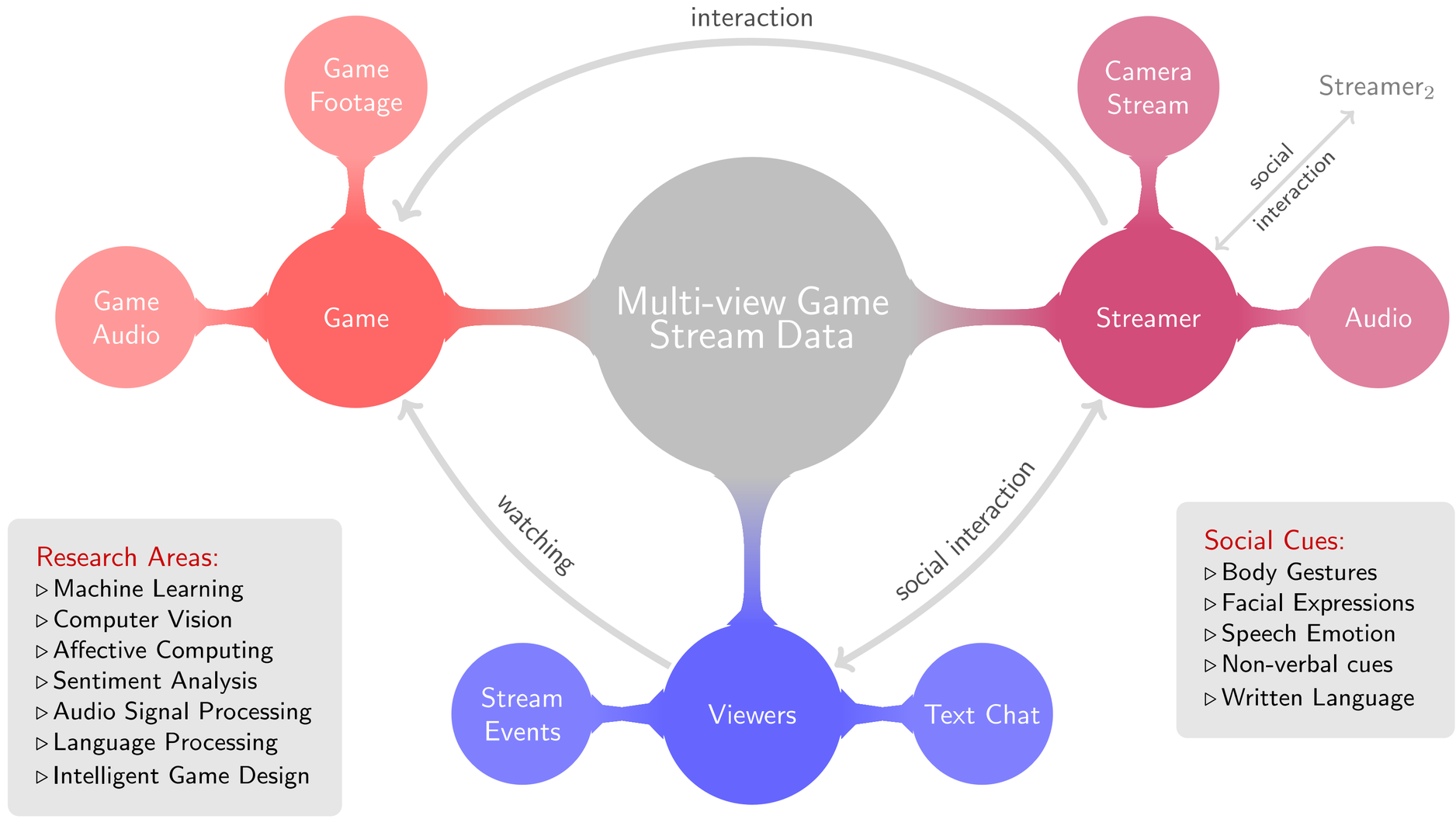
Streaming Behavior: Livestreaming as a Paradigm for Analysis of Emotional and Social Signals

Charles Ringer and Mihalis A. Nicolaou
Goldsmiths, University of London, and The Cyprus Institute

1. MOTIVATION



Research Areas:
- ▷ Machine Learning
- ▷ Computer Vision
- ▷ Affective Computing
- ▷ Sentiment Analysis
- ▷ Audio Signal Processing
- ▷ Language Processing
- ▷ Intelligent Game Design

Social Cues:
- ▷ Body Gestures
- ▷ Facial Expressions
- ▷ Speech Emotion
- ▷ Non-verbal cues
- ▷ Written Language

