Visualizing Multimodal Data

LOT Winter School 2024, Šárka Kadavá

Content of the workshop

01

Why should we care?

02

What is possible?

03

Plots for multimodal data

04

Dashboards

01

Why should we care?

Imagine…

You spend hours and hours on **designing** an experiment, **piloting** the setup, **recording** the data…

You are done, and you start to process the data – **extract** features, **explore** their relationships, building **statistical** models, etc.

And when you are finaly done...

IS GESTURE-SPEECH PHYSICS AT WORK IN RHYTHMIC POINTING? EVIDENCE FROM POLISH COUNTING-OUT RHYMES

Šárka Kadavá[1,3,*], Aleksandra Ćwiek[1], Katarzyna Stoltmann[2], Susanne Fuchs[1], Wim Pouw[3]

[1]Leibniz Center General Linguistics, Berlin, Germany [2]adesso SE, Berlin, Germany [3]Donders Institute for Brain, Cognition and Behaviour, Nijmegen, Netherlands
[*] corresponding author: kadava@leibniz-zas.de

ABSTRACT

'Gesture-speech physics' refers to a possible biomechanical coupling between manual gesture and speech. According to this thesis, rapid gesturing leaves a direct imprint on acoustics (intensity, F0), as gesture accelerations/decelerations increase expiratory forces and therefore subglottal pressure, leading to higher amplitude envelope peaks and higher F0 values. This acoustic effect has been reported in lab experiments, spontaneous speech, clinical studies, and professional vocal performers. The current study investigates this phenomenon in Polish counting-out rhymes, using motion capture data and acoustic recordings from 11 native Polish speakers. Following the gesture-speech physics thesis, we expect acceleration/deceleration peaks to be correlated with speech intensity/F0. Through Bayesian analyses, we obtained a weak but reliable coupling of deceleration of the pointing hand and the nearest peak in the smoothed amplitude envelope.

Keywords: pointing gestures, motion tracking, poetry, prosody, coupling

1. INTRODUCTION

The evidence that gesture-speech coordination on the prosodic level arises out of basic properties of physiology and motor control is increasing [1, 2, 3, 4, 5, 6]. This contrasts with the argument that gesture is a sophisticated cognitive achievement, proliferating due to cultural conventionalization [7].

While not downplaying either of those constraints, according to the gesture-speech physics account (see [6]), there is a biomechanical nudge for aligning peaks in F0 and amplitude envelope with the peak of the physical impulse. As such, the human voice receives an 'imprint' due to the gestural activation of expiration-related muscles. Specifically, upper limb acceleration and deceleration affect rib-cage movement and

studies, showing that neural networks trained on acoustics and body kinematics can come to predict the presence of gesture or kinematic properties of gestures [8, 9, 10].

Although gesture-speech physics seems robust in some tasks, a recent study on leg and arm biking suggests that acceleration may need to reach a certain threshold to affect speech acoustics [11]. This is in line with previous research showing that body parts with lower mass (hand vs. arm) have much weaker effects on speech (e.g., [12]), if at all [13].

The reason why the biomechanical gesture-speech coupling is weak is likely because there must remain the flexibility to speak in certain ways when gesturing. The larynx should indeed be flexible to resist the effect of motion at times it is appropriate to do so. After all, the primary function of the larynx is to act quickly and protect the lungs from inhaling foreign bodies [14].

This study replicates the basic kinematic-acoustic coupling findings from previous research. Our dataset consists of motion data recorded while performing Polish counting-out rhymes involving pointing movement. During a counting-out rhyme game, one person speaks a rhyme while rhythmically moving their index finger between themselves and another person. Having clear turning points, these childhood poems are a valid paradigm to investigate speech-gesture physics, as appreciated by previous studies [15].

We extend previous work by studying forward and backward pointing movement and speech rate as additional factors. So far, only flexion-extension movements have been studied [16] and it is possible that different movements will have different respiratory interactions. Looking at a faster rate is motivated by the fact that this may go hand in hand with larger accelerations/forces. However, the rate may also change the complexity of gesture-speech physics, as different coupling strengths and muscular stiffness are involved.

