Quantifying sign language movement kinematics from video  
Aline Normoyle, Samuel J. Rothstein, and Norman I. Badler

### Abstract
We describe our system for extracting kinematic metrics for sign language gestures from a video. The goals of these analyses are to provide insights into how factors related to physical movements, such as principles of least effort and ease of articulation, might affect how sign languages evolve or differ across individuals.

### Motivation
Evidence suggests that psycho-physical factors influence sign languages.

- Signers might use a one-handed version of a sign in casual conversations, and a two-handed version in formal settings. (Zimmer 2000)
- Signers might also reduce the number of repetitions in a sign, or cut short the articulation of a sign, or replace translational symmetry with reflective symmetry. (Napoli et al. 2014)
- Sign may evolve to move away from peripheral locations and closer towards the face (Frischberg 1975)

Such shortcuts are analogous to co-articulations, contractions, and acronyms used in spoken English. We can use kinematic metrics to identify how sign languages change based on context, differ across individuals, or change to make expressions easier and faster to convey.

Most sign language datasets are based on video. Usually, fluent speakers need to transcribe and annotate these data sets manually. However, recent advances in computer vision give us the opportunity to automate parts of this process. Furthermore, we can measure quantities that would have been prohibitively difficult to measure before.

### ASL movement: Saying chat
A primary goal of our pipeline is to quantify the articulatory effort needed to make different signs. In the examples below, we show several variants for signing the word chat, as in ‘Many people chatted at the cafe’. Some variants require the speaker to move the hands much more than others.

Chat in ASL, signed with the hands near the face using a quick repetitive movement. Images from ASL-LEX (Sehry et al. 2021)

Chat in ASL, signed with the hands below the face using a quick repetitive movement. This gesture indicates a short chat. Images from signingsavvy.com

Chat in ASL, signed with the hands alternating up and down with mirrored symmetry. This gesture indicates a long chat. Images from signingsavvy.com

### Metrics
Our current metrics explore measures of articulatory effort using different complexity of model. Some metrics, such as distance to a rest posture, only rely on the points of the skeleton. Other metrics, such as kinetic energy and torques, require full mass and inertia estimates for the entire skeleton. We compute velocities and accelerations using 5-point central differencing and torques using inverse dynamics. These metrics are lowest when both hands are in the lap and highest when both arms are accelerating quickly.

**Distance to rest posture:** Measures the difference (in meters) between the current pose and a manually chosen rest posture.

**Hand speed:** Hand speed (m/s) is the sum of the speeds of both wrists.

**Hand potential energy:** Hand potential energy (Joules) quantifies the effect of gravity on the hands.

### Ongoing work
Our current work is exploring additional metrics for measuring symmetry, size, and repetitions of movement in the trajectories of the hands. We are also designing experiments to validate such metrics.

Several challenges need to be addressed to complete this system. The body model currently used to estimate mass and inertia does not take into account different body sizes or gender. Furthermore, the distribution of mass across the body requires the weight and height of the individual, which is difficult to discern from a video.

### References

### Acknowledgements
Many thanks go to our collaborators: Matthew Dye (RIT), Naomi Caselli (BU), Corrine Senghas, Matthew Dye (RIT), and Andreas Savakis (RIT) for their work to extract poses from the video; and Matt Huenerfauth for providing the RGB-D data set. Thanks also to our reviewer for their helpful feedback and comments as well as to Effie Lu for her early work on this project. This work is supported by NSF Grant BCS 1748976 and REUs.