

Problem Statement

- Cat vocal communication is poorly understood
- Existing datasets are small and mostly lab-based
- It is unclear whether cats change vocal behavior based on context



DATASET

- ~300k vocalization events
- ~100 hours of audio
- Collected from YouTube and Reddit
- Thousands of cats across real-world environments
- High diversity enables better generalization

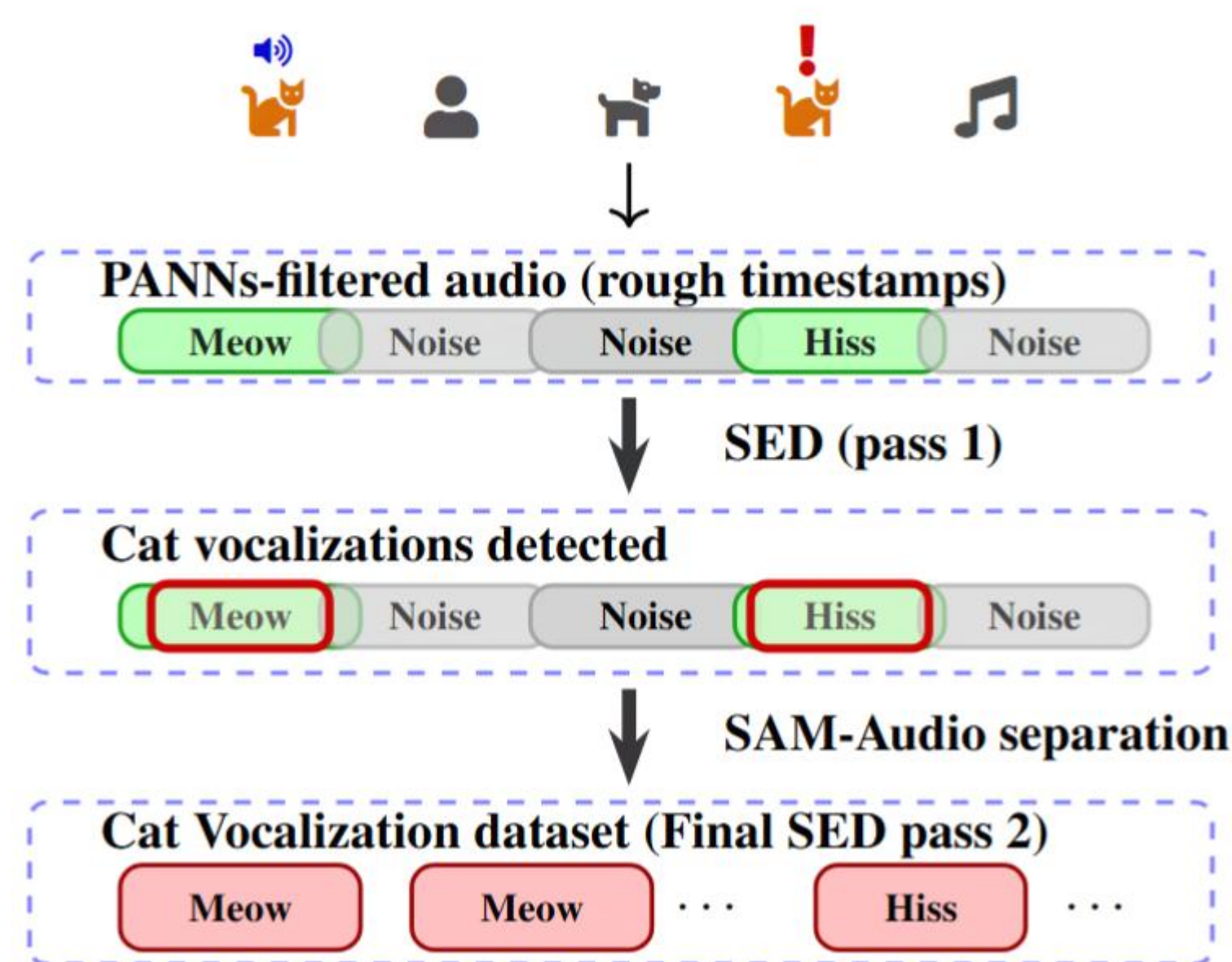
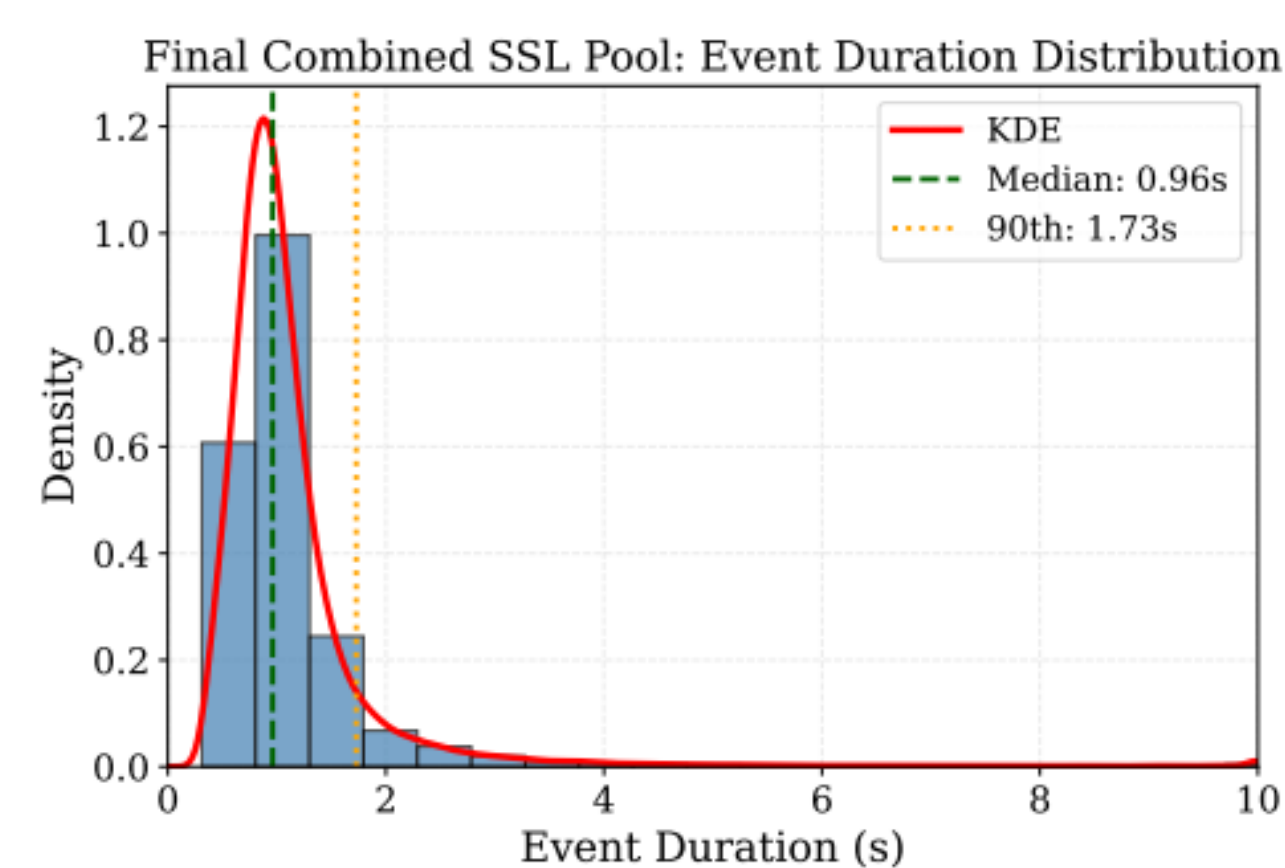