28. Multimodal Phonetics ID: 955

Model	Parameter	Estimate [95% CrI]	$Pr(\beta <$ or $> 0)$
F0 acc	Intercept	185.58[172.59; 198.94]	
	Speech rate	8.16[1.70; 14.51]	0.99
	Sex	104.07[77.81; 129.97]	1.00
F0 dec	Intercept	177.12[162.24; 191.94]	
	Speech rate	5.04[0.20; 9.90]	0.98
	Sex	100.84[71.03; 130.46]	1.00
ENV acc	Intercept	0.82[0.72; 0.93]	
ENV dec	Intercept	0.70[0.54; 0.87]	
	Wrist dec. peak	−0.02[−0.04; −0.00]	0.99

Table 1: The table lists out all intercepts, as well as the parameters with a reliable effect on the outcome variables of the four models, with posterior means and the 95% CrI. The rightmost column is the posterior probability of the effect to be below or above 0, depending on the direction.

on amplitude deceleration peak is not reliable, however, since the task was not heavily cont[illegible] we acknowledge that those relation[illegible] should be further studied in the future. [illegible] be seen in Figure 2, the ma[illegible] *[illegible]eration peak* on the [illegible] envelope peak we [illegible] bove diff[illegible] it is within backward vs. [illegible] ment. Nevertheless, given the priors, the da[illegible] [illegible]d the model, there was no interaction effect of the two parameters ($\beta = -0.00[-0.02; 0.01]$).

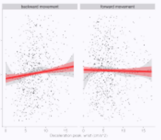


Figure 2: The linear relationship between deceleration peak and amplitude envelope peak. [illegible]shed non-linear 'loess' line reflects possible [illegible]arities. Note that the deceleration val[illegible] have [illegible]lutized.

4. DISCUSSION

The current study goes beyond previous research on gesture-speech physics by assessing statistical coupling in (1) multi-directional (2) pointing movements in (3) the Polish language. Our findings suggest that deceleration peaks scale to their nearest amplitude envelope peak, rather than F0. This

envelope or F0 peaks.

Why did the rapidity by which participants halted a pointing movement (i.e., deceleration) not scale to the nearest F0 peak? The gesture-speech physics thesis proposes that there is a mechanical interaction between an upper limb and the body during acceleration or deceleration. The physical impulse of a upper-limb movement produces a mechani[illegible] loading onto the rib cage, which limits its [illegible]ent and impacts subglottal pressures ne[illegible] for voice production. Subglottal [illegible] are primarily linked with affecting i[illegible], and only secondarily F0, which is u[illegible]re flexible laryngeal control.

That [illegible] a coupling of deceleration rather than [illegible] might look like a counterargument for gesture-speech physics. However, comparing the absolute raw values of deceleration and acceleration peaks, we found 20% lower magnitudes for acceleration than deceleration. In line with [11], we suppose that a certain threshold needs to be reached before a significant effect of physics arises.

As for the deceleration effect alone, it is known that speakers coordinate their emphasis in speech [illegible]ith the moment when the limb movement reaches [illegible]destination [27]. Thus, emphasis is generally n[illegible]cated at the initial stage of pointing; rather, it occ[illegible] when reaching the intended target.

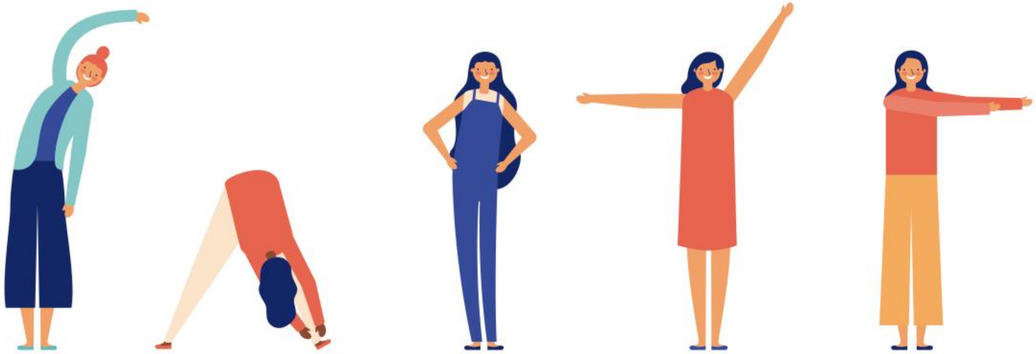
Fu[illegible]r, we did not find that kinematic peaks affect[illegible] speech differently depending on the directi[illegible] of movement. This means forward and ba[illegible]ard movements along the sagittal plane likely p[illegible]urb vocalization by increasing subglottal pressu[illegible] much like flexion-extension movements along [illegible] frontal plane [16].

