



Leibniz-Zentrum
Allgemeine Sprachwissenschaft



DFG Deutsche
Forschungsgemeinschaft



From Visual Salience to Bodily Signal

A Movement Science Approach to Gesture, Discourse and Prominence

Šárka Kadavá

ProDiGe workshop, Cologne, December 2025

Today's agenda

- i. Theoretical foundation
- ii. Dimensions of movement
- iii. (Rather quick) intro to motion tracking
- iv. Hands-on tutorial

What is gesture?

„actions that have the features of manifest deliberate expressiveness“ (Kendon 2004)

„utterance dedicated visible bodily action“ (Kendon 2016)

„motion that embodies a meaning relatable to the accompanying speech“
(McNeill, Enc)

„communicative tools that can enhance or alter the recipient's understanding of spoken utterances (de Ruiter 2003)

„a movement of the hands, arms, or head, etc. to express an idea or feeling“
(CambDic)

“communicative movements of the hands and arms, which, similar to language are used to express the thoughts, feelings, and intentions of a speaker” (Müller 1998)

What is gesture?

„**actions** that have the features of manifest deliberate expressiveness“ (Kendon 2004)

„utterance dedicated visible **bodily action**“ (Kendon 2016)

„motion that embodies a meaning relatable to the accompanying speech“
(McNeill, Enc)

„communicative tools that can enhance or alter the recipient's understanding of spoken utterances (de Ruiter 2003)

„a **movement** of the hands, arms, or head, etc. to express an idea or feeling“
(CambDic)

“communicative **movements** of the hands and arms, which, similar to language are used to express the thoughts, feelings, and intentions of a speaker” (Müller 1998)

What is gesture?

„actions that have the features of manifest **deliberate expressiveness**“ (Kendon 2004)

„utterance dedicated visible bodily action“ (Kendon 2016)

„motion that **embodies a meaning** relatable to the accompanying speech“ (McNeill, Enc)

„**communicative tools** that can enhance or alter the recipient's understanding of spoken utterances (de Ruiter 2003)

„a movement of the hands, arms, or head, etc. **to express an idea or feeling**“ (CambDic)

“communicative movements of the hands and arms, which, similar to language are used to **express the thoughts, feelings, and intentions** of a speaker” (Müller 1998)

What is gesture?

„actions that have the features of manifest deliberate expressiveness“ (Kendon 2004)

„**utterance dedicated** visible bodily action“ (Kendon 2016)

„motion that embodies a meaning **relatable to the accompanying speech**“ (McNeill, Enc)

„communicative tools that can **enhance or alter the recipient's understanding of spoken utterances** (de Ruiter 2003)

„a movement of the hands, arms, or head, etc. to express an idea or feeling“ (CambDic)

“communicative movements of the hands and arms, which, **similar to language** are used to express the thoughts, feelings, and intentions of a speaker” (Müller 1998)

Definition shapes annotation



Gesture as meaningful

← what is the meaning?

iconic, metaphoric, deictic, pragmatic, ...

Gesture relates to speech

← how?

synchrony with prosodic peaks, lexical affiliates, ...

Gesture is (somewhat) movement

← articulatory phases

preparation, **stroke (and apex)**, hold, retraction, ...

Problem I: theory

Meaning emerges from movement...

Iconicity is expressed through shape, velocity, trajectory, direction, etc. (Spruijt et al., 2025)

Beats derive their function from rhythmic, dynamic movement qualities

Emphasis emerges from acceleration, amplitude (and effort)

Trajectory and endpoint position of pointing gesture (i.e., the precision) matter for reference

Movement is continuous

Gesture phases simplify; hide variations, and cannot represent the dynamic nature of gesture; also stroke is not the only meaning-bearing part (Obermeier et al., 2011)

Gesture is not abstract concept, it is a physical mass

Moving with arm is differently costly than moving with wrist only

Timing is only one layer of multimodal coordination

Also: spatial alignment, amplitude, effort distribution, rhythm; Prominence can be distributed across modalities in complex ways; it is not only important when but also how

Problem II: practice



Resources (esp. time)



