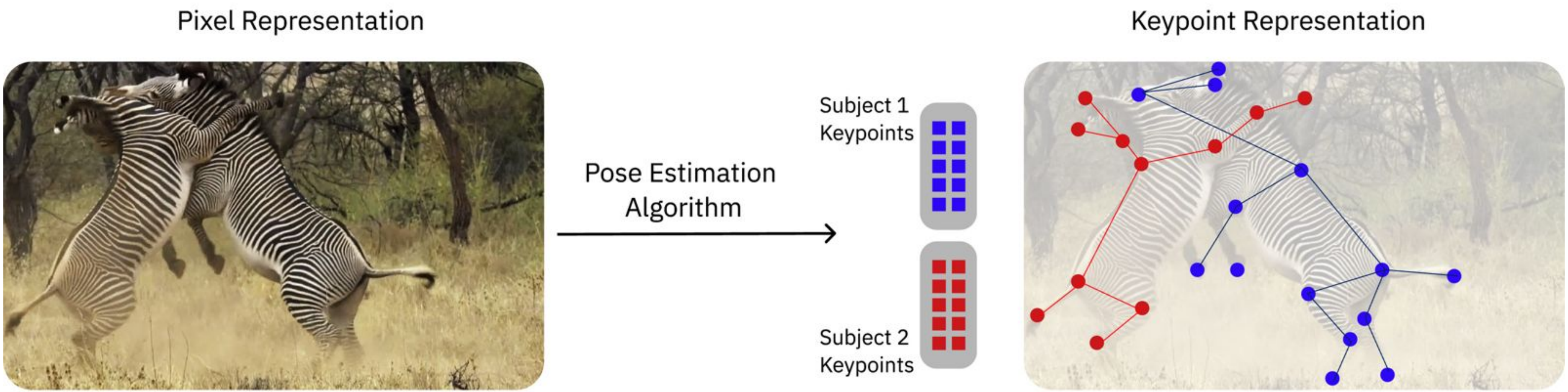
Section: Weinreb et al. 2024

Background Recap:

EPFL

Schematic Overview of Markerless Motion Capture, aka Pose Estimation



Primer

A Primer on Motion Capture with Deep Learning: Principles, Pitfalls, and Perspectives

Alexander Mathis,[1,2,3,*] Steffen Schneider,[3,4] Jessy Lauer,[1,2,3] and Mackenzie Weygandt Mathis[1,2,3,*]
[1]Center for Neuroprosthetics, Center for Intelligent Systems, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland
[2]Brain Mind Institute, School of Life Sciences, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland
[3]The Rowland Institute at Harvard, Harvard University, Cambridge, MA, USA
[4]University of Tübingen and International Max Planck Research School for Intelligent Systems, Tübingen, Germany
[*]Correspondence: alexander.mathis@epfl.ch (A.M.), mackenzie.mathis@epfl.ch (M.W.M.)
https://doi.org/10.1016/j.neuron.2020.09.017

