

Complementing Datasets for Recognition and Analysis of Affect in Speech

Dr. Tal Sobol Shikler

Formerly: Computer Laboratory, University of Cambridge, UK



Department of Industrial Engineering & Management

Ben-Gurion University of the Negev

Requirements

Recognise (and synthesise) affective states occurring in **real** **everyday** scenarios!

- A wide range of affective states
- Natural expressions
- Fully automated
- Flexibility and generality
 - Text and context independent
 - User independent
 - *Suitable for various applications*

Challenges

- General framework rather than specific to predefined emotions
- Handling many affective states that can occur **simultaneously**
- Handling affective states that **change dynamically** over time

Affective states in Time



Personality



Long-term mental state

Attitudes, moods & mental states

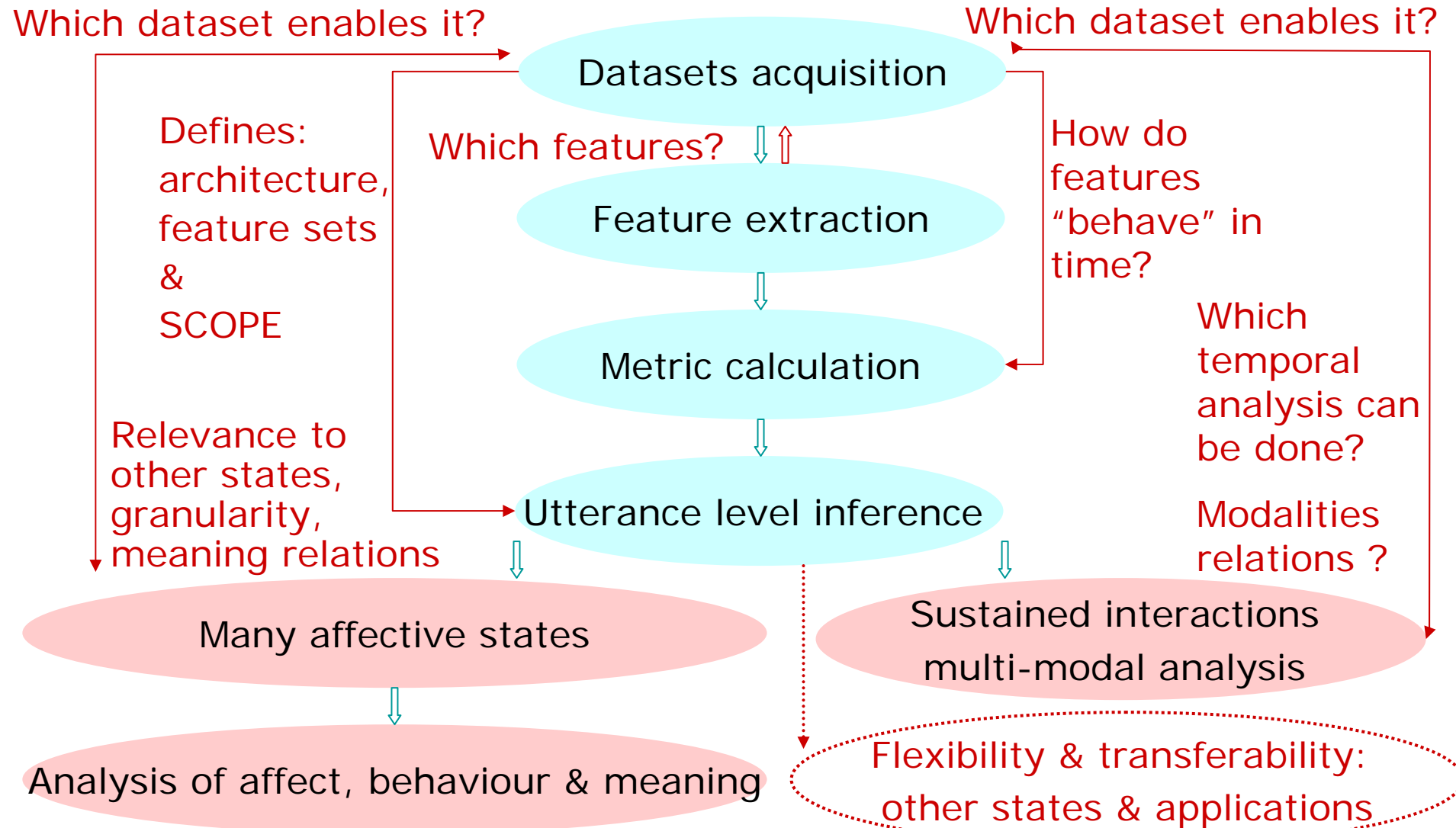


Cognitive & emotional responses to events,
(including the interaction itself)



Interaction

Datasets and the Framework



Datasets

Criteria:

- Scope
- Naturalness
- Context
- Descriptors

The databases set the limitations for the rest of the processing.

Doors

Mind Reading

Hebrew

English

Same text

Different texts

2 sentences, 100 repetitions per person
+unrestricted text, laughter,...

Total ~4400 sentences

15 speakers

10 speakers

children, young & old adults

Natural

Acted (induced)

Sustained HCI

Unrelated utterances

Multi-modal

Speech only

facial expressions, physiological cues,
game events, mouse movements

(video samples are unrelated to speech)