Problem II: practice




Language & Communication

Volume 105, November 2025, Pages 22-36



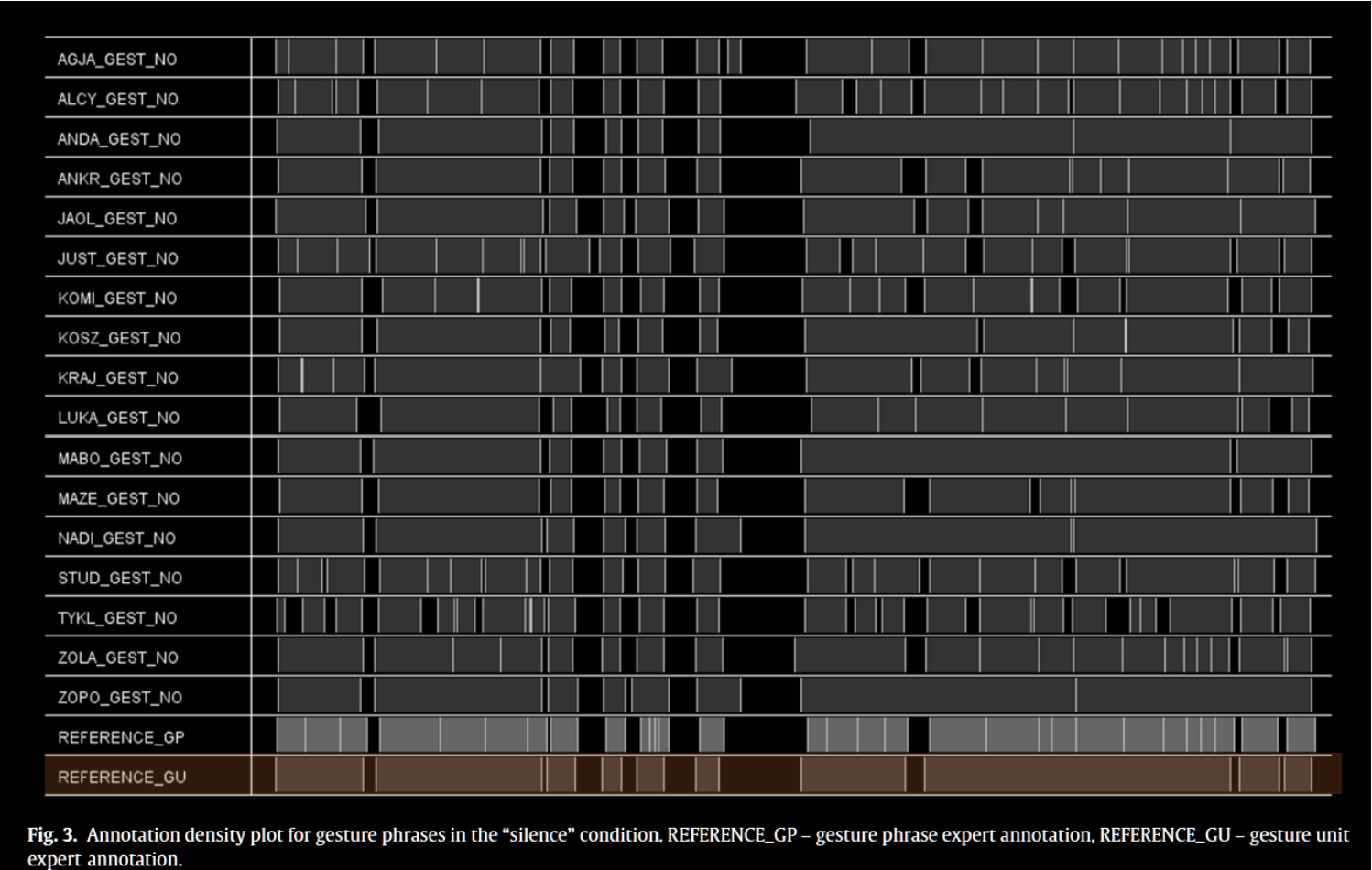
„Reliability“

Boundaries of gestures: Naive segmentation of the stream of human hand movements

Ewa Jaromołowicz-Nowikow  , Maciej Karpiński 

*„Researchers frequently use video annotation software to facilitate this process, leveraging features like multiple playback speeds, slow-motion, and reverse playback. **However, many spatial and visual aspects of annotating videos on a computer screen differ markedly from real-life communicative contexts. These differences include the size and visibility of the observed individual on the screen, the absolute distance from the person, the absence of a fully three-dimensional perspective, and the potential for either more restricted or more expansive access to the communicative situation. Such disparities may influence the annotator’s focus, leading to prioritizing different phenomena or aspects of behavior during the annotation process.**“*

What is actually gesture in a stream of behavior?



What is gestural prominence?

