Leibniz-Zentrum
Allgemeine Sprachwissenschaft

DONDERS INSTITUTE

DFG Deutsche Forschungsgemeinschaft

ViCom

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

To interact is to move

Movement as a core constituent of social interaction

Šárka Kadavá, Marseille 2024

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

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)

Definition shapes annotation

Gesture as meaningful ← what is the meaning?

iconic, metaphoric, deictic, pragmatic, …

Gesture relates to speech ← temporal dimension

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 (Spruijt XX)

Beats derive their function from rhythmic, dynamic movement qualities

Emphasis emerges from acceleration, amplitude (and effort)

Movement is continuous

Gesture phases simplify; hide variations, and cannot represent the dynamic nature of gesture

Timing is only one layer of multimodal coordination

xxx

Problem II: practice

Resources (esp. time)

Problem II: practice

„Reliability“

ELSEVIER

Language & Communication

Volume 105, November 2025, Pages 22-36

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

Ewa Jarmołowicz-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?

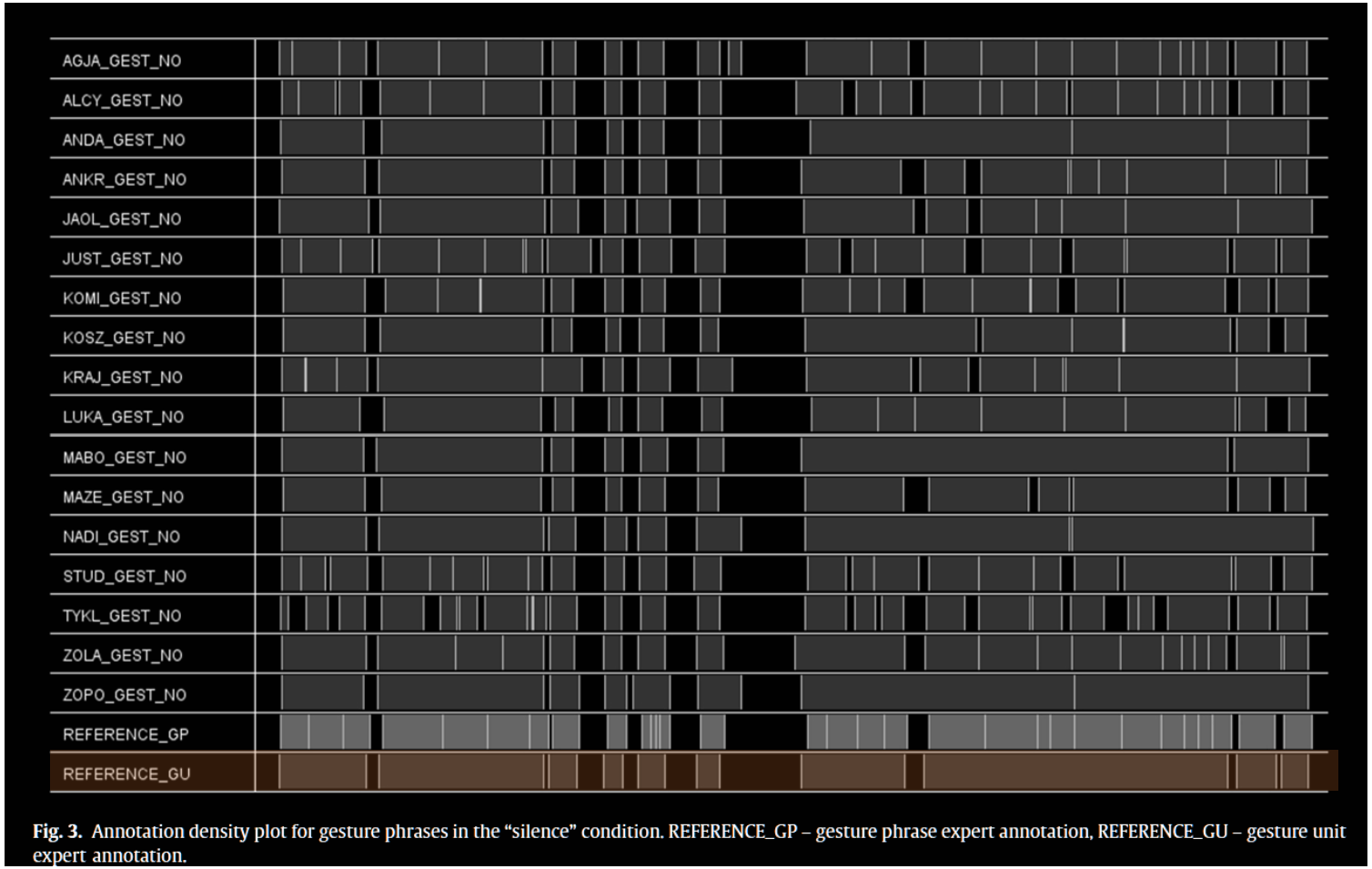


Fig. 3. Annotation density plot for gesture phrases in the "silence" condition. REFERENCE_GP – gesture phrase expert annotation, REFERENCE_GU – gesture unit expert annotation.

What is gestural prominence?