Fo[illegible]uture research, potential alternative hyp[illegible]es should be investigated. For example, oth[illegible]inematic variables (e.g., speed) need to be [illegible]sed for speech coupling. Our study is also [illegible]ited in the number [illegible] [illegible] has a large number of trials and events that have been analyzed, increasing the reliability of the reported effects within our sample. Moreover, since we do not directly measure muscle activity in relation to respiratory-vocal states, it is always possible to maintain that the current kinematic-acoustic effect is solely a neurally controlled achievement. Such an explanation requires an auxiliary hypothesis about why the brain would monitor acceleration peaks and couple them to vocalization. While we deem it possible that the brain is tuned like this, it would be precisely because there is a weak biomechanical coupling to

...this is what remains from my beautiful data

Where are the videos that took so much time to collect?

Where is the movement that we claim to be so important?



02

What is possible?

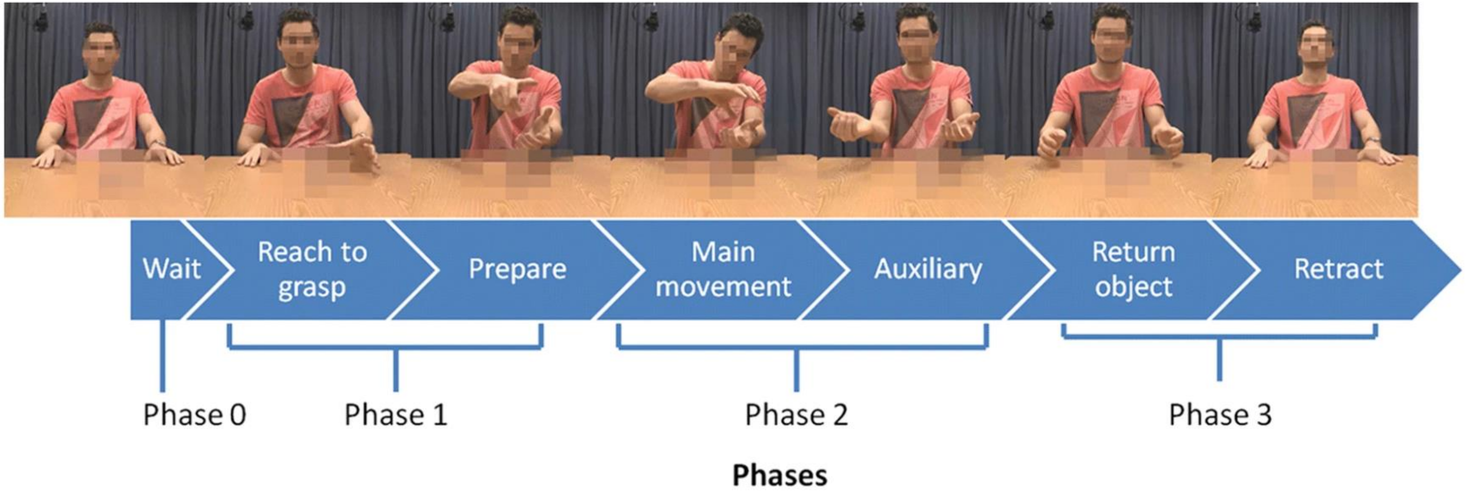
Imagine...

You have an experiment in which people do some movements and sounds. Now you are writing a paper...

How do you make use of these data?