Fig 1. Overview of cat vocalization extraction pipeline



Method

- Detect cat vocalizations from videos
- Encode audio using self-supervised models
- Extract visual context using vision-language models
- Align each vocalization with interaction and activity labels

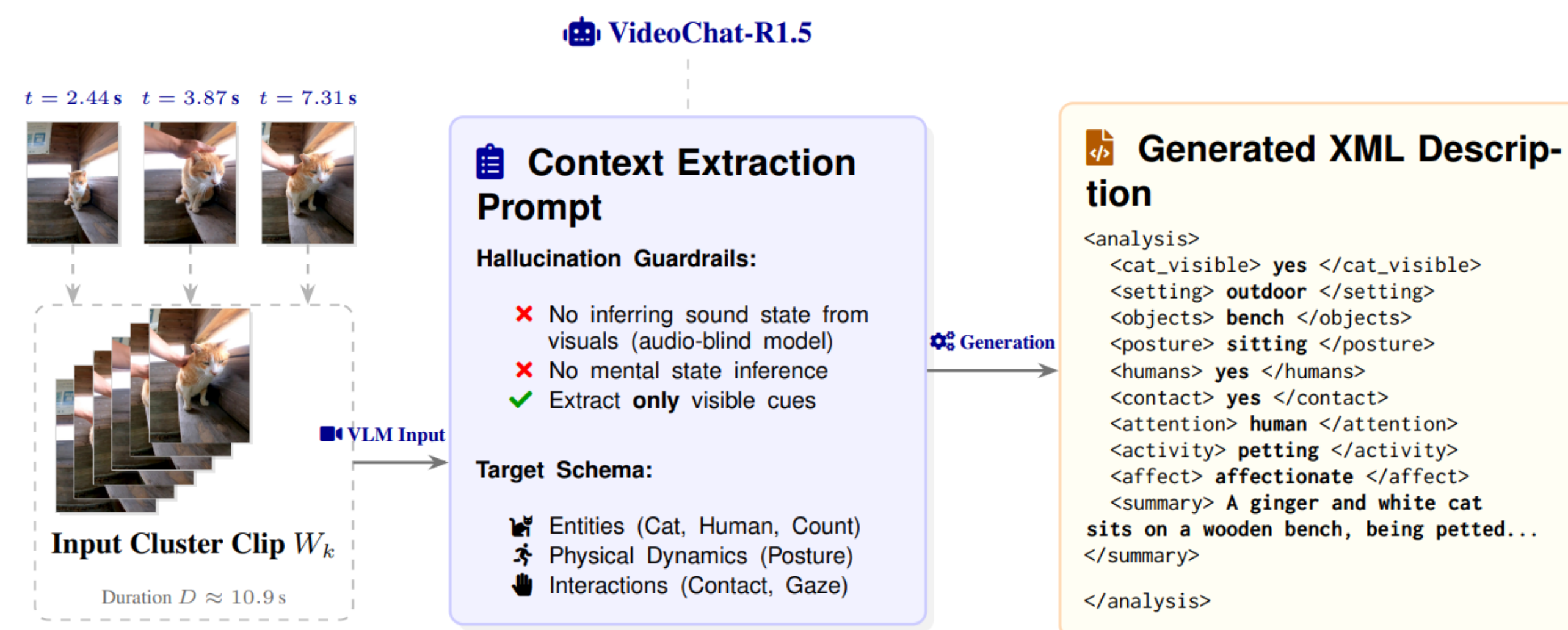


Fig 2. Overview of cat context extraction pipeline

KEY IDEA

- Analyze call types across interaction contexts
- Interaction modes:
 - Cat–Alone
 - Cat–Cat
 - Cat–Human
 - Multi-Cat + Human
- Evaluate distribution shifts in vocal behavior

