs with circular movements can be seen[4]. However, they have different hand synchronization. While in Figure 6 both hands move at the same time in the same direction (ipsilateral, meaning the same "body" direction, not pixel space direction), in Figure 7 they are inverted. This pattern, very common in sign languages, is visible through the time alignment of the plots and the use of color to encode direction information.

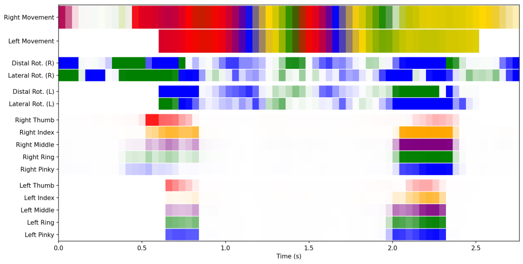


Figure 6: Bandogram analysis of the sign MUEBLES (furniture), a symmetric bimanual circular movement.

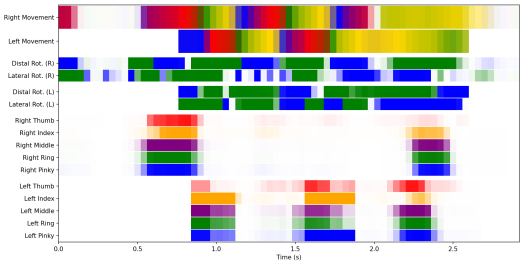


Figure 7: Bandogram analysis of the sign CONFUSION (confusion), an inverted bimanual circular movement.

4.3. Compound sign segmentation

The interactive viewer is particularly useful for segmenting compound signs into their component parts. Figure 5 shows five frames of the compound sign ANÉCDOTA (anecdote), selected by clicking on transitions identified in the bandogram. The sign combines two components: a finger closing internal movement near the nose (related to "memory") and a bimanual closing gesture in neutral space (related to "record/store"). Frames (a) and (b) capture the onset and coda of the first component, with downward movement of the dominant hand seen in yellow in the bandogram. Unfortunately, the index flexion was not captured by MediaPipe, unlike the opening of the other fingers between (b) and (c),

---

[4] https://griffos.filol.ucm.es/signario/signo/12328
https://griffos.filol.ucm.es/signario/signo/12218

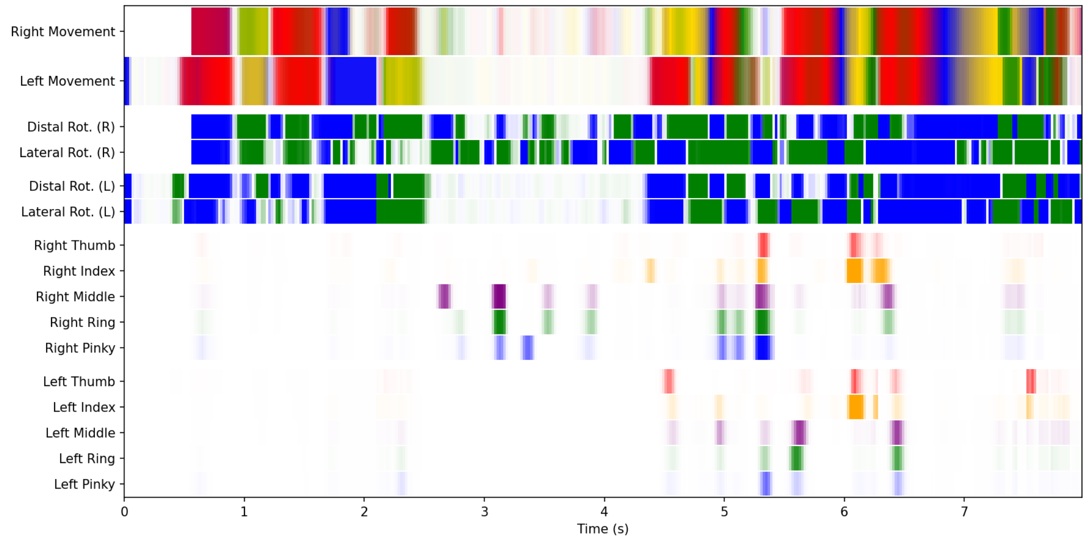


Figure 8: Bandogram of the first 8 seconds of the Public DGS Corpus welcome video, showcasing our method in use with longer, continuous signing.

visible in the bottom purple, green and blue bands.