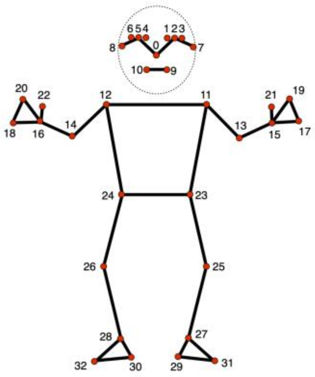
Example

(Trujillo et al., 2020)

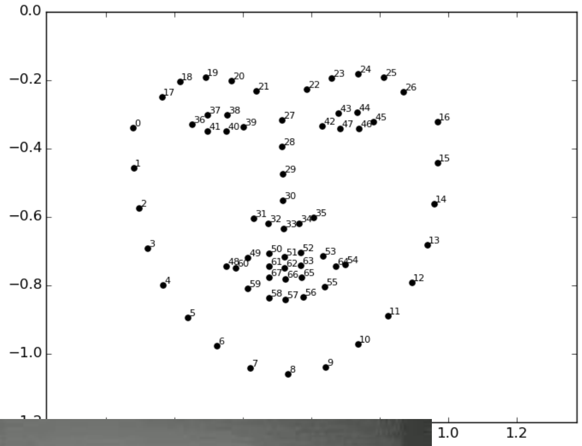


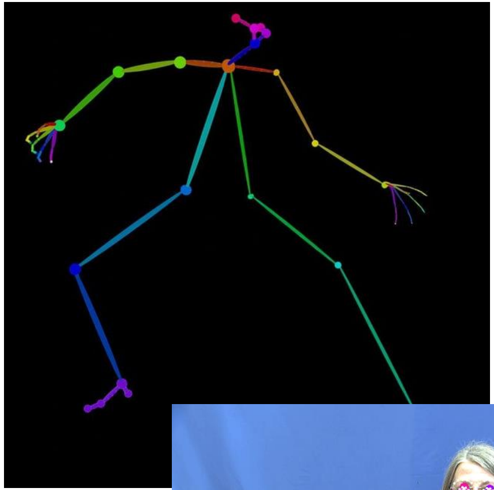
Examples

MediaPipe

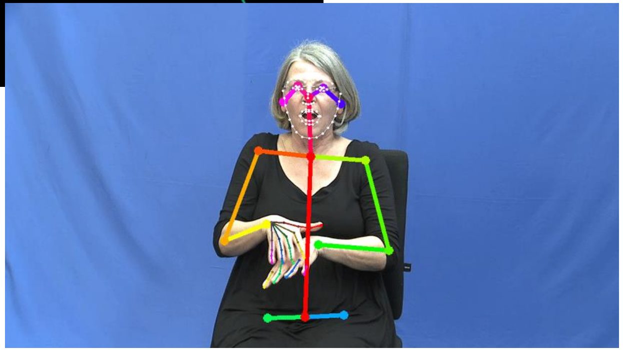


OpenFace

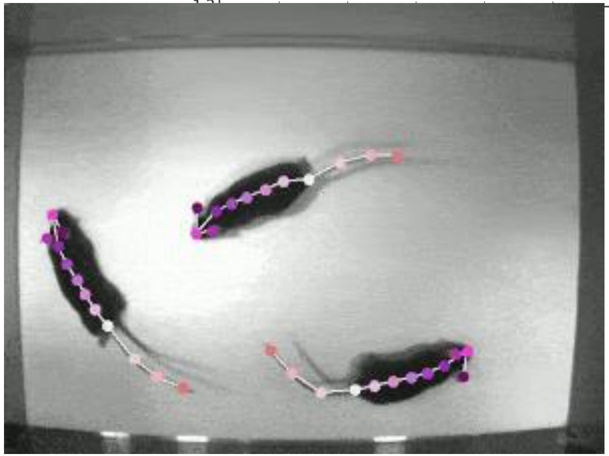




OpenPose



Schulder & Hanke, 2020



DeepLabCut

Example

(Pouw et al., 2023)