Challenges for pose estimation in the laboratory

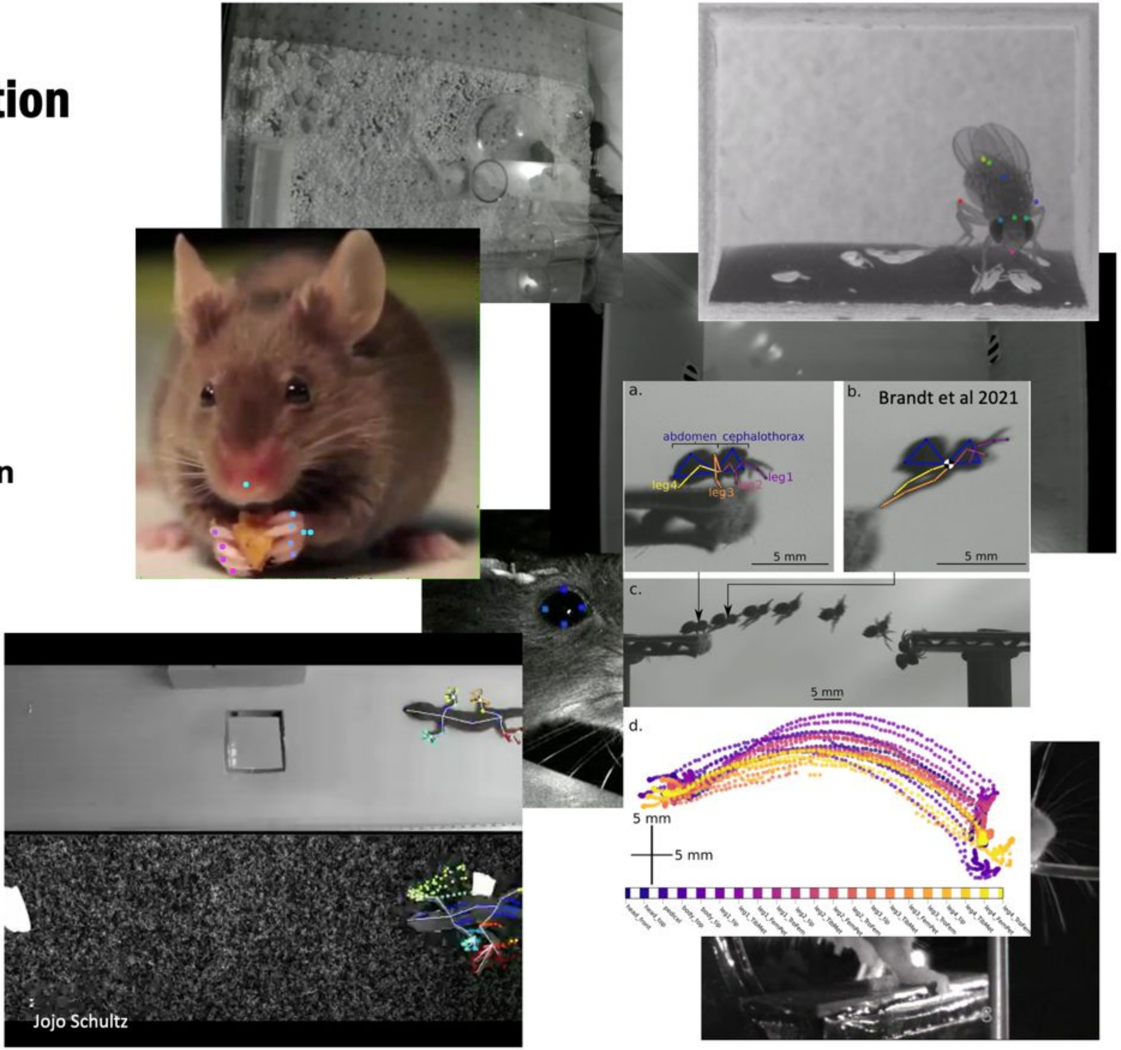
•animals have highly different bodies (i.e., can't leverage a skeleton or pose prior across all species)

•not practical for individuals to label >10,000 frames for training (i.e., human benchmark dataset sizes)

•fast real-time video analysis

•Multi-animal tracking, where animals can look truly identical

•Robust, plug-N-play solutions?



Mathis & Mathis 2020
Current Opinion in Neurobiology

Built on the open source python stack:

StatsModels Statistics in Python

xarray

scikits-image

scikit-learn machine learning in Python

Cartopy

And many, many more...

pandas

matplotlib

SciPy

bokeh

NumPy

jupyter

IP[y]: IPython

python

spyder

User testing/dev & deployment:

DEVELOPMENT

DATA & MODEL STORAGE

REPOSITORY REVISION

DeepLabCut: a software package for animal pose estimation

DLC GUI

jupyter

use our Project Manager GUI, Jupyter Notebooks, Google Colab, or terminal!

Create a project, extract frames, + GUIs to label your data

Select + Train your deep neural network

Evaluate network performance

(active learning + GUIs if improvement needed)

Run inference on new videos, create labeled videos, + plot your results!

Real-time specific tools:

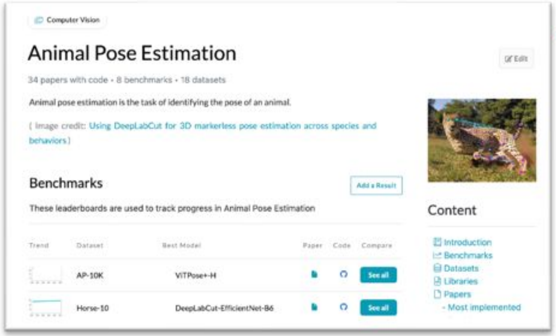
bonsai

VISUAL REACTIVE PROGRAMMING

DeepLabCut-Live! a software package for real-time animal pose estimation

AUTOPILOT

Computer Vision:


Computer Vision

Animal Pose Estimation

34 papers with code • 8 benchmarks • 18 datasets

Animal pose estimation is the task of identifying the pose of an animal.