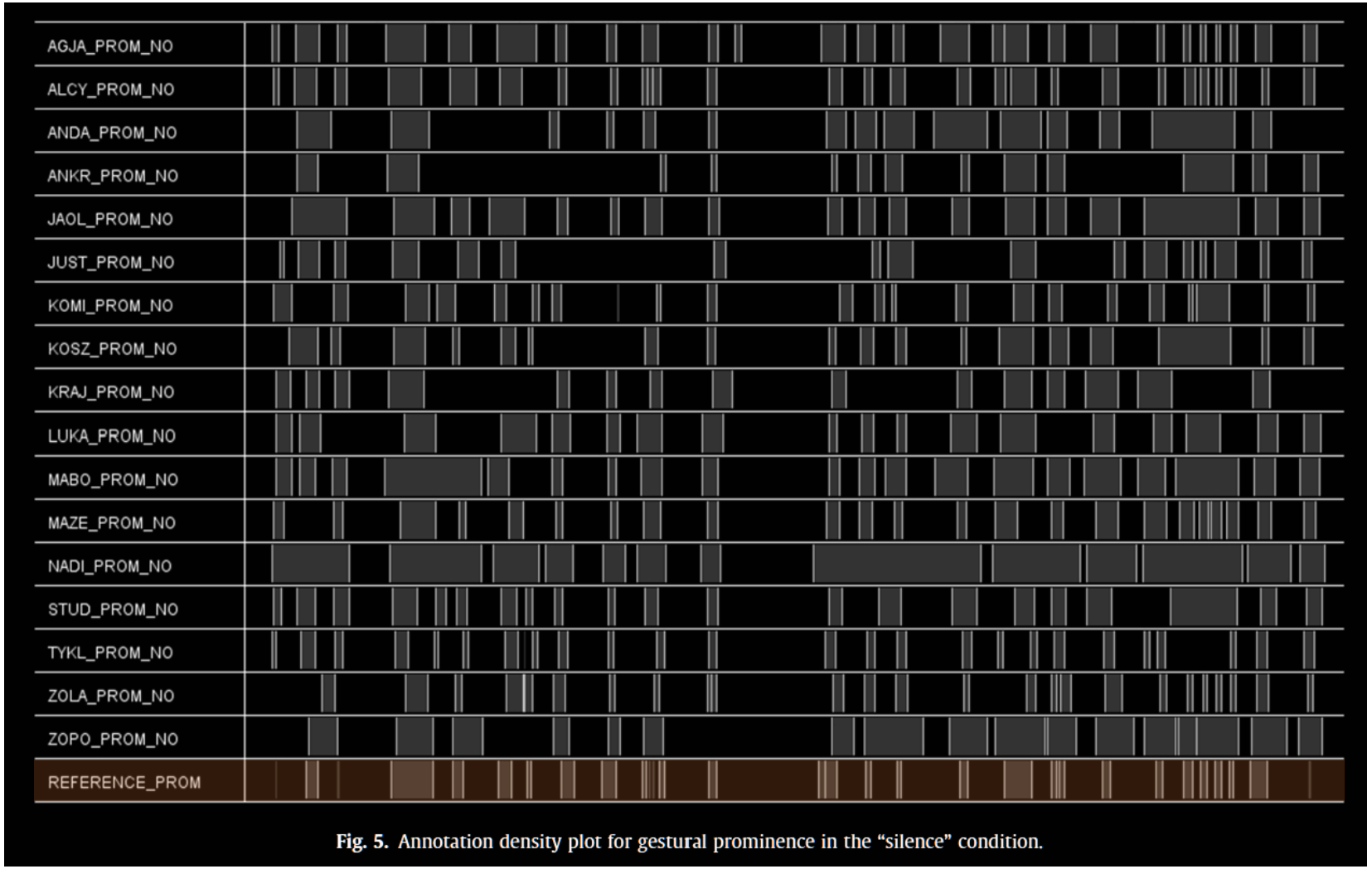


Fig. 5. Annotation density plot for gestural prominence in the "silence" condition.

Interrim summary I

😊

Flexible

Powerful, esp. for quali-analysis

Somewhat consensus around annotation schemes

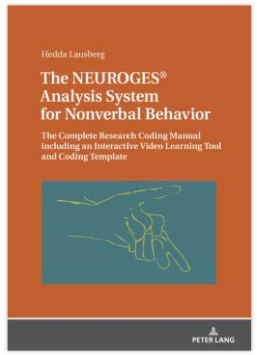
😕

Laborious

Can be quite slow and buggy

Different types of annotation schemes

Perception-dependent



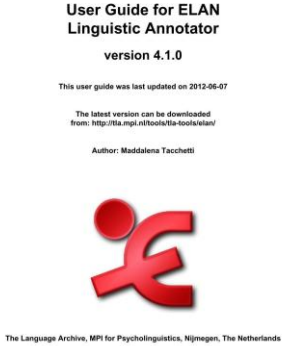
Linguistic Annotation System for Gestures (LASG)

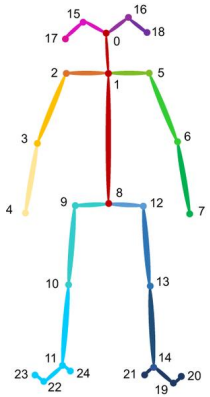
September 2013

DOI: 10.1515/9783110261318.1098

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 Bressem · Silva Ladewig · Cornelia Müller



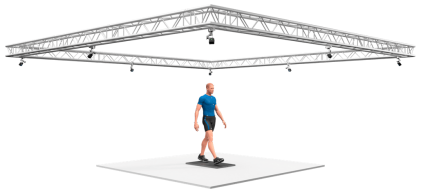


2014

2000

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

Gabriella Vigliocco ✉, Pamela Perniss and David Vinson



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

DONDERS Institute
MOBILE MULTIMODAL LABORATORY

Davide Ahmar (davide.ahmar@ru.nl)
Šárka Kadavá (sarka.kadava@donders.ru.nl)
Wim Pouw (wim.pouw@donders.ru.nl)

EnvisionBOX

New possibilities for gesture!

Automatic detection

Quantification

Precise temporal alignment

What does it mean when I say gesture is movement?

It is material/physical and continuous/dynamic

It has measurable properties

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

Dimensions of movement

II. Kinetics/dynamics describes the cause of motion

How effortful?
How coordinated?