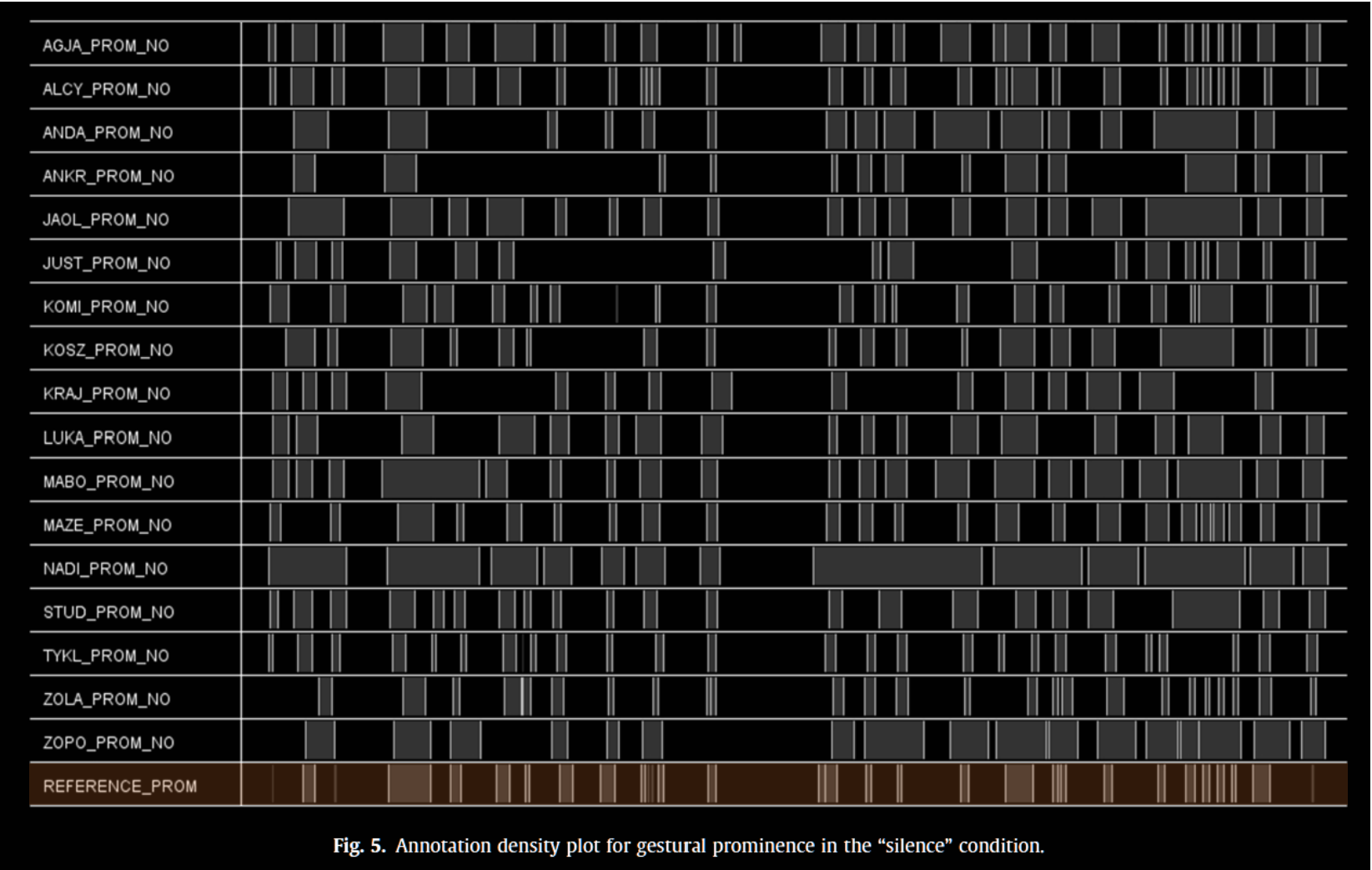


Fig. 5. Annotation density plot for gestural prominence in the “silence” condition.

Working with manual annotation...



Flexible

Powerful, esp. for quali-analysis

Somewhat consensus around annotation schemes

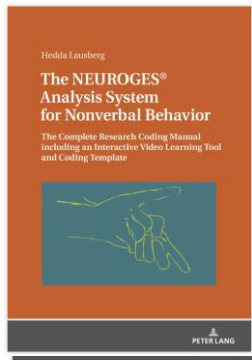


Laborious

Can be quite slow and buggy

Different types of annotation schemes

Perception-dependent



Linguistic Annotation System for Gestures (LASG)

September 2013

DOI: [10.1515/9783110261318.1098](https://doi.org/10.1515/9783110261318.1098)

In book: Body – Language – Communication. An International Handbook on Multimodality in Human Interaction. (Handbooks of Linguistics and Communication Science 38.1.) ·

Publisher: De Gruyter Mouton · Editors: Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill, Sedinha Teßendorf

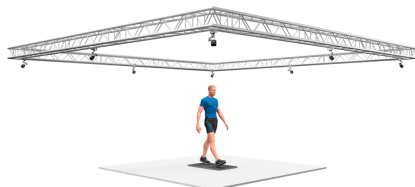
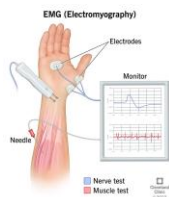
👤 Jana Bressemer · 👤 Silva Ladewig · 👤 Cornelia Müller



For prominence research, which matters more – the perceptual stroke or the biomechanical peak in movement (aligned with speech)?

The dawn of motion tracking

2000

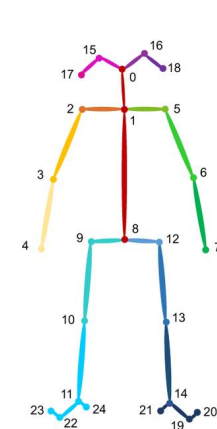


2014



Language as a multimodal phenomenon:
implications for language learning,
processing and evolution

Gabriella Vigliocco, Pamela Perniss and David Vinson



DeepLabCut:
a software package for
animal pose estimation

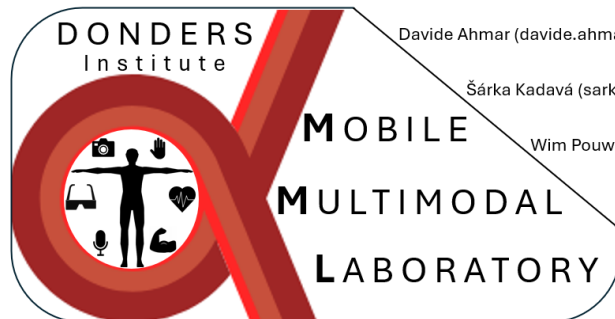


MediaPipe



What is so great about it?

- 1) it allows to **ask** new research questions
- 2) it **democratizes** science
- 3) it **frees** the hands
- 4) it **crosses** the departments & fields



Davide Ahmar (davide.ahmar@ru.nl)

Šárka Kadavá (sarka.kadava@donders.ru.nl)

Wim Pouw (wim.pouw@donders.ru.nl)



New possibilities for gesture!



Automatic detection

Quantification

Precise temporal alignment

Capturing variation

Dimensions of movement

I. Kinematics describes motion of points and bodies

Where?

Position

How far?

Displacement

How fast?

Velocity/speed

How suddenly?

Acceleration

How smoothly?

Jerk

How big?

Volume

How configured?

E.g., hand shape

How regular?

Rhythmicity

How coordinated?

