Speech and Multi-Sensory Data Modeling for Child and Youth Psychiatry
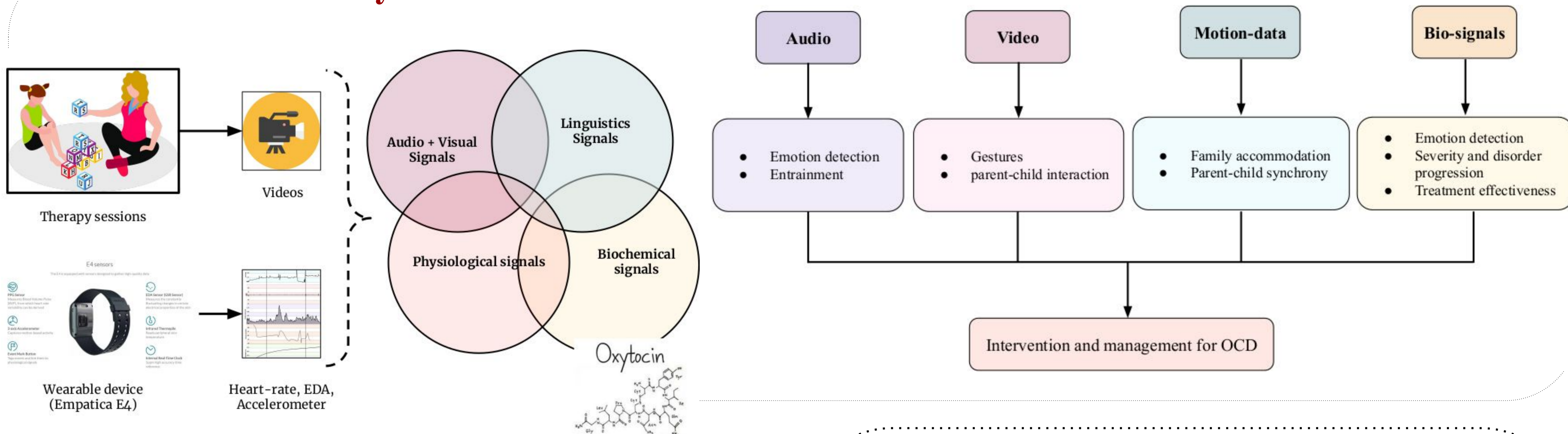
Sneha Das[1] (sned@dtu.dk), Nicole Nadine Lønfeldt[2], Nicklas Leander Lund[1], Anne Katrine Pagsberg[2, 3], Line H. Clemmensen[1]

Multi-sensory Data

Challenges & Motivation

- **Low Resource, In-the-wild, Evolving data**
- Speech Emotion Recognition (SER): Infer emotional state of individuals from speech signals.
- **Trannsferable SER**: *Continuous Metric Learning For Transferable Speech Emotion Recognition and Embedding Across Low-resource Languages*

Objective and Proposal (SER)

- Goal: Obtain emotion representations from speech that are transferable to low-resource (data and labels) languages.
- Proposal: Semi-supervised DAE→to shape the latent space with emotion-relevant information.
- Contributions:
  - Method for <u>continuous metric learning</u> to order samples in latent space.
- Semi-supervision using activation and valence labels.

Conclusions

- Proposed semi supervision method yields more consistent latent space.
- Consistent latent space →transferable to unseen (low-resource) languages.
- Long way to go!

1. Department of Applied Mathematics and Computer Science, Technical University of Denmark
2. Child and Adolescent Mental Health Center, Copenhagen University Hospital, Capital Region, Copenhagen
3. Faculty of Health, Department of Clinical Medicine, Copenhagen University