Forces, joint torques

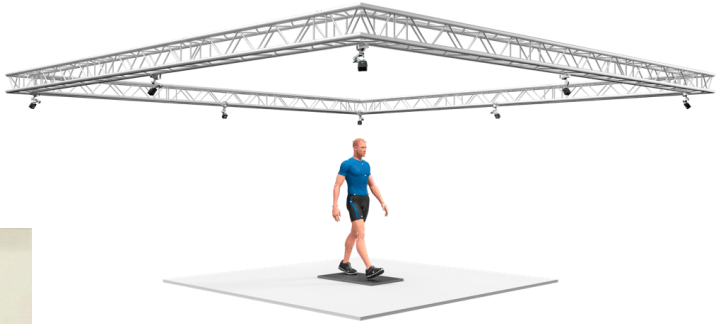
Dimensions of movement

III. Others

Volume
Hand shape configuration
Coordination
Rhythmicity

COMPARISON

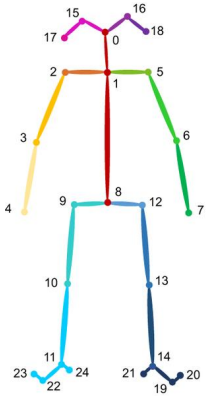
(Rather quick) intro to motion tracking

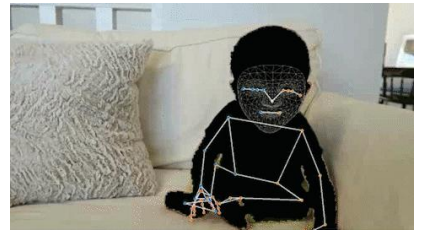


Very high precision (ca. 0.5 mm)

Lab-based

Proprietary (→ expensive)



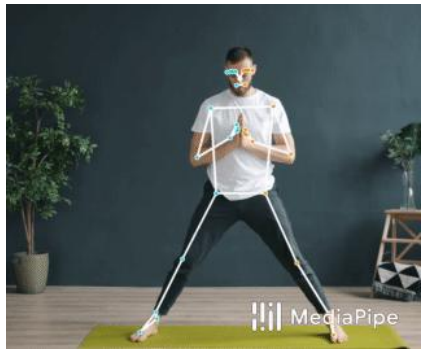


Lower precision (1-2.5 cm)