(Image credit: Using DeepLabCut for 3D markerless pose estimation across species and behaviors)

Benchmarks

Add a Result

These leaderboards are used to track progress in Animal Pose Estimation

Trend	Dataset	Best Model	Paper	Code	Compare
	AP-10K	ViTPose+-H			See all
	Horse-10	DeepLabCut-EfficientNet-B6			See all

Content

Introduction
Benchmarks
Datasets
Libraries
Papers
- Most implemented

Larger scale pipeline computing:

docker

amazon web services

DataJoint

Post- pose estimation tools:

R

scikit learn

Classifiers: SVMs, Random Forrest, ANNs
- B-SOID, ETH-DLC Analyzer, simba

Models: HMMs, decision-trees, ANNs

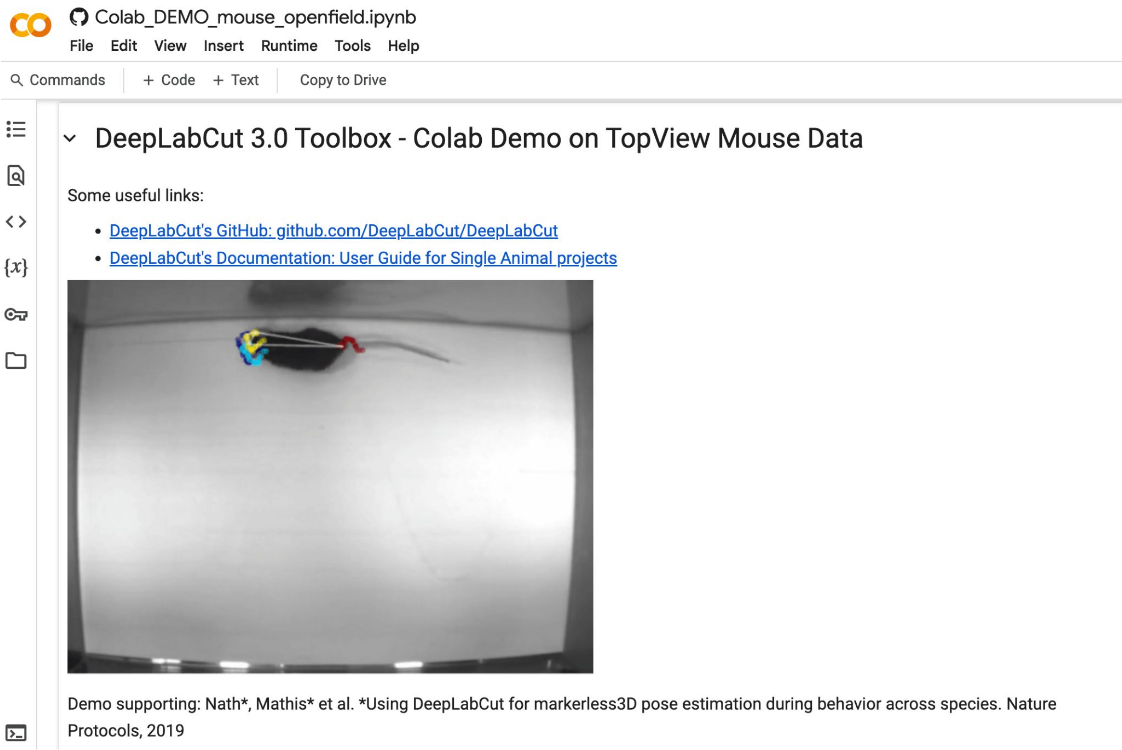
Ethograms: BORIS, BENTO, AmadeusGPT, Keypoint-MoSeq,

Clustering: CEBRA, MotionMapper, JAABA

Motor analysis: DLC2Kinematics

Hands on: DeepLabCut

https://colab.research.google.com/github/DeepLabCut/DeepLabCut/blob/master/examples/COLAB/COLAB_DEMO_mouse_openfield.ipynb



Section Paper:

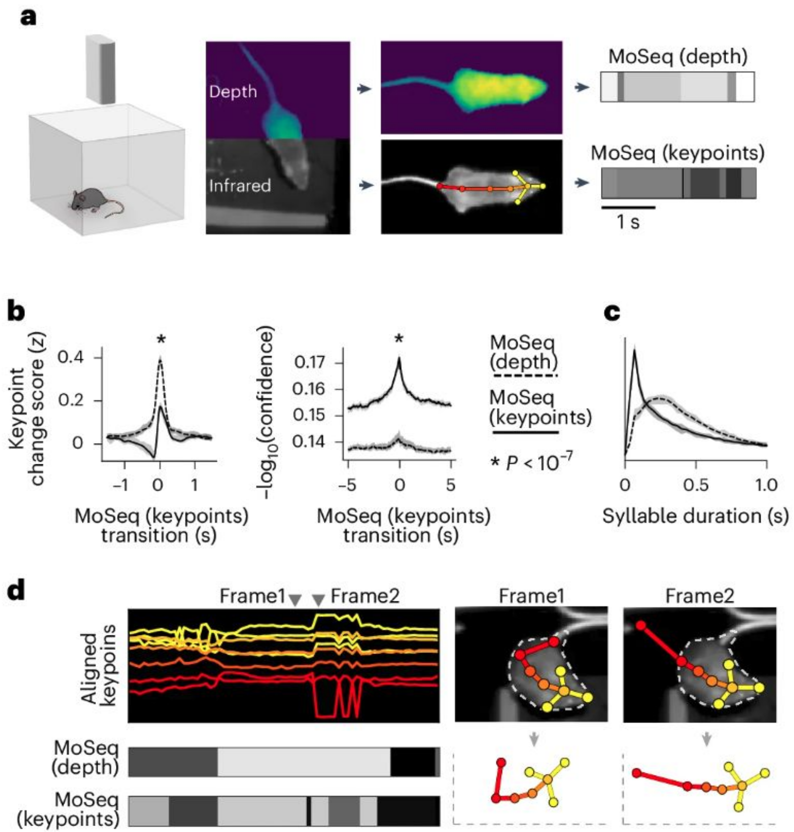
Article | Open access | Published: 12 July 2024