Frame (c), with the right hand already open in preparation for the second sign component, also shows that the left hand has begun rising (red motion in the second band) but has not yet reached its target position. In frame (d) the downward plus closing gesture of the "recording" component has been finished, with the hands touching in this passive non-dominant hand configuration. It is interesting to see that the left hand preparation (moving up from rest position to the neutral signing space) finishes early, but starts as soon as the external movement of the first component has finished (b). The prototypical "hold" segment at the end of signs can be clearly seen between (d) and (e), but not between (b) and (c), strongly indicating that this combination of two already existing meaning signifiers is its own independent lexical item.

4.4. Continuous signing

Our method can be applied to continuous signing as well. In Figure 8, an analysis of the first 8 seconds of the DGS (Hanke et al., 2020) welcome video at the public DGS corpus page[5] can be seen. Using our viewer, the different segments can be analyzed, but even without the video, structure can be clearly read. The fingerspelling section from seconds 2.5 to 4.5 can be seen as a period of inactivity in the left hand but internal movements in the right one, including rotations and fast finger movements typical of fingerspelling. Other typical patterns along the video can be inspected, for example the small stops or holds between signs at the beginning of the video, and the circular movements in the signs after the fingerspelling, clearly visible in the continuous "rainbow" pattern. At second 5.25, a sign with some finger wiggling is performed, clearly visible in the bandogram.

4.5. Limitations

The analyses above also reveal some of our method's limitations. On one hand, it inherits the constraints of the underlying MediaPipe tracker: the hand landmark model was designed for near-frontal views with good lighting, so performance degrades under motion blur, partial occlusion, or unusual hand orientations. While our pipeline interpolates over short tracking failures, more extended occlusions or rapid movements can introduce gaps or artifacts in the plots that must be interpreted with care.

Regarding finger articulation, the finger bands indicate whether individual fingers are moving and at what speed, but they do not capture handshape transitions in fine-grained detail. The tracker provides landmark positions rather than joint angles, so subtle handshape changes may not be reliably reflected, particularly when the hand is moving fast or when fingers overlap from the camera's perspective. With these limitations in mind, we turn to the broader conclusions about the method's potential and directions for future work.

5. Conclusions

The examples presented demonstrate that our method successfully isolates dependent movements within the sign language articulatory chain, separating hand rotation from arm displacement

---

[5] `https://www.sign-lang.uni-hamburg.de/dgs-korpus/public-dgs-corpus.html`

and finger movement from hand movement. The bandogram visualization reveals temporal structure and synchronization patterns that are difficult to observe in raw video alone, including bimanual coordination (symmetric vs. inverted movements), the distinction between linguistic and non-linguistic segments, and the temporal alignment of different articulatory components. The method scales from isolated signs to continuous signing, working with ordinary 2D video through accessible tools like MediaPipe.

These capabilities enable empirical observation of rhythmic and prosodic patterns in sign languages that may address theoretical questions about sign language phonology. By separating movement into distinct channels (displacement, rotation, finger articulation) with both magnitude and direction information, linguistically-relevant organizational patterns become visually accessible. The interactive viewer further supports this empirical work by enabling researchers to efficiently explore movement patterns and their relationship to video, facilitating fluid examination of specific articulatory events.

While the examples shown here demonstrate the method's potential, systematic corpus studies remain future work. Such studies could leverage our tools to investigate questions about syllable structure, prosodic organization, and the coordination of simultaneous articulatory gestures across different sign languages. The combination of automated tracking, movement decomposition, and interactive visualization provides a foundation for such empirical investigations into sign language structure.

6. Acknowledgements

This publication is part of the R&D&I project HumanAI-UI, Grant PID2023-1485 77OB-C22 (Human-Centered AI: User-Driven Adaptative Interfaces-HumanAI-UI) funded by MICIU/AEI/10.13039/ 501100011033 and by FEDER/UE.

We want to acknowledge our sign language informants at "Idiomas Complutense", whose invaluable participation made this work possible.

7. Bibliographical References

Sunusi Bala Abdullahi and Kosin Chamnongthai. 2022. American Sign Language Words Recognition Using Spatio-Temporal Prosodic and Angle Features: A Sequential Learning Approach. *IEEE Access*, 10:15911–15923. Conference Name: IEEE Access.

Mark Aronoff, Irit Meir, and Wendy Sandler. 2005. The Paradox of Sign Language Morphology. *Language*, 81(2):301–344.

Félix Bigand, Elise Prigent, Bastien Berret, and Annelies Braffort. 2021. Decomposing spontaneous sign language into elementary movements: A principal component analysis-based approach. *PLOS ONE*, 16(10).

Ángel Luis Herrero Blanco. 2009. *Gramática didáctica de la lengua de signos española (LSE)*. Sm.