Unlabelled expressions

Labelled

412 affective states, in 24 groups according
to the Mind Reading taxonomy

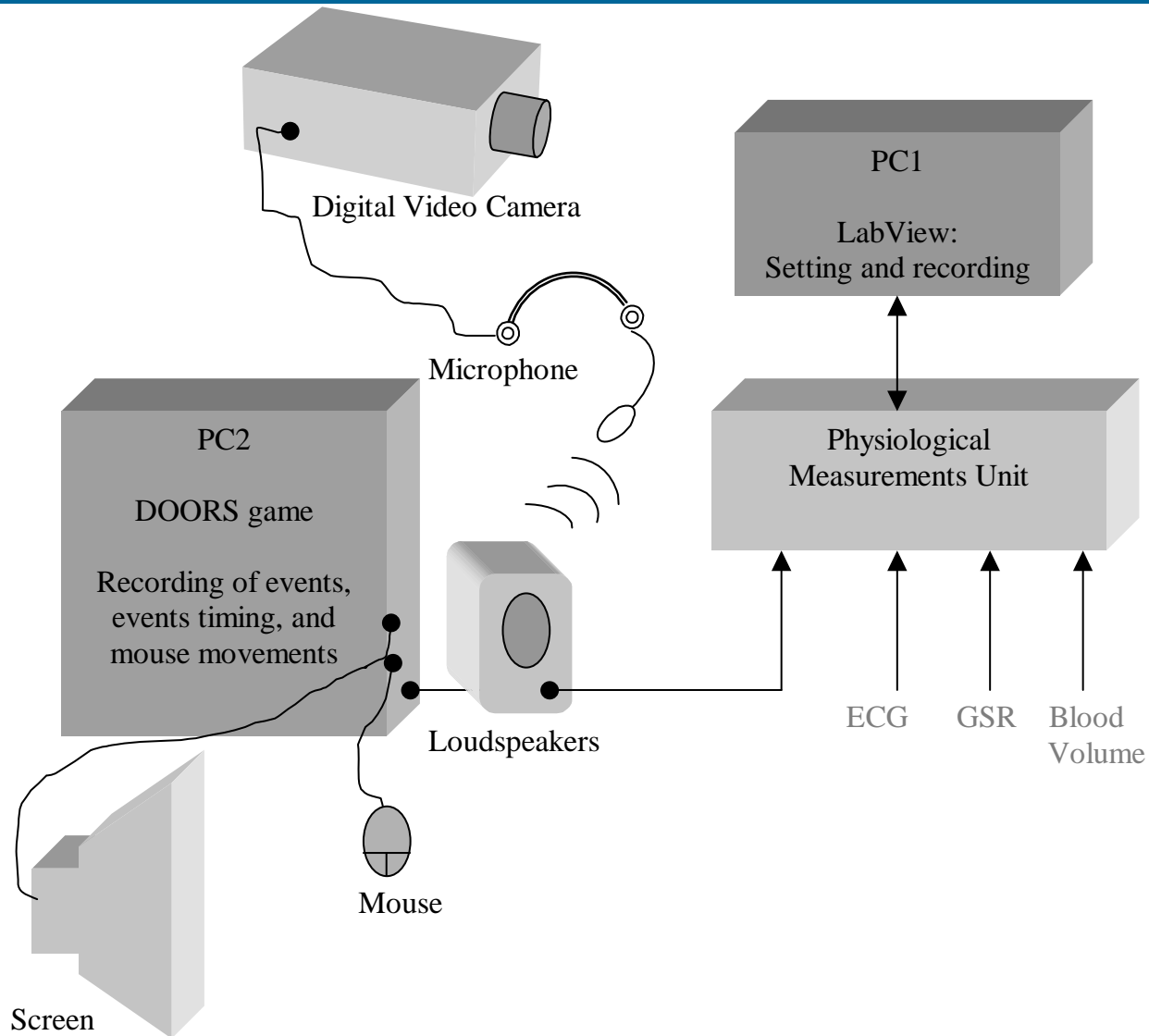
Segmentation to sentences based on
Shen *et al.* (*entropy*), zero crossing and adaptation

Created by Baron-Cohen *et al.*, Autism Research Centre,
University of Cambridge, 2004

Datasets in the Framework

Doors	Mind Reading
	Taxonomy
Feature definition & extraction	Feature definition & extraction
Observations: <ul style="list-style-type: none">• Different sets of features for different pairs of affective states• Thresholds	
	Classification: training and testing sets
	Comparison to human performance
Generalisation: Naturally evoked, subtle affective states & nuances, another language	Generalisation: A VERY wide variety of affective states (over 500 states, 4400 sentences)
Multi-modal analysis: Dynamics of sustained interactions	Analysis: Affective states, taxonomies, meaning, abstraction, word sense disambiguation

Doors



Doors

- Based on the Iowa Gambling Test (IGT)
(Bechara *et al.*, Science 1997)
- Subtle & naturally evoked affective states
- Dynamic changes and nuances
- Controlled environment & text
- Blue background - head detection
- Painted blue dots – detect and track facial features
- Synchronization:
 - Game event = beep
 - Speech recorded into the camera
 - Loudspeaker connected to LabView (physio)

Mind Reading

- A game for people on the Autism spectrum to recognise and understand expressions
 - Realistic expressions
 - A wide variety
- Underlying taxonomy
 - Based on the way people organise knowledge and meaning (the prototypical approach).
 - States that (may) have an emotional aspect.
- Induced emotions
- Neutral text
- Speakers of all age groups
 - Children, young and old adults.

Research Approaches

- **Approaches:** categorical (usually - basic emotions), dimensional (positive-negative, etc.), bi-polar (an emotion exists or not), appraisal model (w3c, inference?)
- **All the approaches are “correct”**
(Harnad, Russel), but partial
- **The prototypical approach**
 - Intelligible - based on the way people/humans organize knowledge
 - **Supports and includes the other approaches**
 - A wider range of affective states
 - An affective state can belong to several groups (depending on the taxonomy in use)