Temporal/spatial relationships

Dimensions of movement

II. Kinetics/dynamics

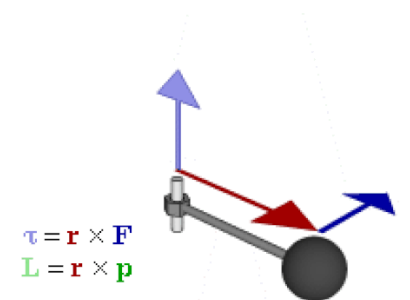
describes the cause of motion

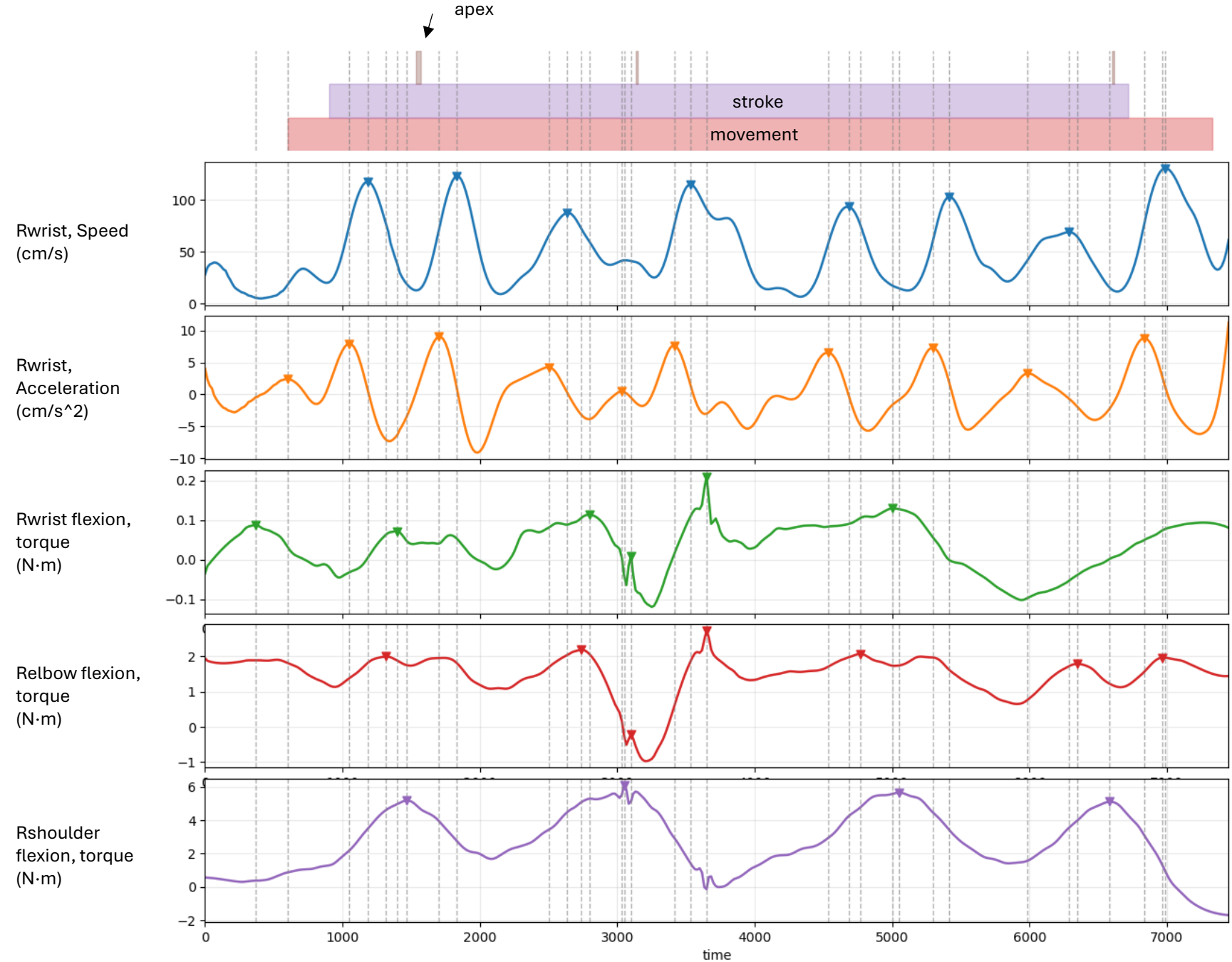
How effortful?

Forces, joint torques

How coordinated?

Synergies







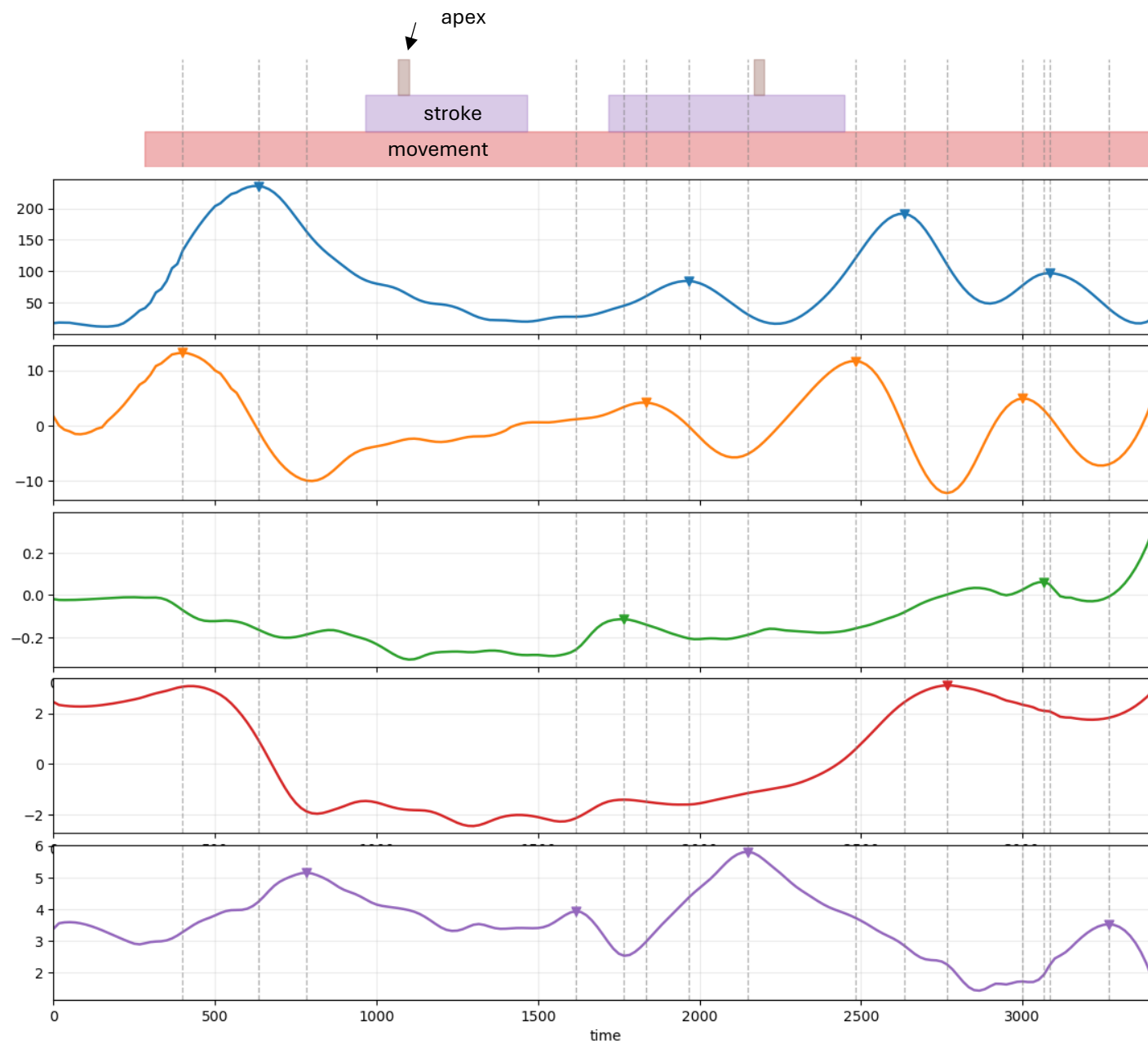
Rwrist, Speed
(cm/s)