No markers needed

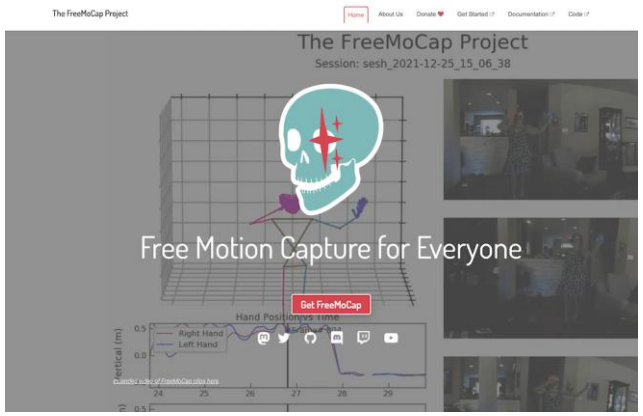
Open-source

Most common marker-less tracking

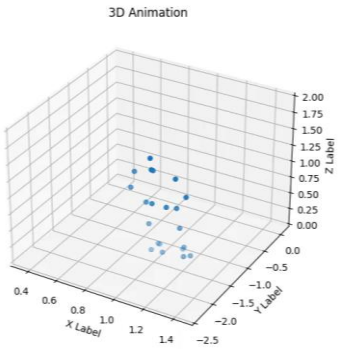


MediaPipe

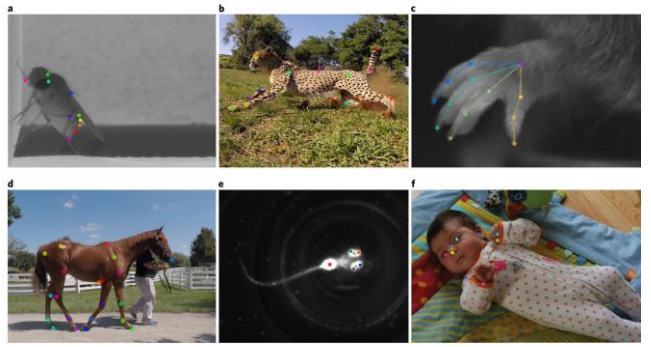
OpenPose



FreeMoCap



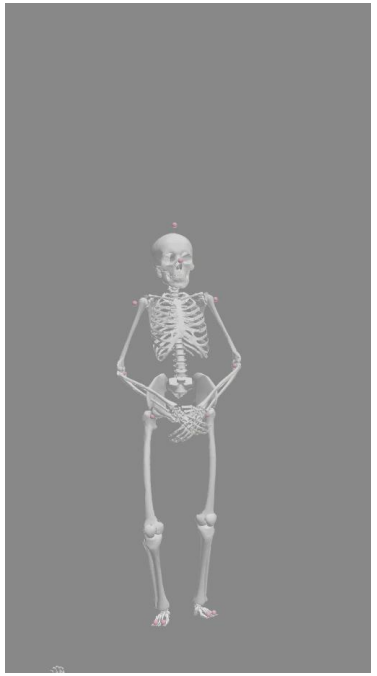
Pose2sim



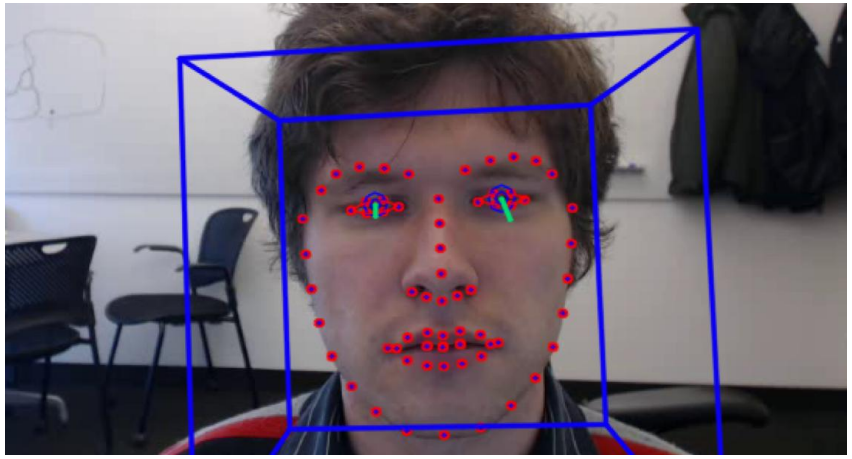