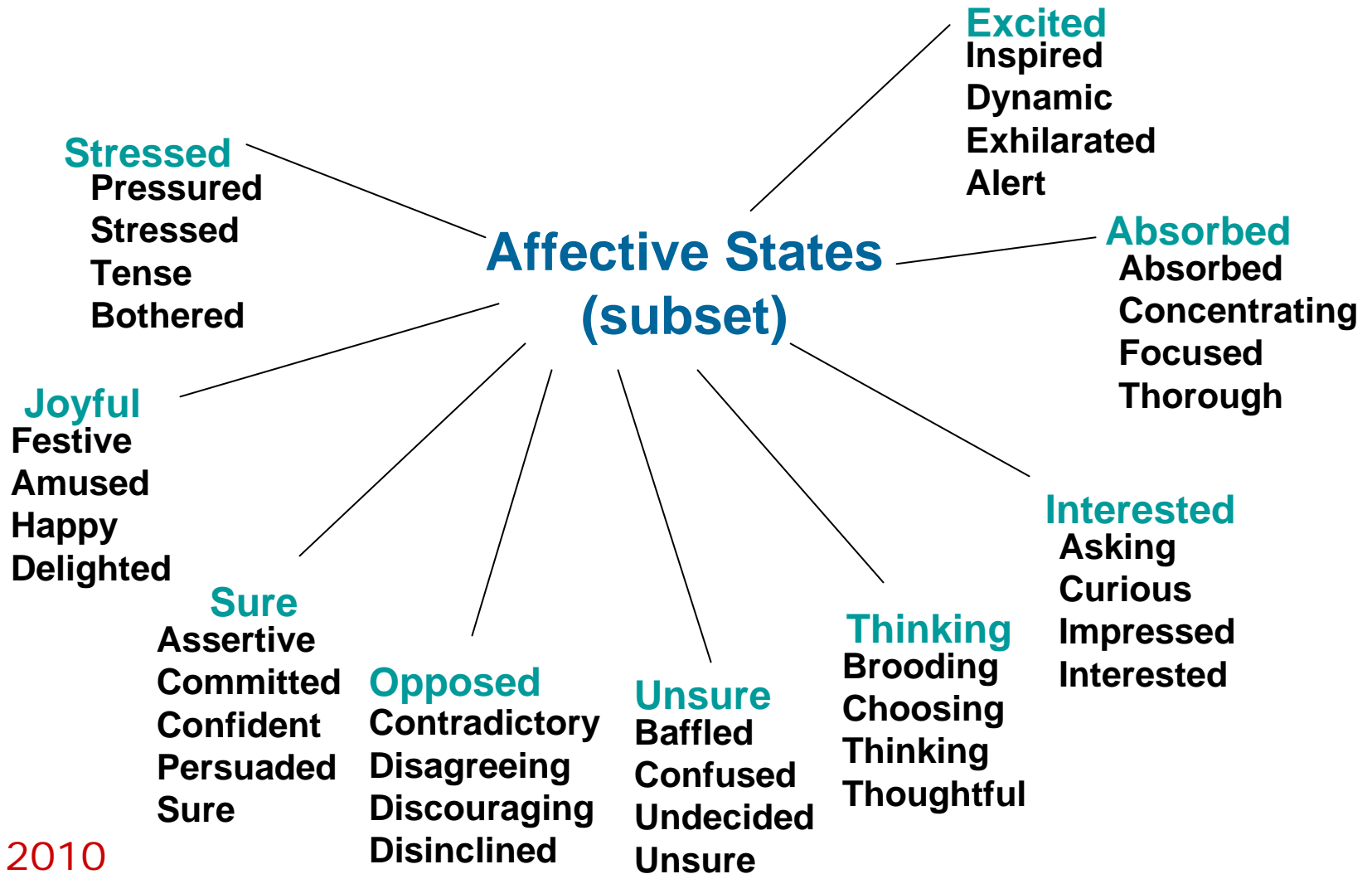
Mind Reading

The groups of the Mind Reading taxonomy:

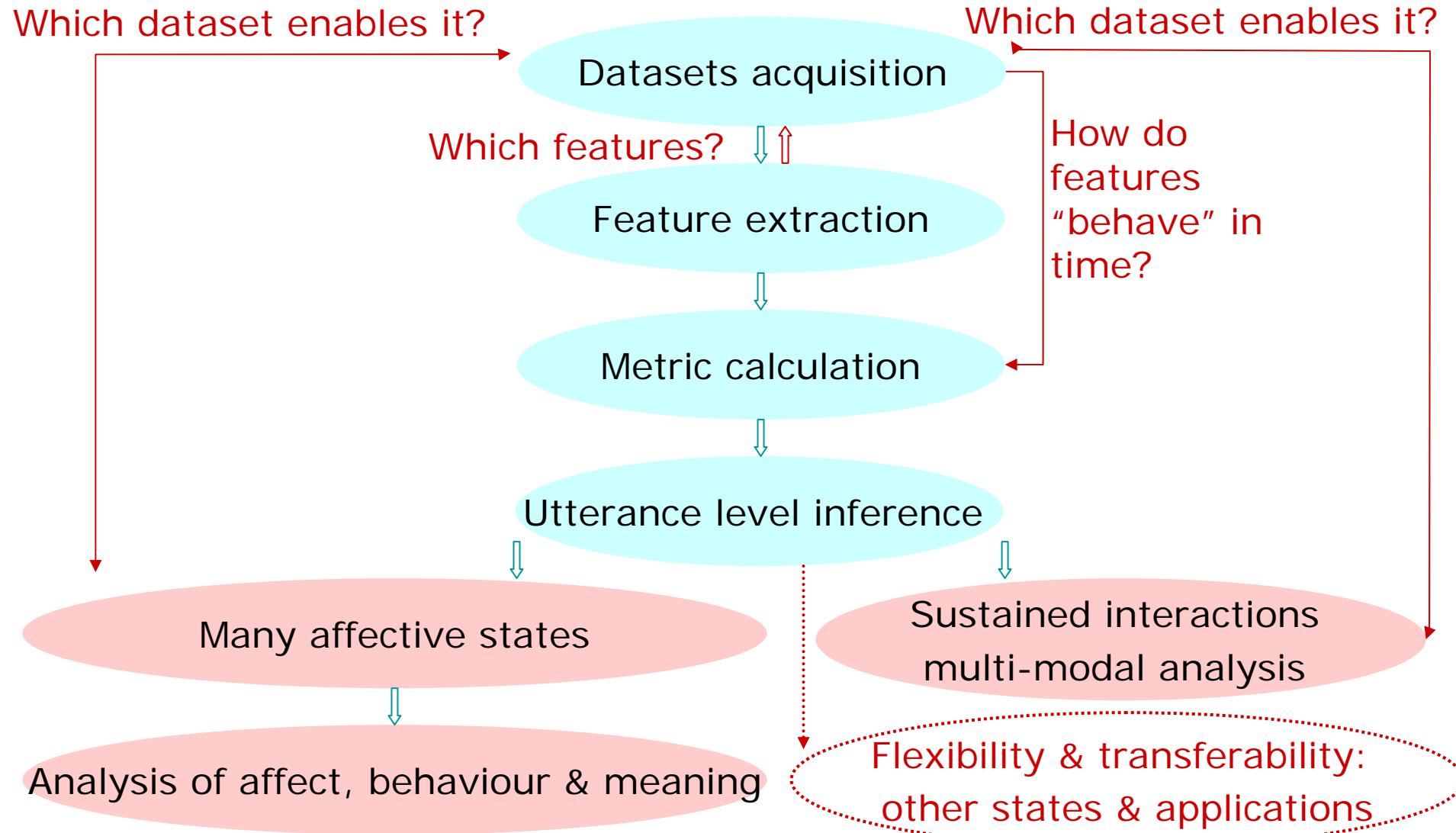
afraid	touched	bothered*	unfriendly*	thinking*
surprised	fond	hurt	sneaky	interested**
angry	liked	sorry	bored	excited*
sad	kind	disbelieving	wanting	sure*
happy*	romantic	unsure*		
disgusted				

Each group comprises many affective states that share a meaning.

Taxonomy



Framework - Features



Vocal Features

What are the features?

How to extract them?

Extraction algorithms often require manual “cleaning”



How to do it automatically?

Features in the psycho-acoustic literature are often qualitative.

Extraction algorithms require manual adjustments.

Features of voice quality are not well defined .

Defining expressive vocal features

Doors	Mind Reading
<ul style="list-style-type: none">• multiple text repetitions by each speaker• subtle affective states• natural transitions between affective states during sustained interactions	<ul style="list-style-type: none">• a VERY large variety of affective states• different speakers

