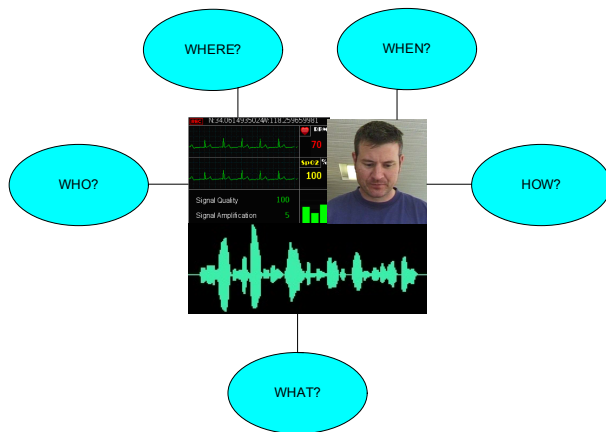# Representation, classification and information fusion for robust and efficient multimodal human state recognition

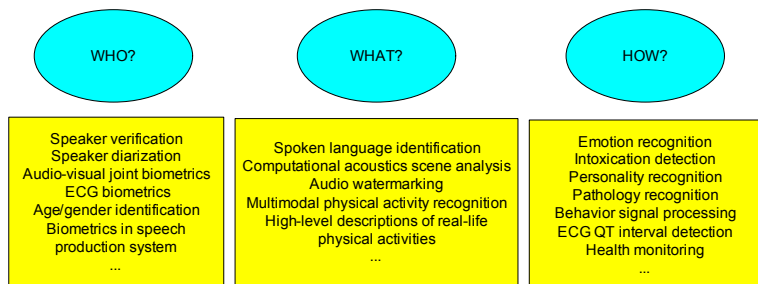Ming Li[1]

Signal Analysis and Interpretation Lab

2013 Research Festival

Identifying or verifying various human states from human centered multimodal signals
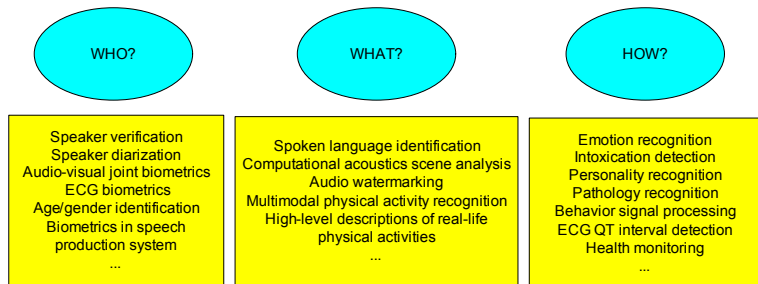

Each signal source (e.g. speech) offers a part of the story about each of the "state" of interest

Identifying or verifying various human states from human centered multimodal signals

WHO?

Speaker verification
Speaker diarization
Audio-visual joint biometrics
ECG biometrics
Age/gender identification
Biometrics in speech
production system
...

WHAT?

Spoken language identification
Computational acoustics scene analysis
Audio watermarking
Multimodal physical activity recognition
High-level descriptions of real-life
physical activities
...

HOW?

Emotion recognition
Intoxication detection
Personality recognition
Pathology recognition
Behavior signal processing
ECG QT interval detection
Health monitoring
...

A variety of technology applications in the state of the art

each with different levels of process

Identifying or verifying various human states from human centered multimodal signals

WHO?

Speaker verification
Speaker diarization
Audio-visual joint biometrics
ECG biometrics
Age/gender identification
Biometrics in speech
production system
...

WHAT?

Spoken language identification
Computational acoustics scene analysis
Audio watermarking
Multimodal physical activity recognition
High-level descriptions of real-life
physical activities
...

HOW?

Emotion recognition
Intoxication detection
Personality recognition
Pathology recognition
Behavior signal processing
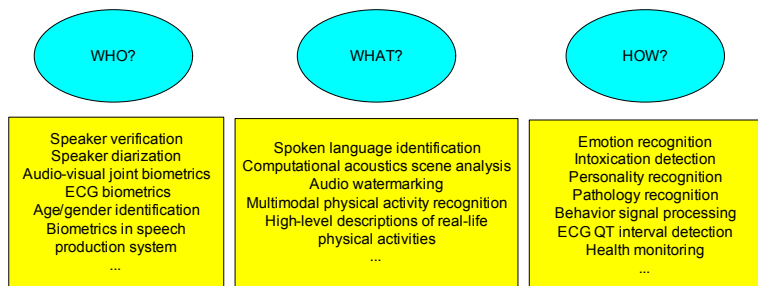ECG QT interval detection
Health monitoring
...

Human "state" recognition: a broad range of associated information:

biometric identity
language identity
physical activity
health state
emotional state
cognitive functioning

...

Identifying or verifying various human states from human centered multimodal signals

| WHO?	WHAT?	HOW?
Speaker verification Speaker diarization Audio-visual joint biometrics ECG biometrics Age/gender identification Biometrics in speech production system ...	Spoken language identification Computational acoustics scene analysis Audio watermarking Multimodal physical activity recognition High-level descriptions of real-life physical activities ...	Emotion recognition Intoxication detection Personality recognition Pathology recognition Behavior signal processing ECG QT interval detection Health monitoring ...