DeepLabCut

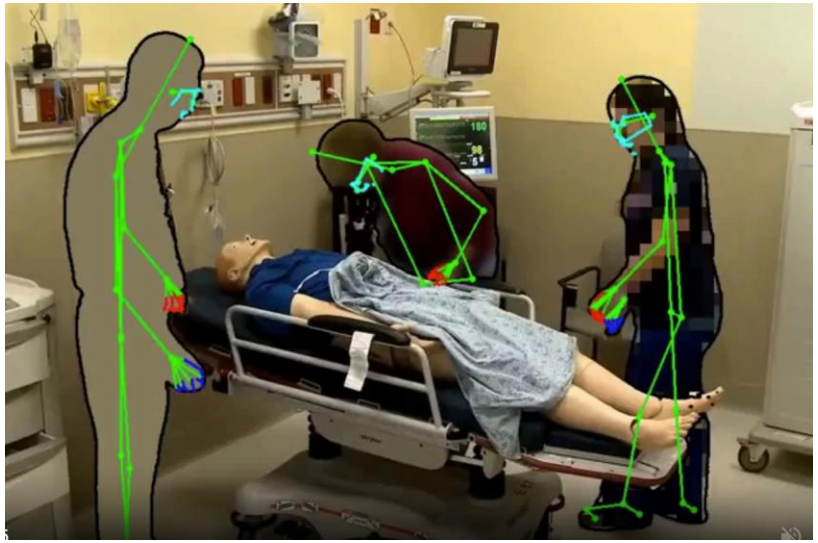
Add-ons

OpenSim



OpenFace





MaskAnyone

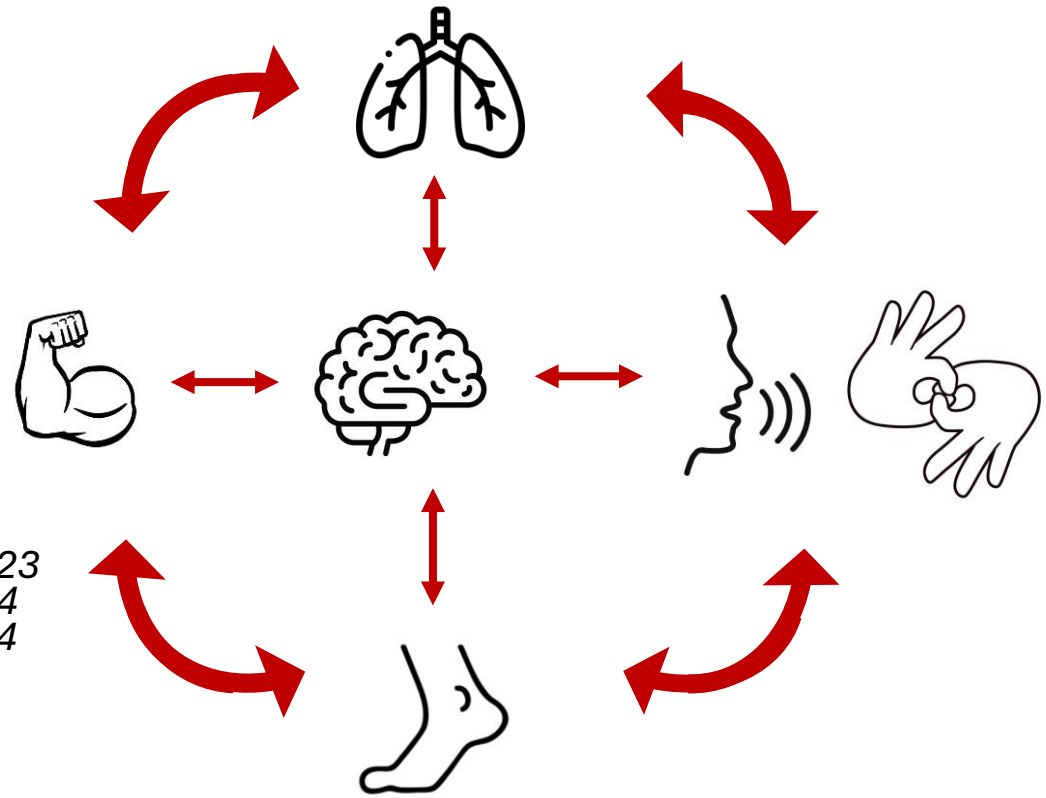
Considerations

HANDS-ON

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

~ don‘t forget about **other parts of body** involved in meaning-making

Speech Prosody 2010