Defined features & extraction algorithms

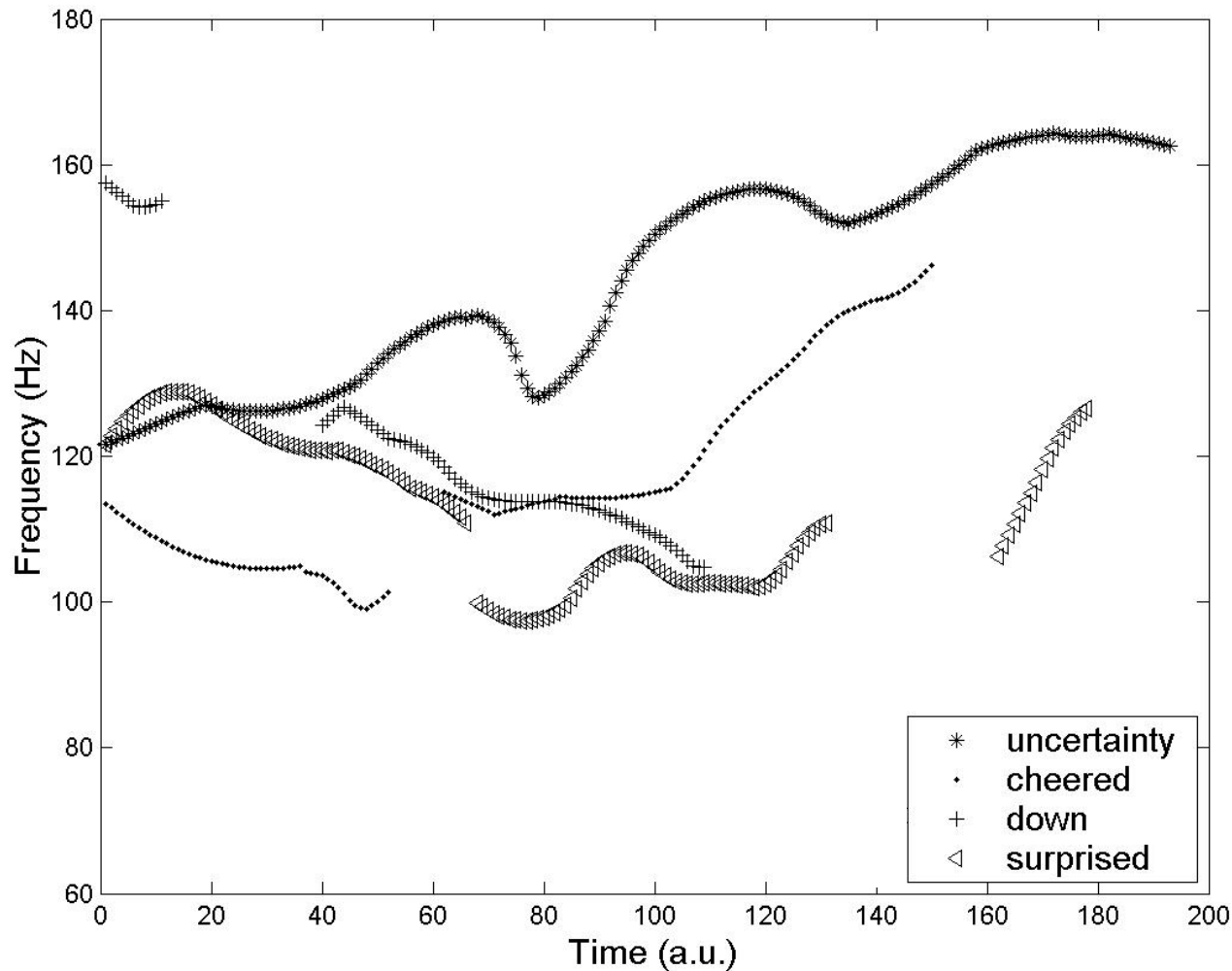
- **Fundamental frequency** (f_0 , pitch or intonation),
Based on Boersma's algorithm (PRAAT) **with adaptations**
- **Smoothed energy curve**
Average of the energy over a time frame combined with Hamming window
- **Spectral content**
A filter-bank up to 9 kHz (Bark scale based)
- **Harmonic properties**
Voice quality, consonance and dissonance.
Based on findings from physics, musicology and neuro-science.

All the vocal features were extracted **automatically**



Both datasets are “annotated” with features

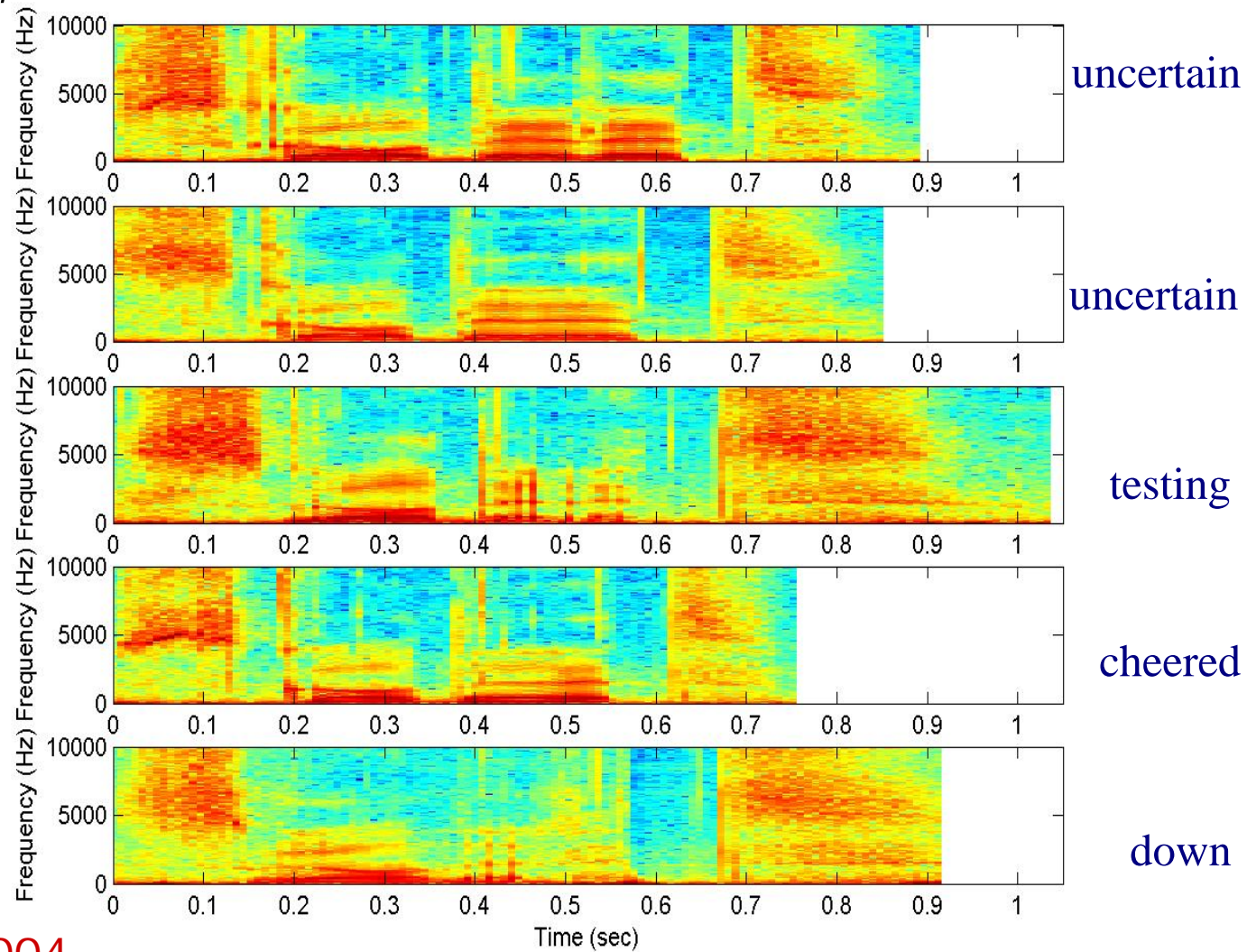
Features in time (f_0 , intonation)



The same speaker,
text & interaction
(Doors)