Joshua D. Borneman, Evie Malaia, and Ronnie B. Wilbur. 2018. Motion characterization using optical flow and fractal complexity. *Journal of Electronic Imaging*, 27(5).

Diane Brentari, Jordan Fenlon, and Kearsy Cormier. 2018. Sign Language Phonology. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press.

Carl Börstell. 2023. Extracting Sign Language Articulation from Videos with MediaPipe. In *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 169–178, Tórshavn, Faroe Islands. University of Tartu Library.

Perrine Chassat, Juhyun Park, and Nicolas Brunel. 2022. Analysis of variability in sign language hand trajectories: Development of generative model. In *Proceedings of the 8th International Conference on Movement and Computing*, pages 1–8, Chicago IL USA. ACM.

Jordan Fenlon and Diane Brentari. 2021. Prosody: Theoretical and experimental perspectives. In Josep Quer, Roland Pfau, and Annika Herrmann, editors, *The Routledge Handbook of Theoretical and Experimental Sign Language Research*.

Thomas Hanke, Marc Schulder, Reiner Konrad, and Elena Jahn. 2020. Extending the public dgs corpus in size and depth. In *Proceedings of the LREC2020 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*, pages 75–82, Marseille, France. European Language Resources Association (ELRA).

Nikita Karaev, Ignacio Rocco, Benjamin Graham, Natalia Neverova, Andrea Vedaldi, and Christian Rupprecht. 2024. CoTracker: It is better to track together. In *Computer Vision – ECCV 2024*.

Chemuttaai Koech. 2007. A kinematic analysis of sign language. Master's thesis, New Jersey Institute of Technology.

Scott K. Liddell and Robert E. Johnson. 1989. American Sign Language: The Phonological Base. *Sign Language Studies*, 64(1):195–277.

Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Yong, Juhyun Lee, et al. 2019. Mediapipe: A framework for perceiving and processing reality. In *Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR)*, volume 2019.

Irma Muñoz Baell. 1999. *¿Cómo se articula la LSE?* CNSE, Spain.

Takehiko Ohkawa, Ryosuke Furuta, and Yoichi Sato. 2023. Efficient Annotation and Learning for 3D Hand Pose Estimation: A Survey. *International Journal of Computer Vision*.

Jefferson Rodríguez, Juan Chacón, Edgar Rangel, Luis Guayacán, Claudia Hernández, Luisa Hernández, and Fabio Martínez. 2020. Understanding motion in sign language: A new structured translation dataset.

Abraham Savitzky and Marcel JE Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639.

Antonio F. G. Sevilla, José María Lahoz-Bengoechea, and Alberto Diaz. 2024. Automated Extraction of Prosodic Structure from Unannotated Sign Language Video. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 1808–1816, Torino, Italia. ELRA and ICCL.

Kelly M Stack. 1988. Tiers and Syllable Structure in American Sign Language: Evidence from Phonotactics. Master's thesis, University of California, Los Angeles.

William C. Stokoe. 1960. *Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf*, volume 8 of *Studies in Linguistics: Occasional Papers*. University of Buffalo, Buffalo, NY.

Martha E Tyrone, Hosung Nam, Elliot Saltzman, Gaurav Mathur, and Louis Goldstein. 2010. Prosody and Movement in American Sign Language: A Task-Dynamics Approach. In *Proceedings of Speech Prosody 2010*, pages 1–4, Chicago, IL. ISCA.

Andrey Vakunov, Chuo-Ling Chang, Fan Zhang, George Sung, Matthias Grundmann, and Valentin Bazarevsky. 2020. Mediapipe hands: On-device real-time hand tracking. In *Workshop on Computer Vision for Augmented and Virtual Reality at CVPR 2020*.

Ronnie B Wilbur and George D Allen. 1991. Perceptual evidence against internal structure in American Sign Language syllables. *Language and Speech*, 34(1):27–46.

Ronnie B. Wilbur and Aleix M. Martinez. 2002. Physical correlates of prosodic structure in American Sign Language. In *Chicago Linguistic Society*, volume 38, pages 693–704.

Ronnie B Wilbur and Howard N Zelaznik. 1997. Kinematic correlates of stress and position in asl. In *Annual Meeting of the Linguistic Society of America (LSA), Chicago, Illinois*.

Kayo Yin, Amit Moryossef, Julie Hochgesang, Yoav Goldberg, and Malihe Alikhani. 2021. Including Signed Languages in Natural Language Processing. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7347–7360, Online. Association for Computational Linguistics.