Chicago, IL, USA
May 10-14, 2010

Are torso movements during speech timed with intonational phrases?

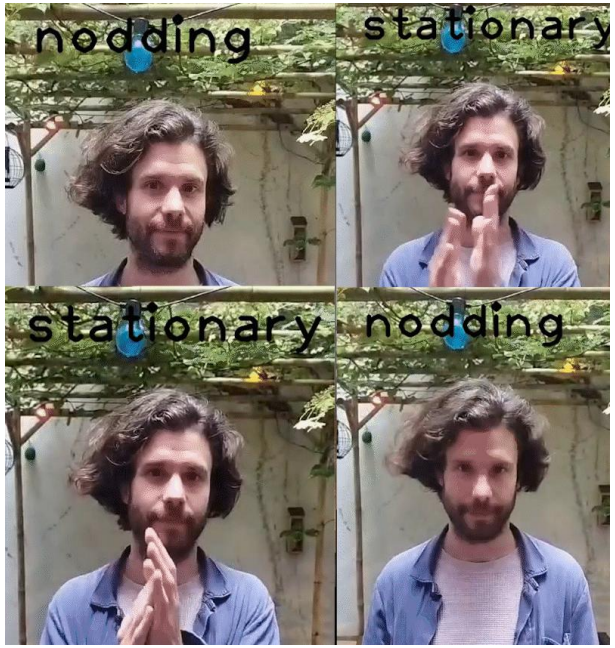
Stefanie Shattuck-Hufnagel[1], *Pei Lin Ren*[1] *and Elizabeth Tauscher*[2]

[1] Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge MA USA [2] Wellesley College, Cambridge MA USA