Keypoint-MoSeq: parsing behavior by linking point tracking to pose dynamics

Caleb Weinreb, Jonah E. Pearl, Sherry Lin, Mohammed Abdal Monium Osman, Libby Zhang, Sidharth Annapragada, Eli Conlin, Red Hoffmann, Sofia Makowska, Winthrop F. Gillis, Maya Jay, Shaokai Ye, Alexander Mathis, Mackenzie W. Mathis, Talmo Pereira, Scott W. Linderman & Sandeep Robert Datta

Keypoint tracking algorithms can flexibly quantify animal movement from videos obtained in a wide variety of settings. However, **it remains unclear how to parse continuous keypoint data into discrete actions**. This challenge is particularly acute because keypoint data are susceptible to high-frequency jitter that clustering algorithms can mistake for transitions between actions. Here we present keypoint-MoSeq, a machine learning-based platform for identifying behavioral modules (‘syllables’) from keypoint data without human supervision. **Keypoint-MoSeq uses a generative model to distinguish keypoint noise from behavior, enabling it to identify syllables whose boundaries correspond to natural sub-second discontinuities in pose dynamics.** Keypoint-MoSeq outperforms commonly used alternative clustering methods at identifying these transitions, at capturing correlations between neural activity and behavior and at classifying either solitary or social behaviors in accordance with human annotations. Keypoint-MoSeq also works in multiple species and generalizes beyond the syllable timescale, identifying fast sniff-aligned movements in mice and a spectrum of oscillatory behaviors in fruit flies. Keypoint-MoSeq, therefore, renders accessible the modular structure of behavior through standard video recordings.

Figure 1: Keypoint trajectories exhibit sub-second structure



Explain how Motion Sequencing (MoSeq) works.

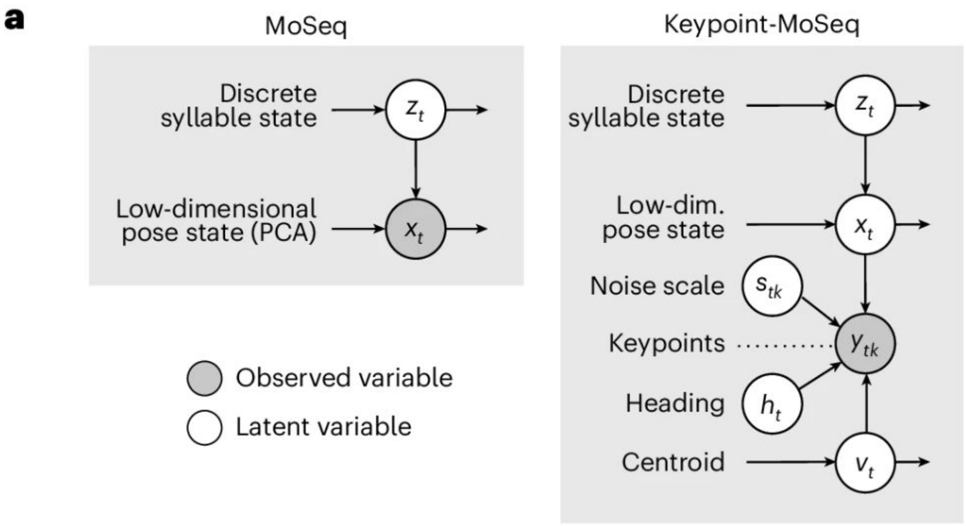
It uses unsupervised ML to transform inputs (3D depth video or keypoints) into a set of behavioral motifs (=syllables) by searching for discontinuity in the behavioral data at a customizable timescale.

Panel (b:d): What weakness of MoSeq is demonstrated here?