Features in time

The same speaker,
text & interaction
(Doors)



Metrics

- **Rule-based Parsing**

Divides the duration of the utterance

- **Temporal characteristics**

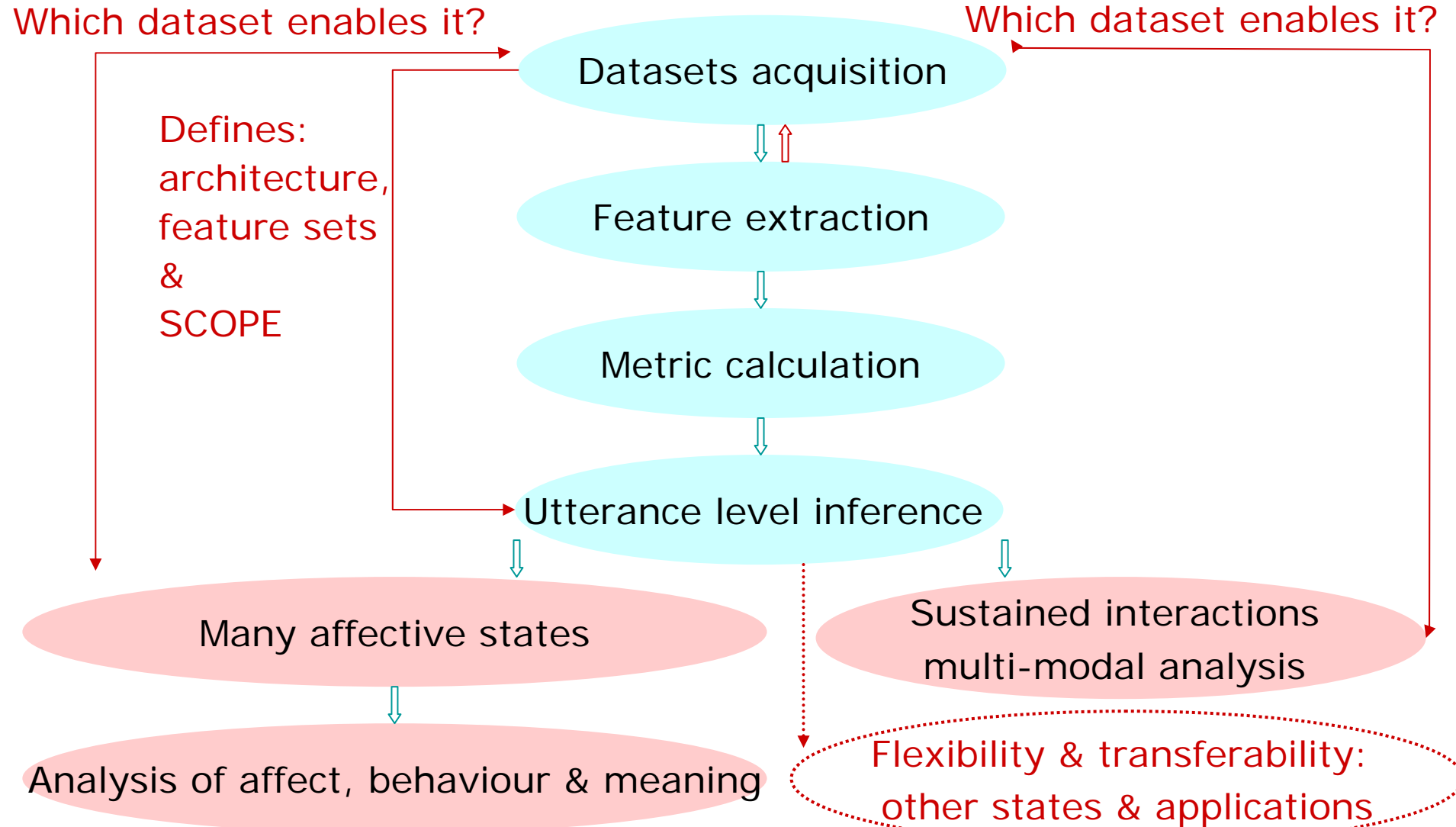
Draw on terms from disciplines such as linguistics and musicology

- Syllables, consonants and vowels
- Tempo and melody .

- **Statistical properties**

173 metrics for each speech signal
(sentence/utterance)

Framework - Inference



Defining expressive vocal features

Doors

Mind Reading

Complementing information for feature extraction, rule-based parsing, metric definition & normalization

Observations :

- Thresholds
- Different features <-> different pairs

The goals (levels of affect, co-occurring affects) and the observations defined the architecture

- Taxonomy
- Training and testing data for inference

Defined the scope

Normalisation

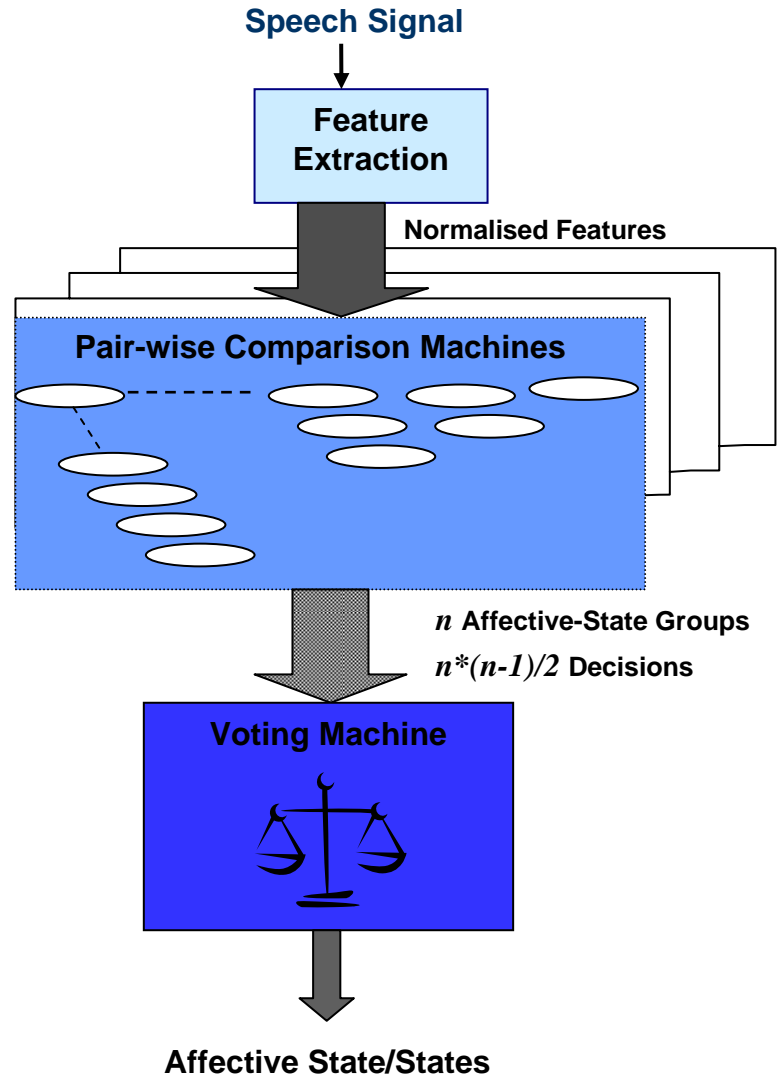
Normalise each feature by the speaker

- Common approach: normalise only by features
- Minimise the effect of speaker variability – **two languages and cultures**
- Possible generalisation to new speakers with no further training