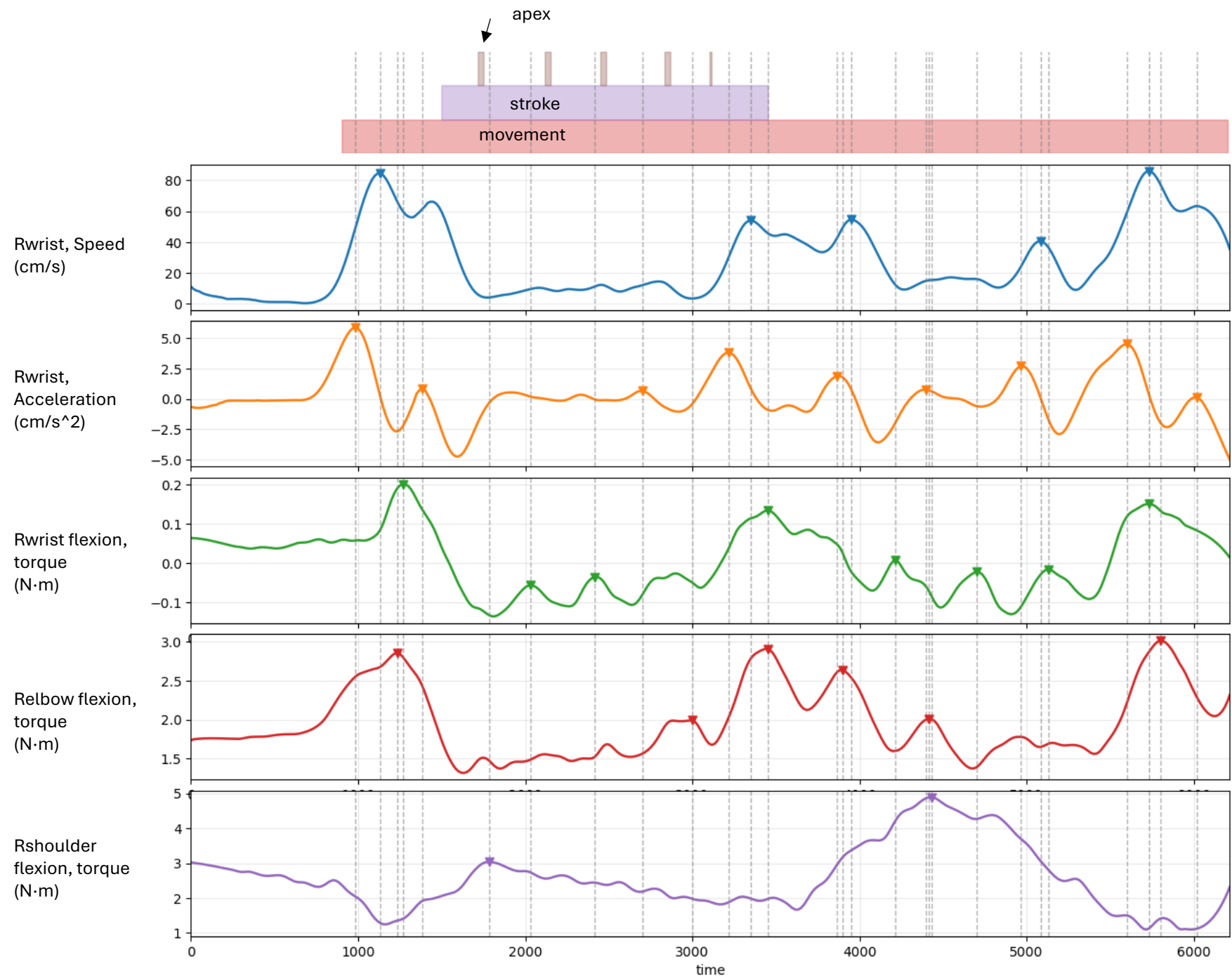
Rwrist, Acceleration
(cm/s²)

Rwrist flexion, torque
(N·m)