- *Depth MoSeq: works well but difficult to deploy, high reflection sensitivity, limited temporal resolution.*
- *Keypoints MoSeq: inability to differentiate noise (especially keypoint jitter artifact) from behavior when presented with keypoint data.*

Figure 2: Hierarchical modeling of keypoint trajectories decouples noise from pose dynamics

a



How is keypoint-MoSeq addressing this limitation? Explain the model's architecture.

Builds on SLDS

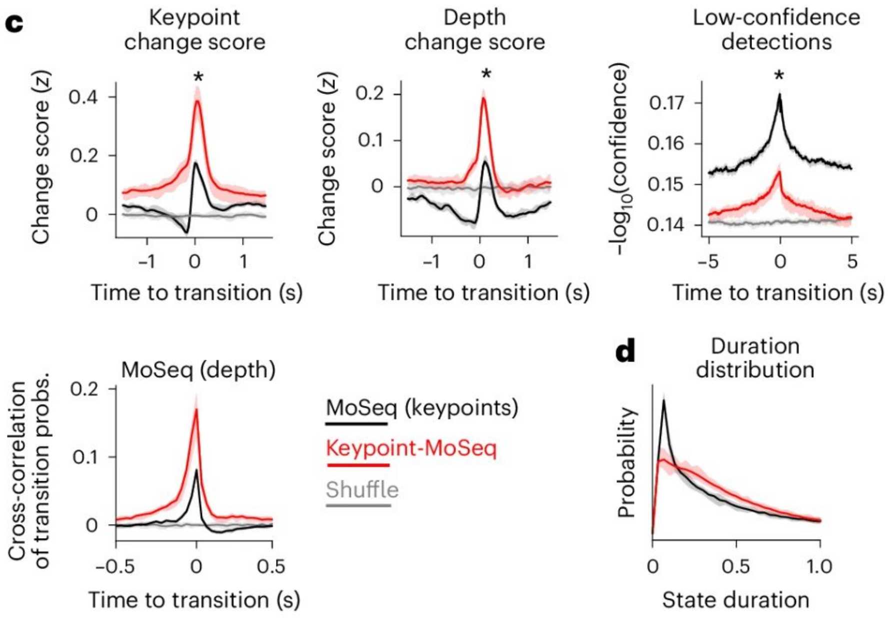
Hierarchical levels to form a dynamical system:

- *Discrete syllable state space (state z_t)*
- *Low-dimensional pose space (pose x_t)*
- *Location (centroid) information (v_t)*
- *Heading information (h_t)*

Output: keypoints coordinate (y_tk)

If jitter, it can be attributed to noise (s_tk).

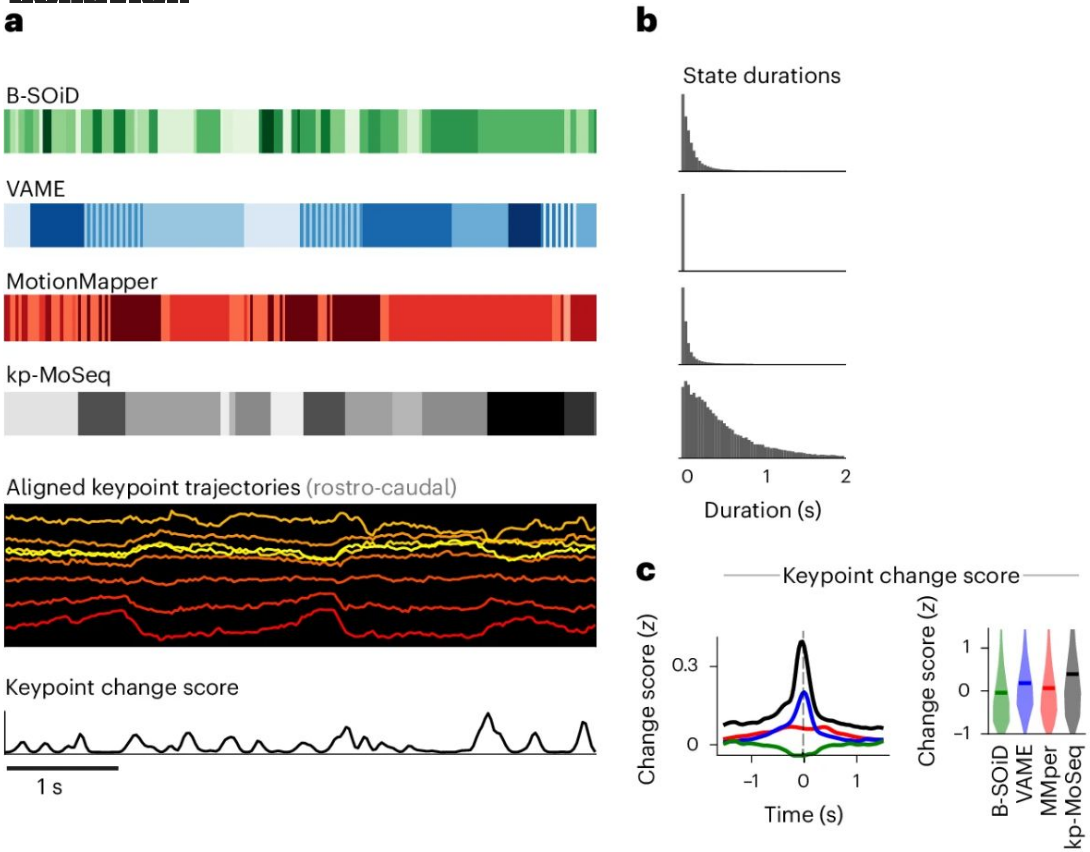
Figure 2: Hierarchical modeling of keypoint trajectories decouples noise from pose dynamics