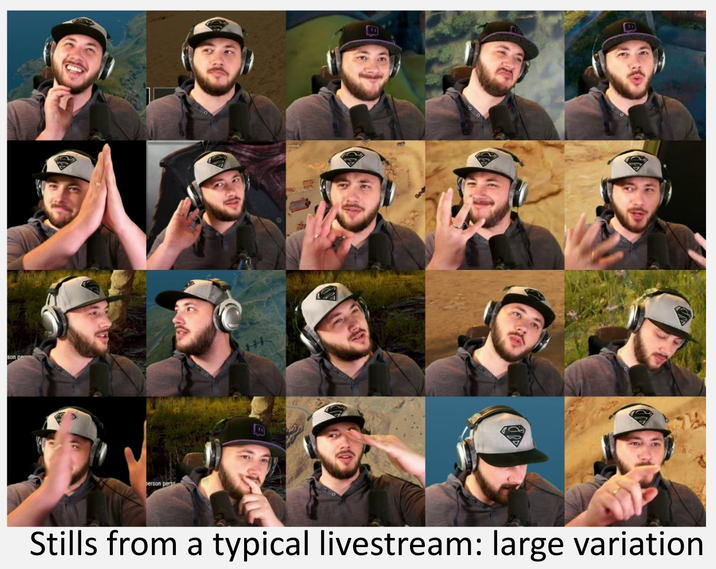
2. NOVELTY AS A PROXY FOR HIGHLIGHT DETECTION

- Consider reconstruction error as a proxy indicating interesting segments in a livestream.
- Train a deep network with convolutional and recurrent layers to do so (unsupervised).

Modalities	No. Videos	Highlight				No Highlight
		Funny %	Action%	Interaction%	Total%	Total%
Face, Game, Audio	98	0.27	0.26	0.24	0.77	0.23
Face, Audio	95	0.22	0.23	0.28	0.74	0.26
Face Only	96	0.14	0.14	0.24	0.52	0.48
Game Only	94	0.04	0.18	0.07	0.29	0.70
Audio Only	126	0.08	0.29	0.18	0.56	0.44

- Including the streamer face and audio significantly decreases the error rate.

Sequences from detected episodes.



Stills from a typical livestream: large variation in pose, expression, illumination, occlusions.

3. JOINT EMOTION AND GAME CONTEXT RECOGNITION

- Dataset (faces, game footage, audio, annotations) from League of Legends livestreams (>10h, 7K videos)
- Evaluate fusion techniques in a deep learning scenario.

F1 Scores for Each Label Across All Models. (V=Valence, A= Arousal)

Model	Task	Neg V	Neut V	Pos V	Neut A	Pos A	In Lane	Shopping	Returning	Roaming	Fighting	Pushing	Defending	Dead
Early Fusion	Joint	0.194	0.911	**0.362**	0.969	0.509	0.778	0.797	0.496	0.667	0.515	0.568	0.544	0.899
	Single	0.206	0.905	0.345	0.966	**0.540**	**0.842**	0.724	0.591	0.794	0.565	0.610	**0.667**	0.924
Late Fusion	Joint	**0.286**	**0.925**	0.297	0.964	0.465	**0.840**	**0.828**	**0.615**	**0.805**	**0.635**	**0.652**	0.557	0.906
	Single	0.088	0.918	0.340	0.968	0.491	0.791	0.776	0.513	0.774	0.581	0.582	0.452	0.937

Distribution of classes.

Valence			Arousal		Game Context								
Neg	Neut	Pos	Neut	Pos	In Lane	Shopping	Ret. to Lane	Roaming	Fighting	Pushing	Defending	Dead	Misc.
246	6,227	727	6,755	445	2,418	294	591	1,422	892	213	233	831	308

PAPER

https://github.com/charlieringer/LoLEmoGameRecognition

4. CONCLUSIONS

There is a great underlying potential in utilizing the vast amounts of audio-visual data broadcasted during a typical game stream. As we show, this setting is enriched with emotional and social signals, thus providing a vast resource of multimodal data. The interaction between different groups (streamers, viewers) can also facilitate the study of complex behaviours such as mimicry and conflict. Nevertheless, the analysis of livestreams presents several challenges to researchers working in areas such as affective computing, machine learning, audio signal processing and NLP.

Thanks to Mats Kathage (P4wnyhof) for permission to present images from his stream.

EPSRC Engineering and Physical Sciences Research Council

iGGi

Goldsmiths University of London

University of York

The Cyprus Institute Research·Technology·Innovation