Architecture

- Underlying taxonomy for all affective states
- Choice of a sub-set of affective states for inference (flexible)
- Classification of **co-occurring** affective- state groups:
 - **Pair-wise comparisons**
 - **Different metrics for different pairs of states**
=> metrics & algorithm optimization per pair
 - **Voting algorithm to consolidate comparisons into a single ranked list**
=> levels of recognition for all (9) inferred states (vs. exists or not)
=> possible: several dominant states

Architecture



Inference Results

- Inference of **a single** affective state:

Average accuracy of system over 70% (random probability 11%)

However

More than one affective state can exist and be chosen by 6-8 of the 8 pair-wise machines

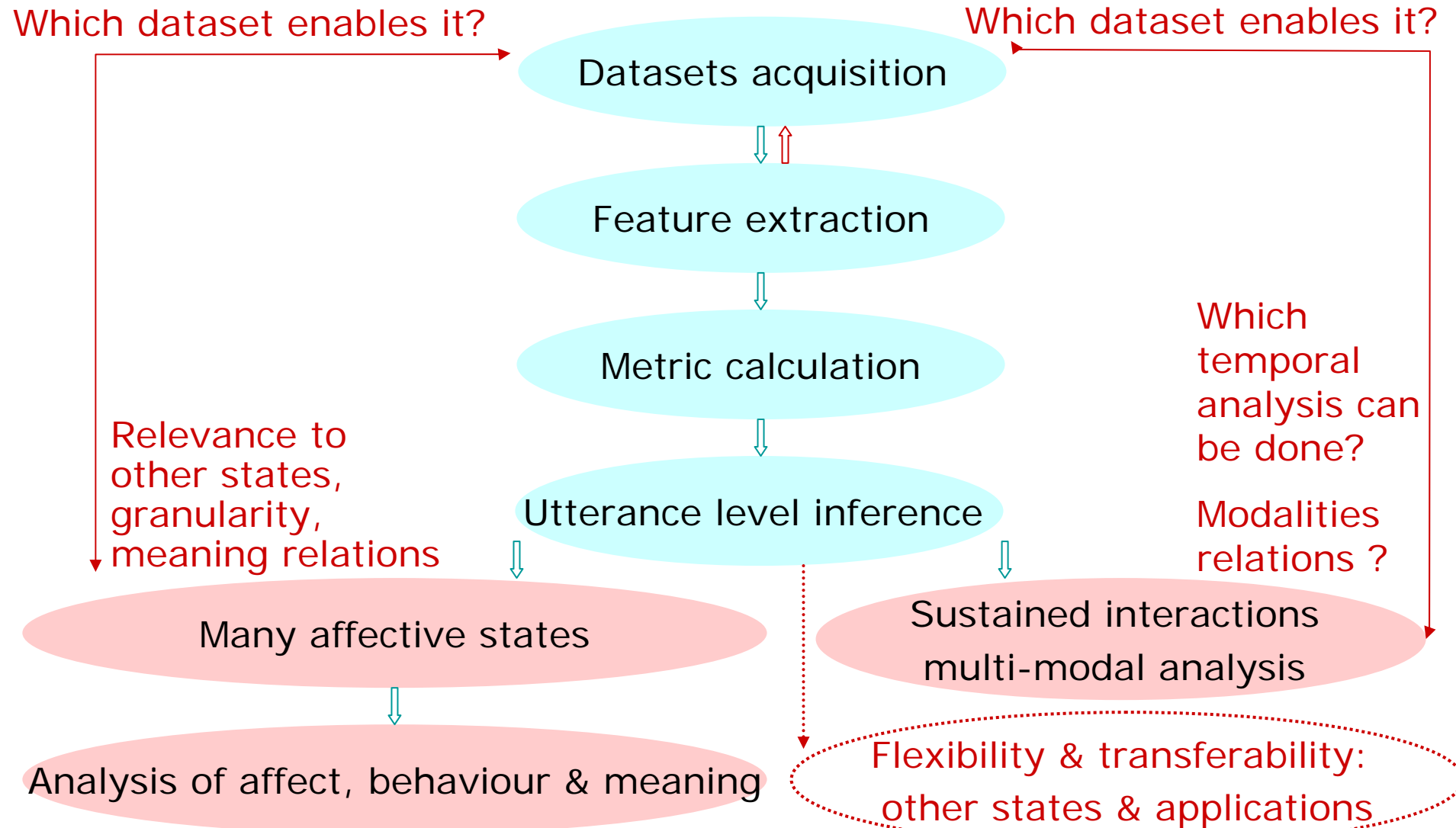


- Inference of **co-occurring** affective states:

Overall accuracy of the system **83%** true-positive recognition

- Recognition of each affective-state group >75% (random probability 14%)
- 60%-40% split between training and testing

Framework - Generalisation



Datasets in the Framework

Doors	Mind Reading
<p>Generalisation:</p> <ul style="list-style-type: none">• Naturally evoked, subtle affective states & nuances• Statistically significant correlation to events• Dynamic changes over sustained interactions – short/long terms• Human-computer & human-human• Controlled and unrestricted text• Multi-modal analysis – relations to events and to other cues – short and long term• Another language	Taxonomy
	<p>Evaluation & Generalisation:</p> <ul style="list-style-type: none">• Evaluation on the nine inferred affective-state groups• Comparison to human performance on an independent test (the CAM Battery Test)• Application to the entire Mind Reading database <p>A VERY wide variety of affective states (over 500 states, 4400 sentences)</p>
	<p>Analysis:</p> <p>Affective states, taxonomies, meaning, abstraction, word sense disambiguation</p>

Comparison to Human Performance

Cam Battery Test

[1] IEEE TPAMI, 2010

- 20 complex affective states
- 50 sentences
- Distinguish concept from 3 other concepts
 - AS group (21 people),
 - Control group (17 people),
 - Inference machine – same test (sentences and choice)



	AS group	Control group	Total	Inference machine
Mean no of sentences	35.7	42.8	38.9	(total) 49

“Mapping”

Define affective states and the relations between them (using the Mind Reading dataset)