How does it affect the the behavioral syllables transition?

Syllable transitions overlap more strongly with change points in pose as well as with syllable transition from depth MoSeq. Distribution duration overall longer.

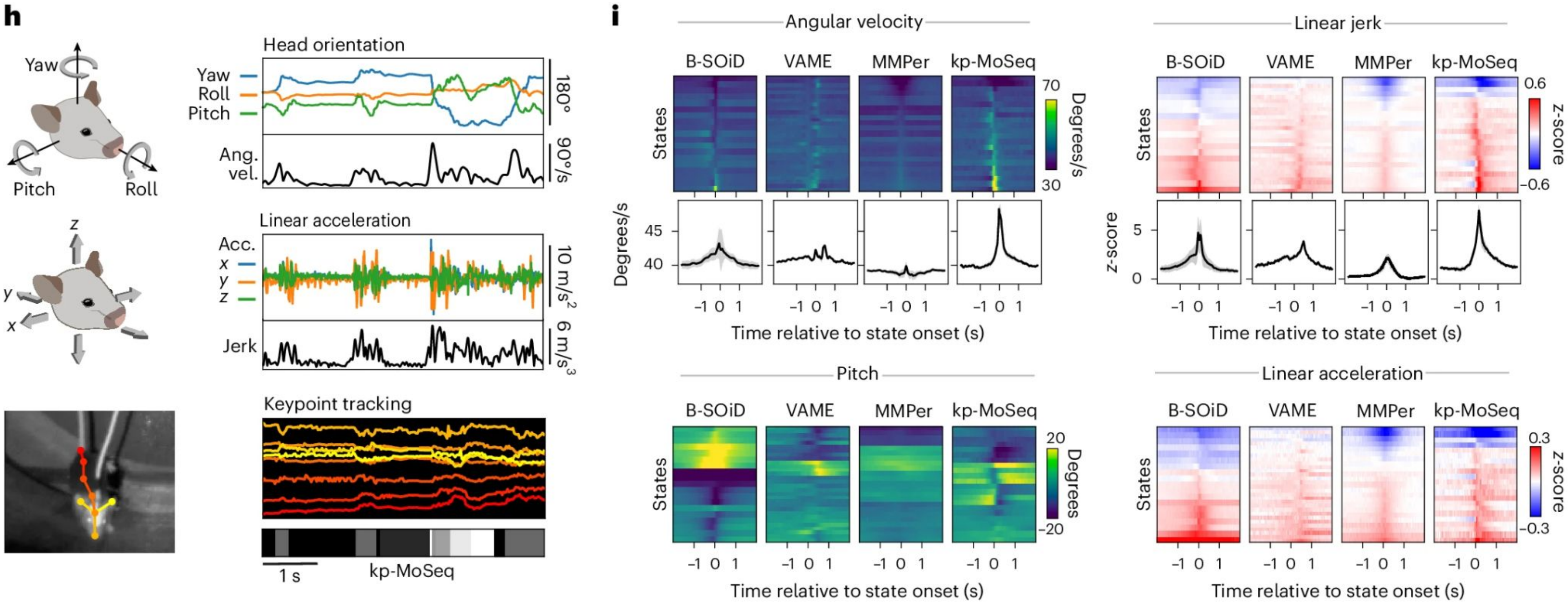
Figure 3: Keypoint-MoSeq captures the temporal structure of behavior.



How does kpts-MoSeq compare to existing methods?

Behavioral states overall longer, and transitions aligned better with keypoints change score.