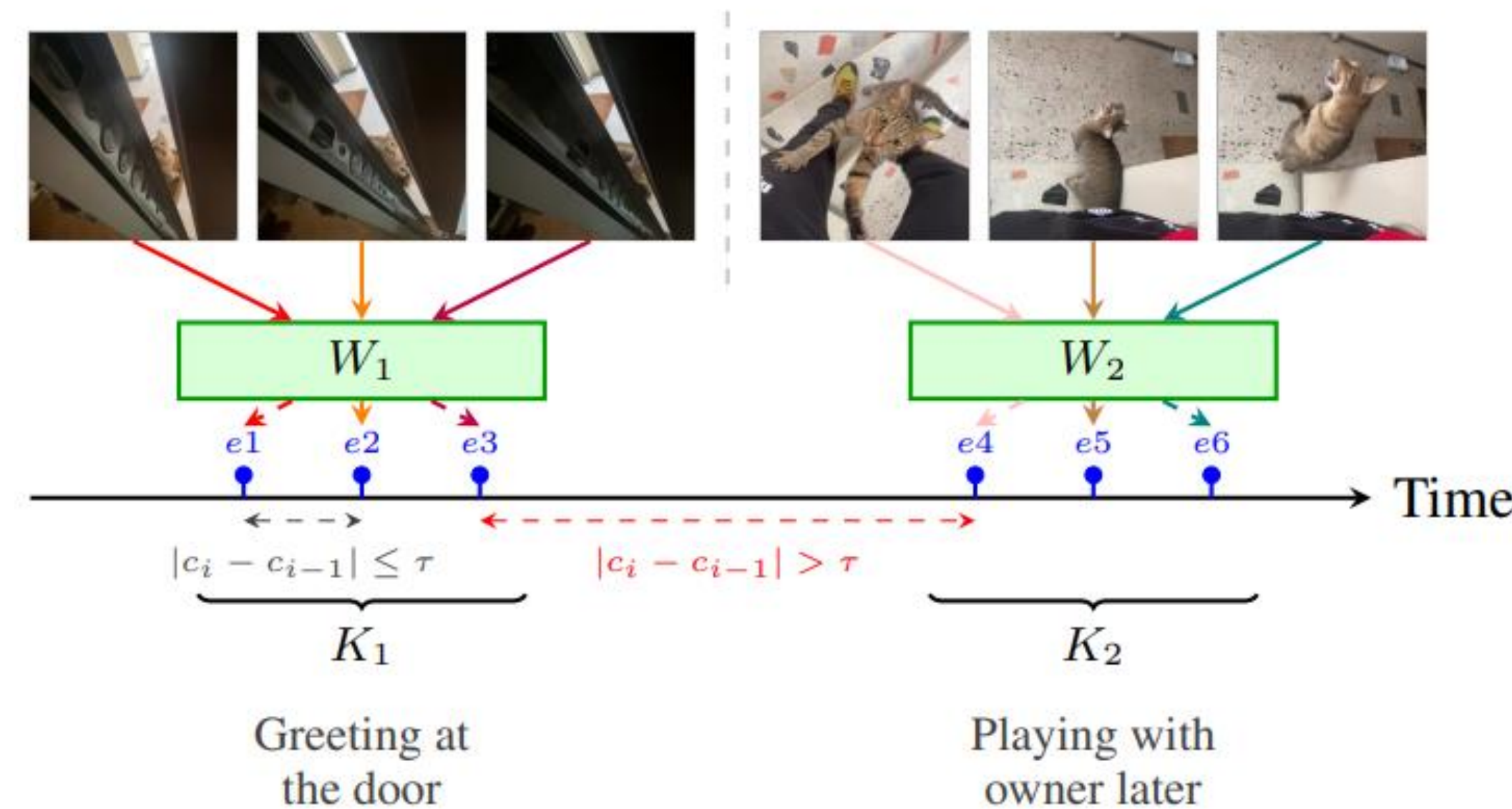


Fig 3. Showing cat event clustering algorithm

Experiment

Key Findings

- Cats are ~2× more likely to meow with humans than when alone
- Cats are ~1.6–1.7× more likely to produce snarl/pain cries when alone
- Cats are ~1.3–1.4× more likely to yowl in multi-cat settings
- Cats are ~2.2–2.8× more likely to purr in one-on-one human interactions



Interpretation

- Meows are strongly human-directed signals (supported by existing literature)
- Distress vocalizations occur more in solitary contexts
- Group settings increase arousal-related calls (yowling)
- Purring is primarily a dyadic (one-on-one) interaction signal

Conclusions

- Cat vocalizations vary systematically with interaction context
- Different call types are associated with specific social situations
- Meows are primarily human-directed, while distress calls occur more in isolation
- Large-scale multimodal data enables studying animal communication in real-world settings

References

1. Chen et al., 2022 — BEATS: Audio Pre-training with Acoustic Tokenizers
2. Hsu et al., 2021 — HuBERT: Self-Supervised Speech Representation Learning
3. Baevski et al., 2020 — wav2vec 2.0
4. Kong et al., 2020 — PANNs: Large-Scale Audio Neural Networks
5. Pandeya et al., 2018 — Domestic Cat Sound Classification