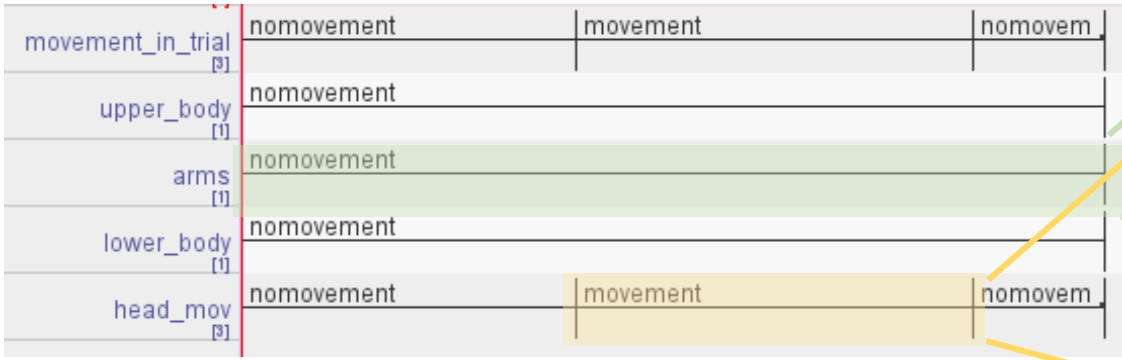
sshuf@mit.edu, pelire@mit.edu, etausch@wellesley.edu

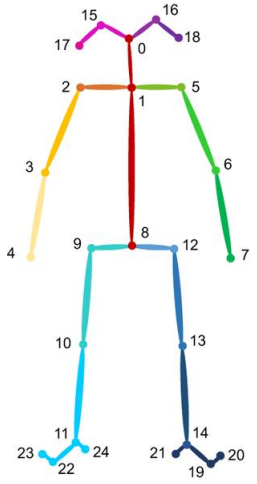
Final remarks

~ 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?





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 Ripperda[1] · Linda Drijvers[1,2] · Judith Holler[1,2]

Published online: 23 January 2020

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

Walter Dych[1], Karee Garvin[1,2], Kathryn Franich[1,2]

[1]University of Delaware, [2]Harvard 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

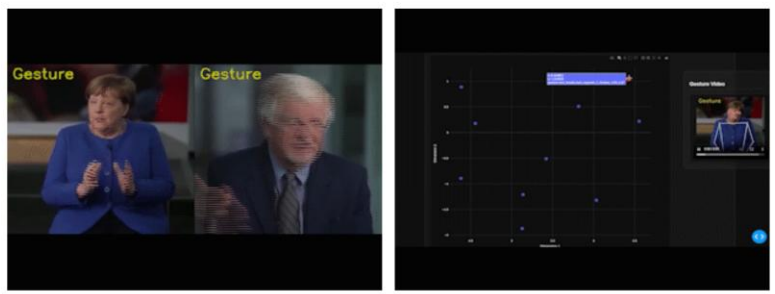
Marlou Rasenberg
Meertens Institute
The Netherlands

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

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

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)



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

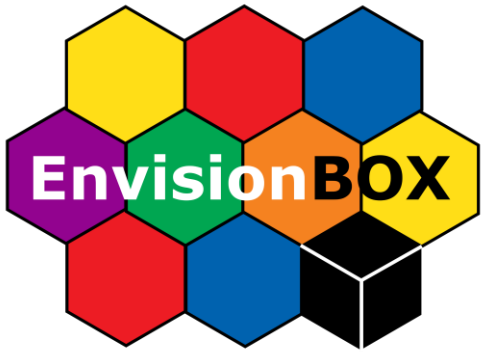
… data are just data without theory

Let‘s think (and act) together…

~ how can we utilize our data

1) to make our life (relatively) easier in the future
2) to enrich our own analyses

~ how can we use movement to study social interaction across groups and/or species on a larger scale



stay tuned for summer school (June 2025)

Leibniz-Zentrum
Allgemeine Sprachwissenschaft

DONDERS INSTITUTE

DFG Deutsche Forschungsgemeinschaft

ViCom

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Thank you for your effort!

kadava@leibniz-zas.de

Sketches were retrieved from sketchyideas.co, people illustrations from openpeeps.com