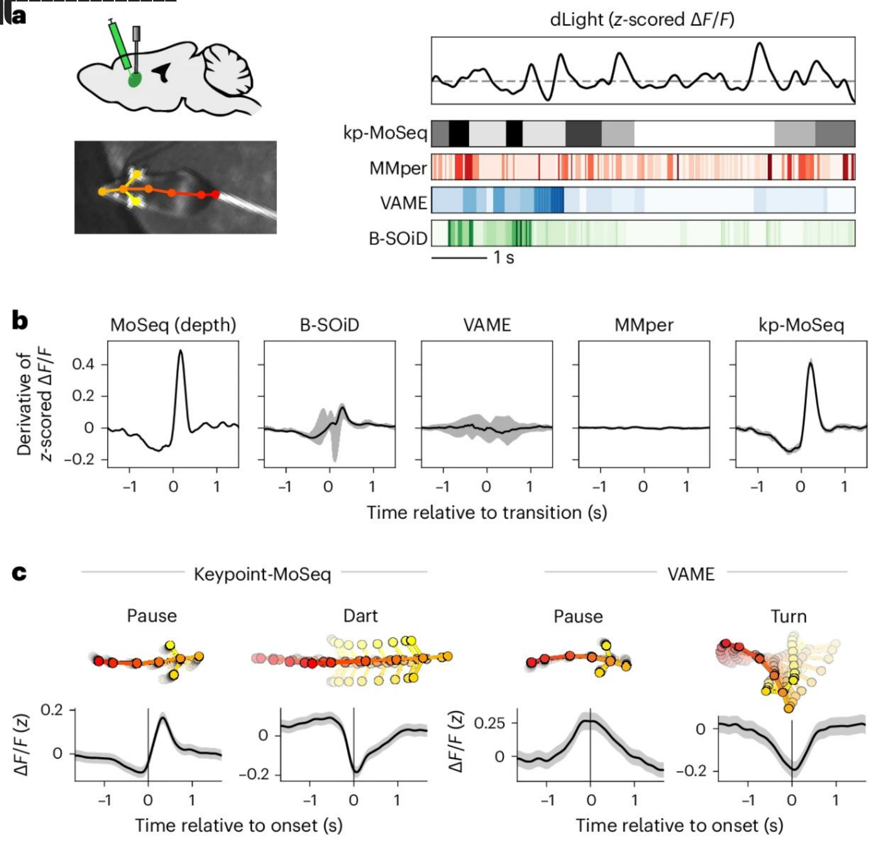
Figure 3: Keypoint-MoSeq captures the temporal structure of behavior.



Why are they looking so closely at kinematic measurements to evaluate their method?

Here, they show that it is accurately capturing changes in kinematic, which provides a reason why it more clearly identify behavioral boundaries: it represents the temporal structure of the behavior differently from the others.

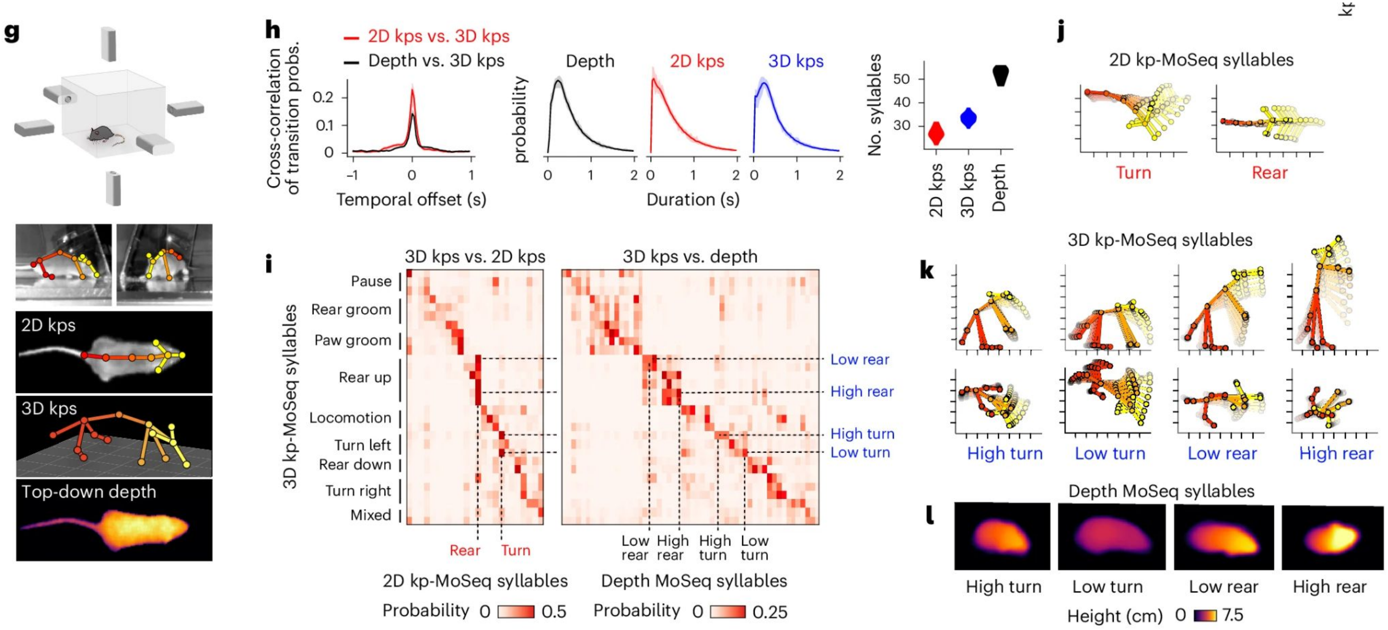
Figur e 4: Keypoint-MoSeq syllable transitions align with fluctuations in striatal dopamine



How does dopamine fluctuate around syllables transition of different models ?