https://tsg-131-174-75-200.hosting.ru.nl/siamang/

03

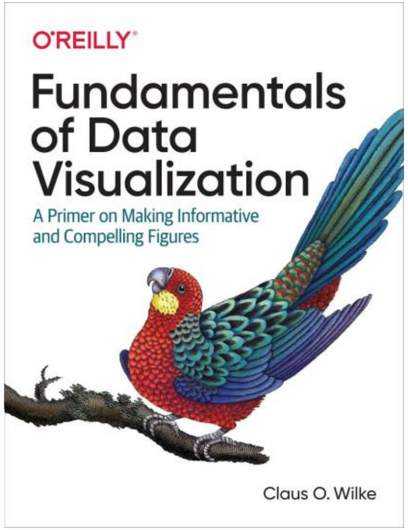
Plots for multimodal data

Friends Don't Let Friends Make Bad Graphs

DOI 10.5281/zenodo.7542491

Friends don't let friends make certain types of data visualization - What are they and why are they bad.

- Author: Chenxin Li, postdoctoral associate at Center for Applied Genetic Technologies, University of Georgia.
- Contact: Chenxin.Li@uga.edu | @ChenxinLi2



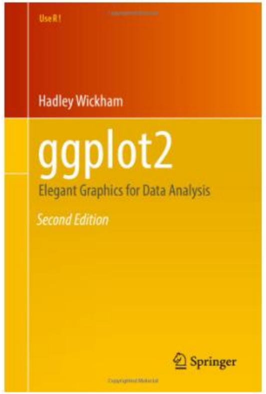
Charming Data

@CharmingData · 37.3K subscribers · 170 video

Charming Data is a Community of people inspire

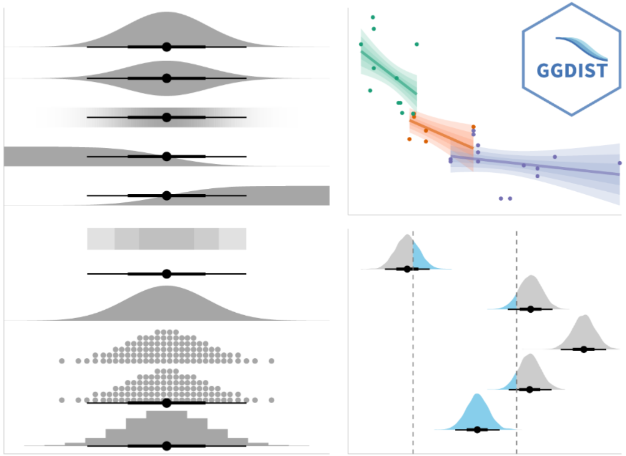
github.com/Coding-with-Adam/Dash-by-Plotly ar

Subscribed   Join



ggdist: Visualizations of distributions and uncertainty

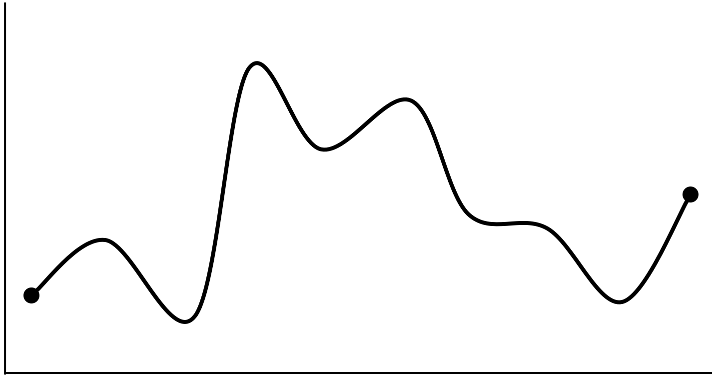
R-CMD-check failing | codecov 98% | CRAN 3.3.1 | downloads 13K/month | DOI 10.5281/zenodo.3879620