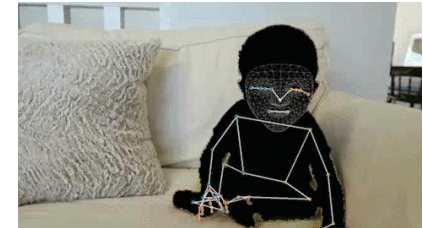
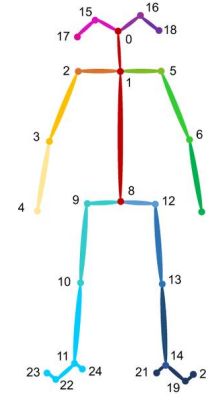
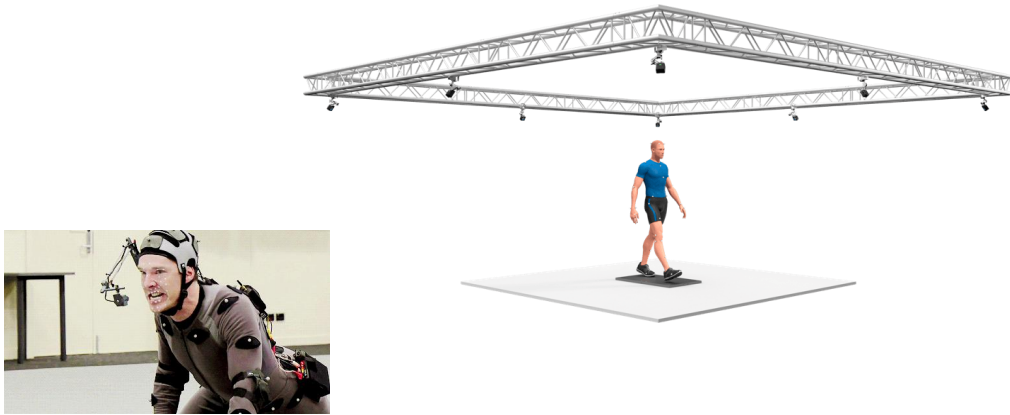
Relbow flexion, torque
(N·m)

Rshoulder flexion, torque
(N·m)





(Rather quick) intro to motion tracking



Very high precision (ca. 0.5 mm)

Lab-based

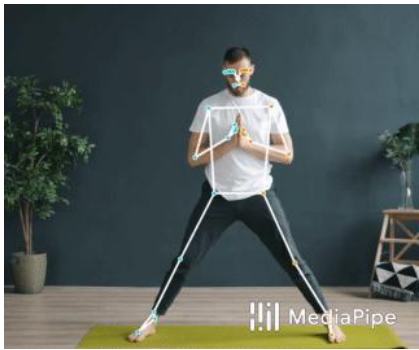
Proprietary (→ expensive)

Lower precision (1-2.5 cm)