[1] IEEE TPAMI, 2010

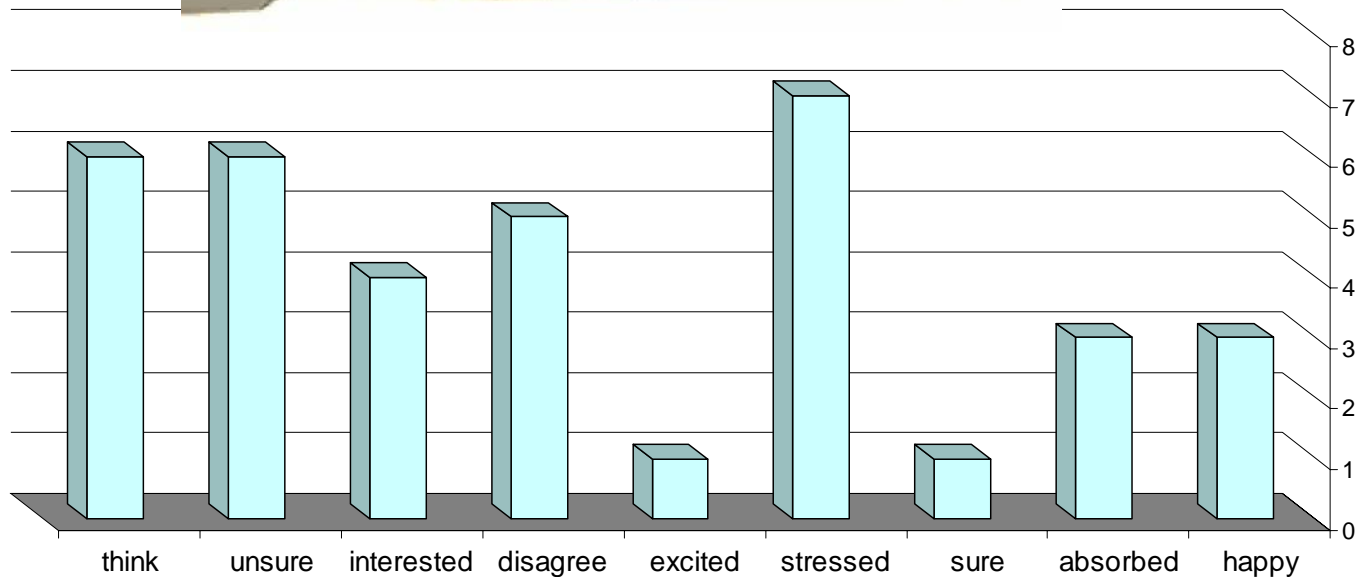
Example – the ‘Thinking’ group

joyful	absorbed	sure	stressed	excited	opposed	interested	unsure	think	
	x						x	x	Agonising, considering, brooding, dreamy, debating
			x				x	x	realising
	x							x	Calculating, fantasising
							x	x	Choosing
								x	Comprehending, deciding, regarding
			x				x		preoccupied

Sustained Interactions (Doors)

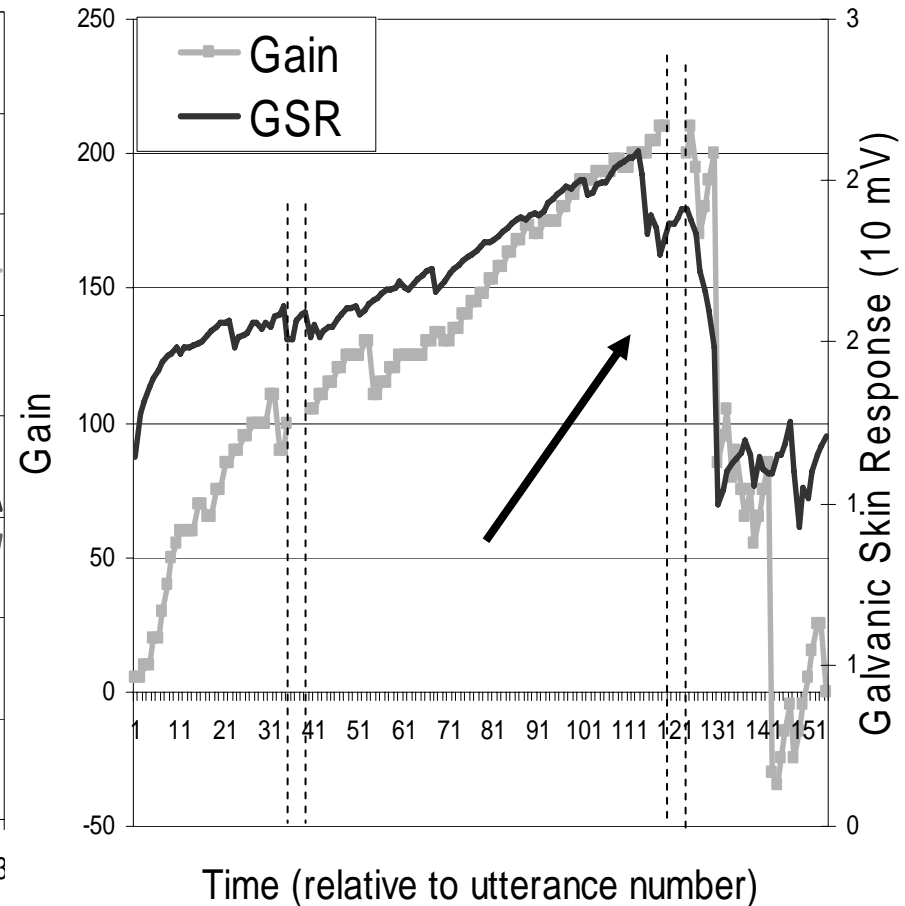
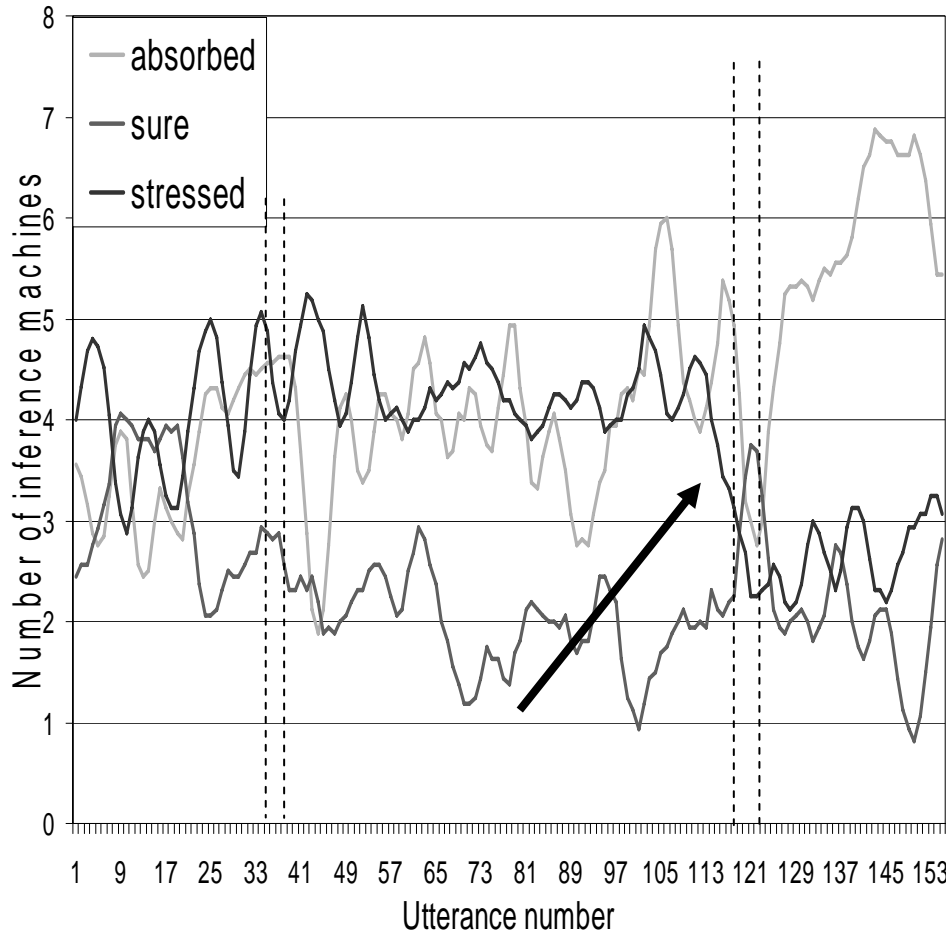