Multimodal essentials

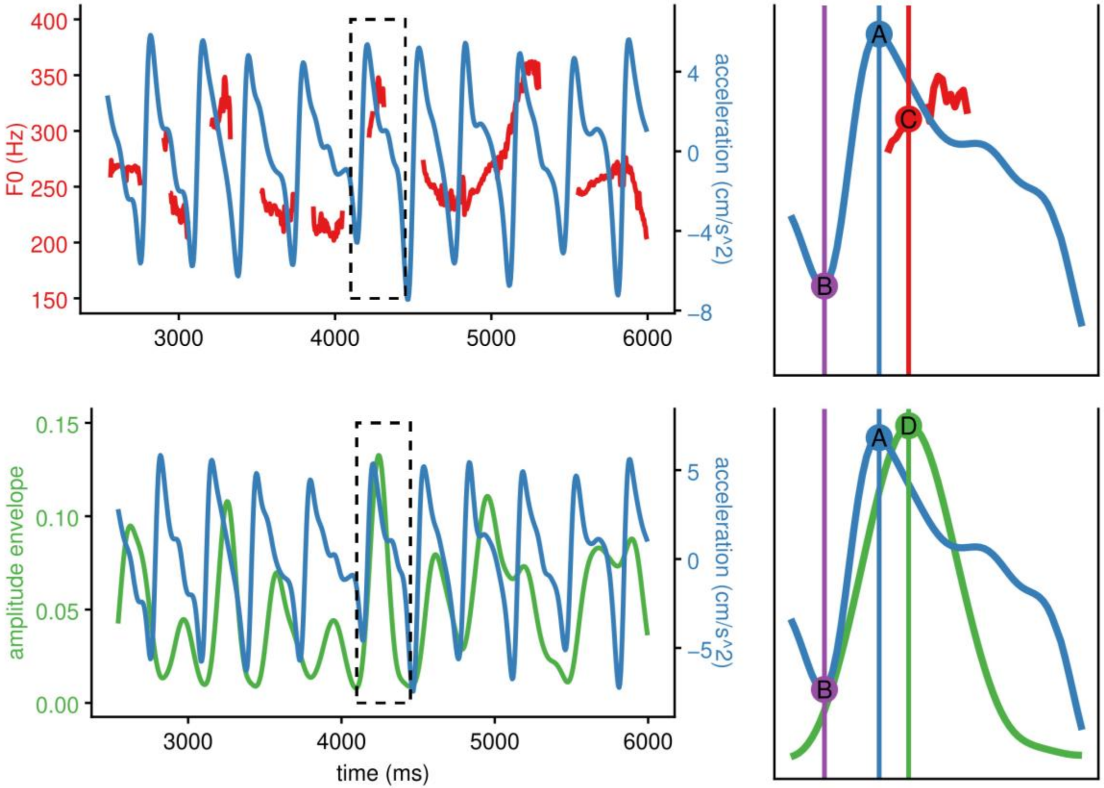
Our data

- change **in time**
- consists of **various signals**
- that have **variety of characteristics**