No markers needed

Open-source

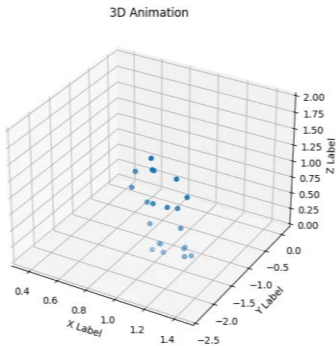
Most common marker-less tracking



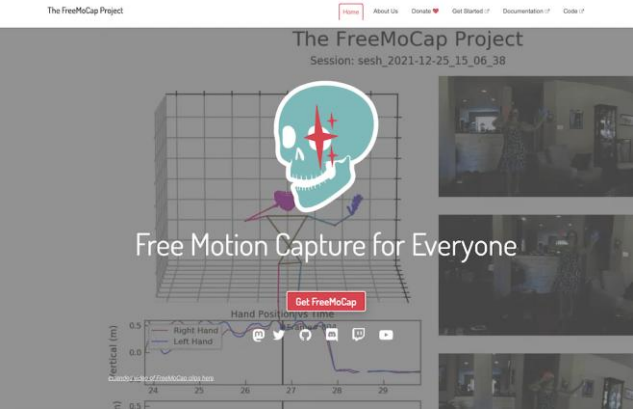
MediaPipe



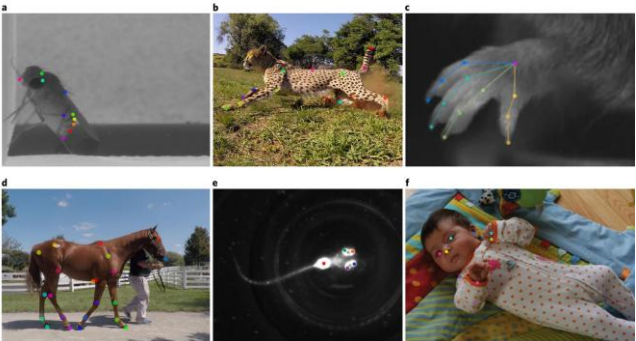
OpenPose



Pose2sim

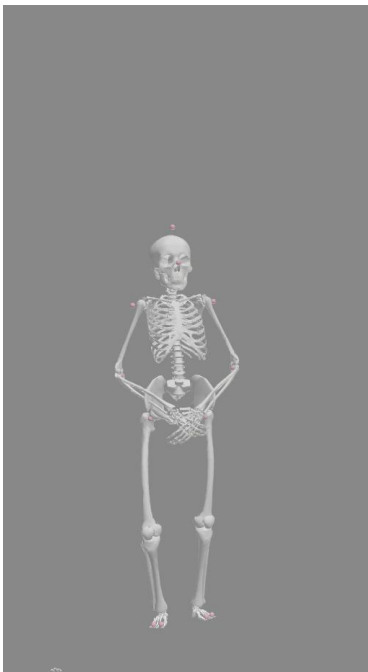


FreeMoCap

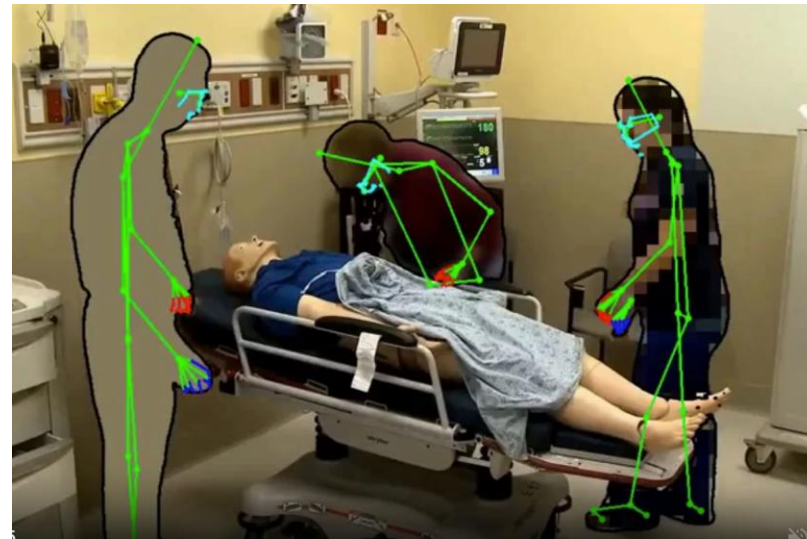
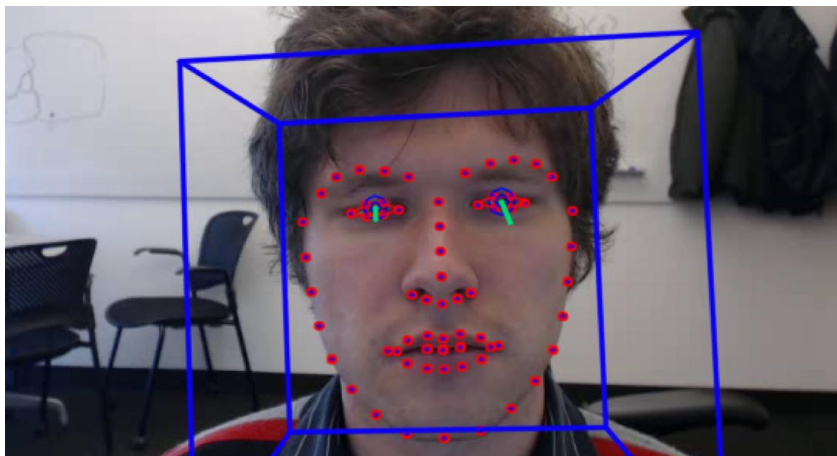


DeepLabCut

Add-ons



OpenFace



MaskAnyone

Considerations

Do I already have data, or am I just about to collect them?

What is my goal and how much is movement crucial for my RQ?

How much precision do I need?

How easy should I make my life?

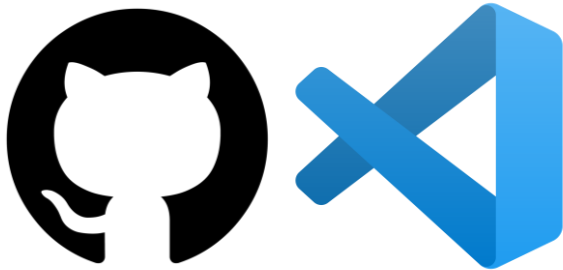


Let's get to work





<https://tinyurl.com/prodige-script-1>
<https://tinyurl.com/prodige-script2>

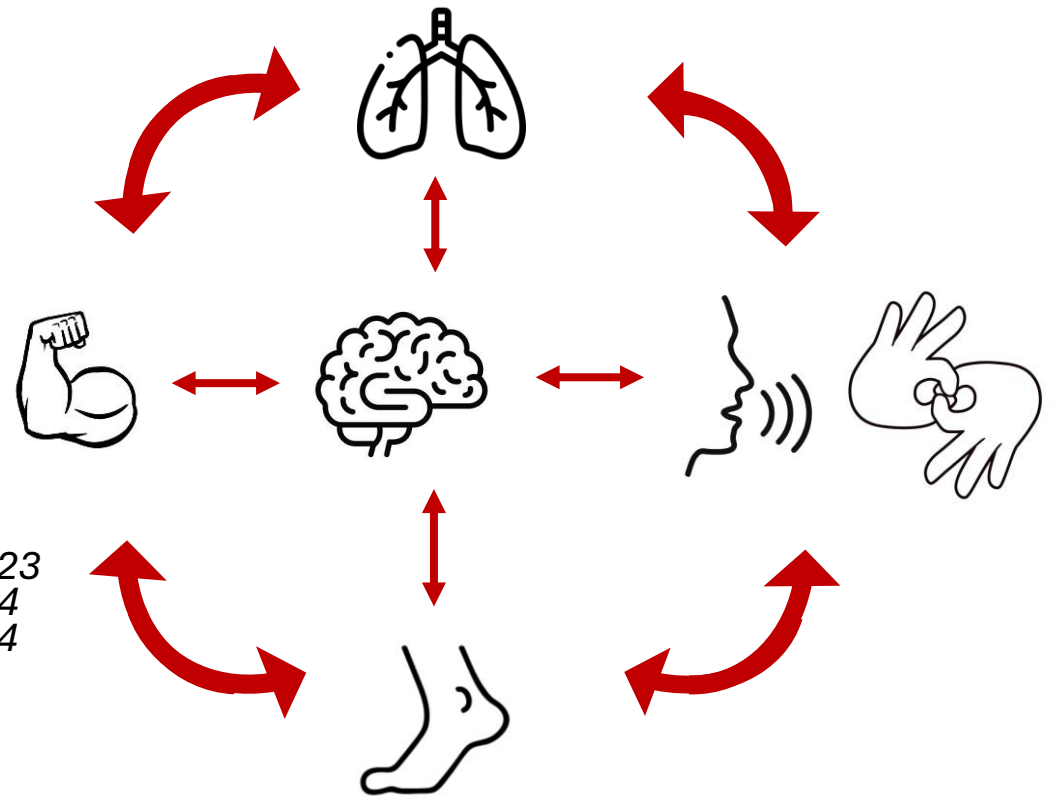


https://github.com/sarkadava/MotionTrackingPipeline_ProDiGe2025

Final remarks

gesture does not
operate on an
abstract level, it's
**profoundly
physical and
entangled with
whole-body
physiology**

Fuchs & Rochet-Capellan 2014



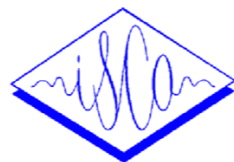
*Pouw et al. 2020
Serré et al. 2022
Pouw & Fuchs 2023
Weston et al. 2024
Offrede et al. 2024*

*Sundberg et al. 1991
Yardley et al. 1999
Lagier et al. 2010*

Final remarks

other parts of body are also involved in meaning-making





Are torso movements during speech timed with intonational phrases?