Characteristics

▶ states are relatively stable

▶ fit a general supervised learning classification framework

USC Viterbi
School of Engineering

Representation, classification and information fusion
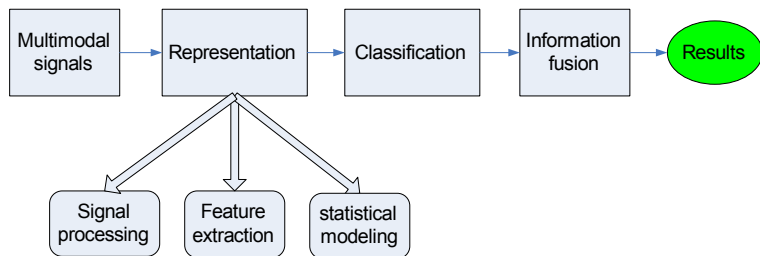
- ▶ time varying property $\Longrightarrow$ short time frame level features
- ▶ generative model for data description $\Longrightarrow$ features (supervectors) in model parameters' space for classification

Representation, classification and information fusion
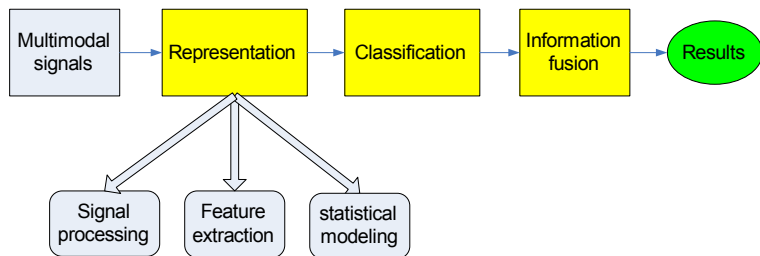
> ## The focus of this proposal:
>
> Robustness: improve the performance
> Efficiency: reduce computational cost
> on top of the state of the art systems
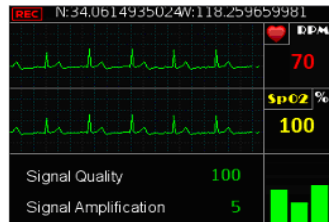
Working on top of the state of the art systems
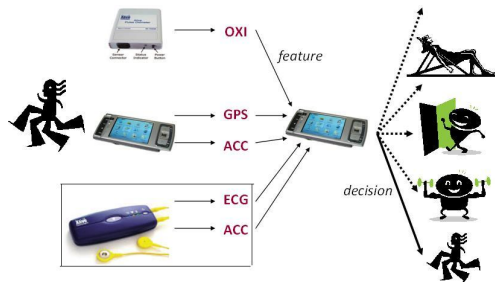
- A very active and competitive research area
  - NIST Speaker Recognition Evaluation (SRE) 1997-2012 (13)
  - NIST Language Recognition Evaluation (LRE) 1996-2011 (5)
  - DARPA RATS LRE evaluation 2011-2014
  - Interspeech challenges 2009-2012
    (emotion, paralinguistic, intoxication, personality, pathology) (5)

- Standard databases and tasks, Engineering skills required

- Multimodal signals

Working on top of the state of the art systems

- A very active and competitive research area
  - NIST Speaker Recognition Evaluation (SRE) 1997-2012 (13)
  - NIST Language Recognition Evaluation (LRE) 1996-2011 (5)
  - DARPA RATS LRE evaluation 2011-2014
  - Interspeech challenges 2009-2012
    (emotion, paralinguistic, intoxication, personality, pathology) (5)

- Standard databases and tasks, Engineering skills required

- Multimodal signals

Working on top of the state of the art systems

- A very active and competitive research area
    - NIST Speaker Recognition Evaluation (SRE) 1997-2012 (13)
    - NIST Language Recognition Evaluation (LRE) 1996-2011 (5)
    - DARPA RATS LRE evaluation 2011-2014
    - Interspeech challenges 2009-2012
      (emotion, paralinguistic, intoxication, personality, pathology) (5)

- Standard databases and tasks, Engineering skills required
- Multimodal signals

Working on top of the state of the art systems

- A very active and competitive research area
  - NIST Speaker Recognition Evaluation (SRE) 1997-2012 (13)
  - NIST Language Recognition Evaluation (LRE) 1996-2011 (5)
  - DARPA RATS LRE evaluation 2011-2014
  - Interspeech challenges 2009-2012
    (emotion, paralinguistic, intoxication, personality, pathology) (5)

- Standard databases and tasks, Engineering skills required

- Multimodal signals

Working on top of the state of the art systems

- A very active and competitive research area
    - NIST Speaker Recognition Evaluation (SRE) 1997-2012 (13)
    - NIST Language Recognition Evaluation (LRE) 1996-2011 (5)
    - DARPA RATS LRE evaluation 2011-2014
    - Interspeech challenges 2009-2012
      (emotion, paralinguistic, intoxication, personality, pathology) (5)

- Standard databases and tasks, Engineering skills required

- Multimodal signals

IEMOCAP database
collected by SAIL lab at USC

- Motion Capture Facial Information
- Speech
- Video
- Transcriptions
- http://sail.usc.edu/iemocap
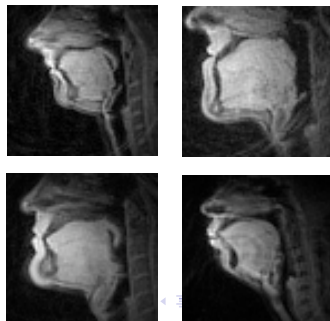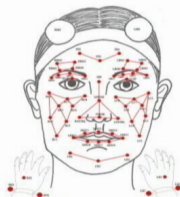
Thank you for your attention.

- ► We welcome your questions, suggestions and comments!
- ► Personal website (http://www-scf.usc.edu/ mingli/)
- ► SAIL lab website (http://sail.usc.edu/)