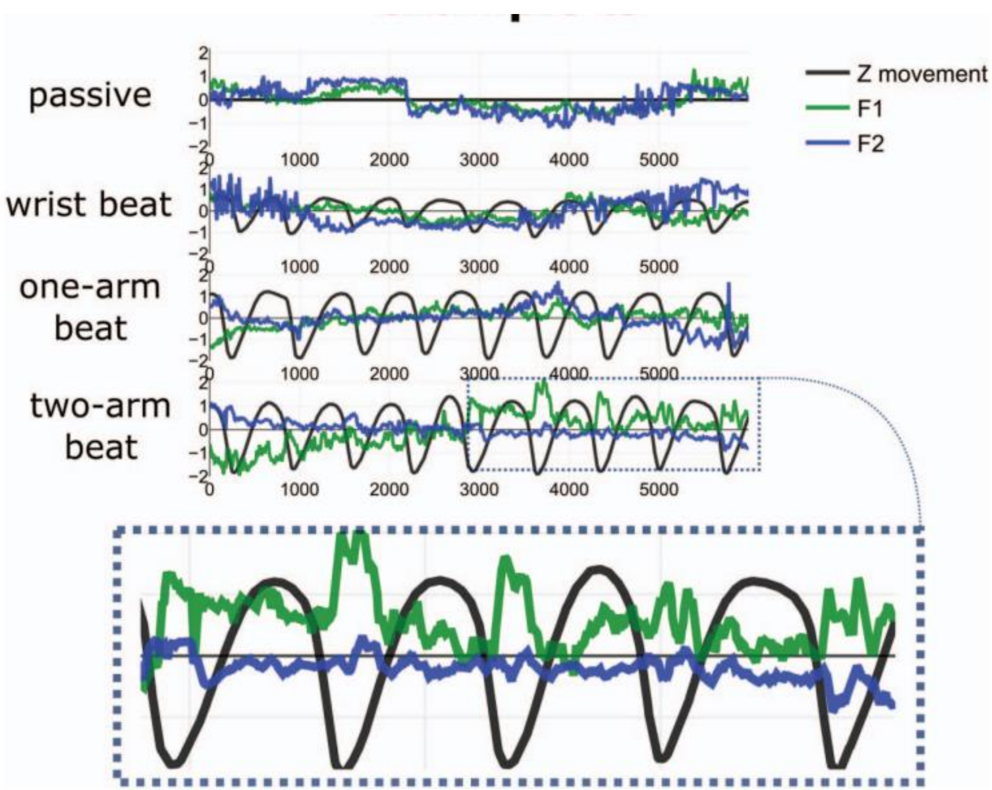
Example

(Kadavá et al., 2023)



Example

(Pouw et al., 2020)



04

Dashboards

Dashboards

App that allows **user** to display various types of **(visual) data**

→ it provides **direct access** to our (visual) data

Key components

Data

Html/dash components

Default interface

Update

Let’s get to work

Last but not least

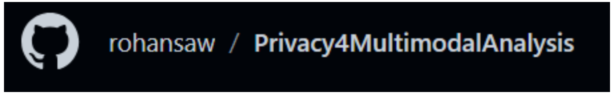
You need to upload the app to server to be able to use it online

Last but not least

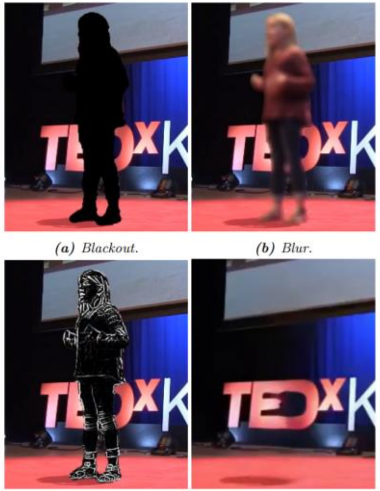
Be aware of privacy!

Masking tools

MaskAnyone



(a) Original. (b) Olaf Scholz.



(a) Blackout. (b) Blur.

(c) Contours. (d) Inpainting.

Masked-Piper (Owoyele et al., 2022)

Thank you

kadava@leibniz-zas.de

sarkadava

References

Kadavá Š., Ćwiek A., Stoltmann K., Fuchs S., Pouw W. (2023). Is gesture-speech physics at work in rhythmic pointing? Evidence from Polish counting-out rhymes. In: R. Skarnitzl & J. Volín (Eds.), Proceedings of the 20th International Congress of Phonetic Sciences. Guarant International.

Owoyele, B., Trujillo, J., De Melo, G., & Pouw, W. (2022). Masked-Piper: Masking personal identities in visual recordings while preserving multimodal information. SoftwareX, 20, 101236.

Pouw, W., Harrison, S. J., & Dixon, J. A. (2020). Gesture–speech physics: The biomechanical basis for the emergence of gesture–speech synchrony. Journal of Experimental Psychology: General, 149(2), 391.

Pouw, W., Kehy, M., Gamba, M., & Ravignani, A. (2023). Cross-modal constraints in multimodal vocalizations in Siamang (Symphalangus syndactylus).

Schulder, M., & Hanke, T. (2019). OpenPose in the Public DGS Corpus. Hamburg: University of Hamburg.

Trujillo, J. P., Simanova, I., Bekkering, H., & Özyürek, A. (2020). The communicative advantage: How kinematic signaling supports semantic comprehension. Psychological research, 84, 1897-1911.