Stefanie Shattuck-Hufnagel¹, Pei Lin Ren¹ and Elizabeth Tauscher²

¹ Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge MA USA ² Wellesley College, Cambridge MA USA
sshuf@mit.edu, pelire@mit.edu, etausch@wellesley.edu



Psychiatry >

Interpersonal and Biological Processes

Volume 27, 1964 - [Issue 4](#)

[Submit an article](#)

[Journal homepage](#)

Enter keywords, authors, C

547

Views

368

CrossRef
citations to date

6

Altmetric

Original Articles

The Significance of Posture in Communication Systems†

Albert E. Scheflen

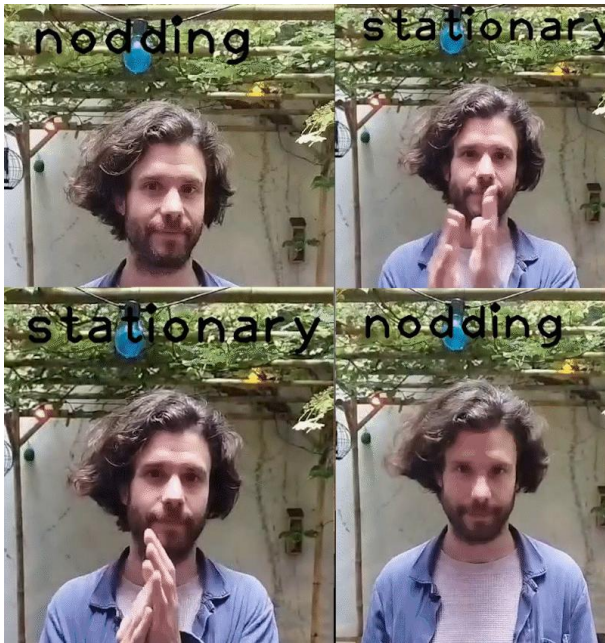
Pages 316-331 | Published online: 07 Nov 2016

“ Cite this article

“ <https://doi.org/10.1080/00332747.1964.11023403>

Final remarks

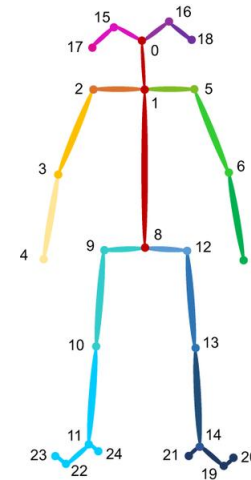
motion capture does not have to be the **endpoint**



What if all the time spent in [fill in a name of an annotation software that you really hate] in the past can contribute to less time spent in it in the future?



movement_in_trial [2]	nomovement	movement	nomovem
upper_body [1]	nomovement		
arms [1]	nomovement		
lower_body [1]	nomovement		
head_mov [2]	nomovement	movement	nomovem



Co-Speech Gesture Detection Through Multi-Phase Sequence Labeling

Esam Ghaleb, Ilya Burenko, Marlou Rasenberg, Wim Pouw, Peter Uhrig, Judith Holler, Ivan Toni, Aslı Özyürek, Raquel Fernández; Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2024, pp. 4007-4015

Behavior Research Methods (2020) 52:1783–1794
<https://doi.org/10.3758/s13428-020-01350-2>

Speeding up the detection of non-iconic and iconic gestures (SPUDNIG): A toolkit for the automatic detection of hand movements and gestures in video data

Jordy Ripperda¹ • Linda Drijvers^{1,2} • Judith Holler^{1,2}

Published online: 23 January 2020
© The Author(s) 2020



Using Envision's Automatic Hand Gesture Detection PyPi package (envisionhgdetector version 2.0.0+)



Wim Pouw (wim.pouw@donders.ru.nl), Bosco Yung, Sharjeel Shaikh, James Trujillo, Antonio Rueda-Toicen, Gerard de Melo, Babajide Owoyele (Babajide.Owoyele@hpi.de)



COMPARING MANUAL VS. SEMI-AUTOMATED METHODS FOR THE CODING OF CO-SPEECH GESTURES

Walter Dych¹, Karee Garvin^{1,2}, Kathryn Franich^{1,2}

¹University of Delaware, ²Harvard University
wdych@udel.edu, kareegarvin@fas.harvard.edu, kfranich@fas.harvard.edu

Learning Co-Speech Gesture Representations in Dialogue through Contrastive Learning: An Intrinsic Evaluation

Esam Ghaleb
University of Amsterdam
Amsterdam, The Netherlands
e.ghaleb@uva.nl

Bulat Khaertdinov
Maastricht University
The Netherlands

Wim Pouw
Radboud University
The Netherlands

Marlou Rasenberg
Meertens Institute
The Netherlands

Judith Holler & Aslı Özyürek
Radboud University & MPI for Psycholinguistics
The Netherlands

Raquel Fernández
University of Amsterdam
The Netherlands
raquel.fernandez@uva.nl

Final remarks

motion capture is a great accelerator of our field, but it does not do miracles

- ... there is some accuracy trade-off
- ... it requires some thinking about the ,infrastructure‘
- ... it requires lots of processing steps (and thus, too, time)
- ... it requires constant sanity checks (!!!)
- ... sometimes coding is as annoying as ELAN
- ... data are just data without theory

What's next?



Modules covering MediaPipe, OpenPose, Pose2sim any many more



**... And maybe see you in
Tilburg, June 2026?**



Leibniz-Zentrum
Allgemeine Sprachwissenschaft



DFG Deutsche
Forschungsgemeinschaft



Thank you!

kadava@leibniz-zas.de

On the FLExibility and Stability of gesture-speech coordination (FLESH)



Aleksandra Ćwiek



Susanne Fuchs



Šárka Kadavá



Wim Pouw



**Melissa
Ebert**



**Justin
Snelders**



**Jet
Lambers**



**Hamza
Nalbantoğlu**



**Gillian
Rosenberg**



Markus Steinbach



Judith Holler