Dopamine fluctuation aligns best to MoSeq and Keypoint-MoSeq transitions whereas the alignment to B-SOiD and VAME shows a lot of variability.
There is an increase in dopamine after a pause and a decrease around a dart shown for syllables identified by both Keypoint-MoSeq and VAME but Keypoint-MoSeq transition alignment shows ***dopamine dynamics that are more time-locked to the onset of the syllable.***

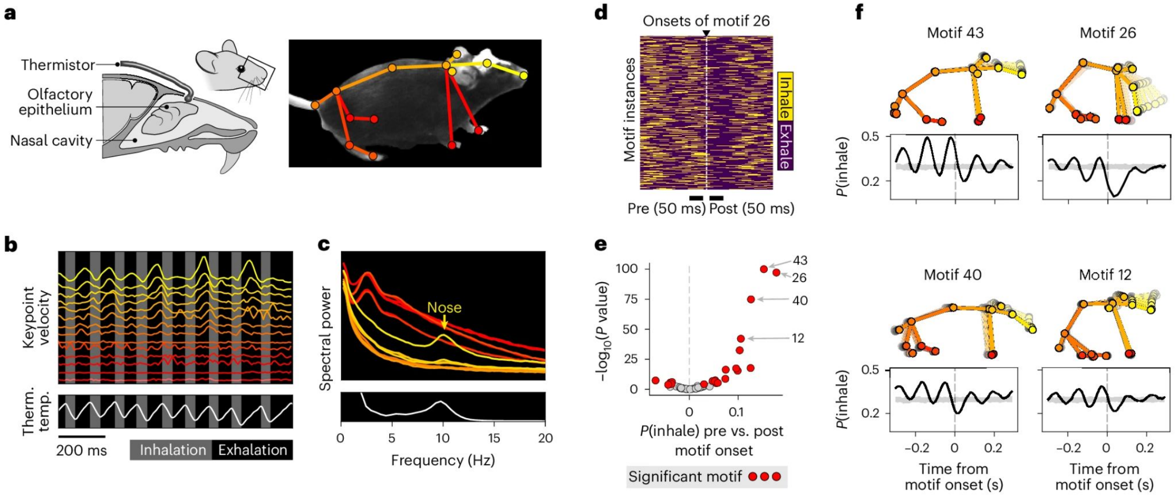
Figure 5: Keypoint-MoSeq generalizes across experimental setups



What difference in tracking appears when switching from 3D vs 2D data with Keypoint-MoSeq ?

There is no difference in the timing of the transition but more fine grained distinctions in behavior (higher number of syllables per behavior like turning of rearing).

Figure 6 : Keypoint-MoSeq segments behavior at multiple timescales



Why is Keypoint-MoSeq now picking up on such fast behavior changes ?

Median duration of syllables is decided by the stickiness hyperparameter.

How does respiration correlates with other behaviors ?

Nose velocity is highly aligned with respiration, and some motifs are respiration coupled (they aligned with transition in respiration state).

Paper round-up

- Keypoint jitter can be disentangled from behavior with a switching linear dynamical system
- This keypoint-MoSeq model captures the temporal structure of behavior
- Syllable transitions align with striatal dopamine fluctuations
- Keypoint-MoSeq is generalizable to setups, different behaviors, different species and different timescales

What did we learn? What questions do we have?

- **What points do they make in the discussion?**

- **Is anything unclear?**

- **What would you do next?**

- *Generalize to multi-animals tracking to be able to better characterize and track different social behaviors: https://keypoint-moseq.readthedocs.io/en/latest/FAQs.html#multiple-animals.*
- *Get more insights in neural activity during spontaneous behavior and how it aligns to syllable transition and for which timescales => Neuropixels implants or fiber photometry of other neuromodulators*
- *Investigate the hierarchical nature of behavior: "simultaneously analyze behavior across multiple timescales or explicitly represent the hierarchical nesting of behavior motifs": hBehaveMAE (https://www.biorxiv.org/content/10.1101/2024.08.06.606796v1)*
- *Unsupervised labelling of behavioral syllables with LLMs?*