[4] IEEE SMC 2008



Sustained interactions

[4] IEEE SMC 2008

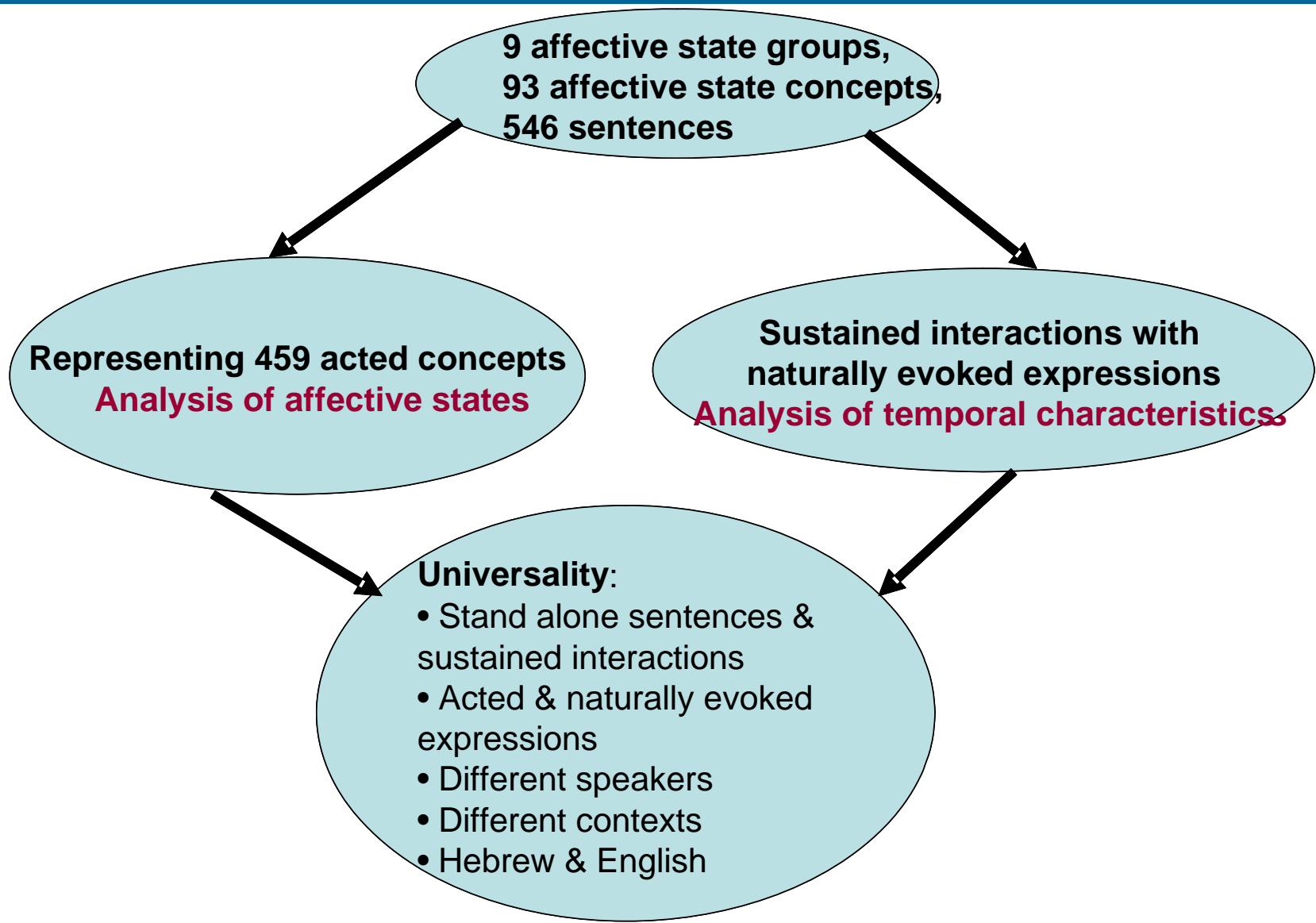


* GSR – Galvanic Skin Response (skin conductivity)

Analysis

1. Statistically significant correlation (2-level ANOVA) between events and the recognized levels of affective states.
2. Temporal changes during interactions and comparison to other behavioural and physiological cues:
 - Tendencies over time vs. other cues (long term)
 - Co-occurrences of special events, text, and changes in physiological cues with specific inferred affective states (short term)

Generalisation



Datasets in the Framework

Doors	Mind Reading
	Taxonomy
Feature definition & extraction	Feature definition & extraction
Observations: <ul style="list-style-type: none">• Different sets of features for different pairs of affective states• Thresholds	
	Classification: training and testing sets
	Comparison to human performance
Generalisation: Naturally evoked, subtle affective states & nuances, another language	Generalisation: A VERY wide variety of affective states (over 500 states, 4400 sentences)
Multi-modal analysis: Dynamics of sustained interactions	Analysis: Affective states, taxonomies, meaning, abstraction, word sense disambiguation

References (available online)

Full references to other works and more details on this research can be found in:

1. T. Sobol-Shikler and P. Robinson, '**Classification of complex information: Inference of co-occurring affective states from their expressions in speech**', in press in *IEEE TPAMI*, available online (DOI:10.1109/TPAMI.2009.107)
2. T. Sobol-Shikler, '**Automatic Inference of Complex Affective States**', in press in *Computer, Speech and Language*, available online (DOI:10.1016/j.csl.2009.12.005)
3. T. Sobol-Shikler, '**Complementing Datasets for Recognition and Analysis of Affect in Speech**', *proc. of the 2nd Emotions Workshop, LREC2010*, 2010, Valletta, Malta.
4. T. Sobol-Shikler, '**Multi-modal analysis of human computer interaction using automatic inference of aural expressions in speech**', *proc. of IEEE SMC*, 2008, Singapore.
5. T. Sobol-Shikler, P. Robinson, '**Visualizing Dynamic Features of Expressions in Speech**', *proc. of ICSLP (INTERSPEECH) 2004*, Jeju, Korea.
6. T. Sobol-Shikler, '**Analysis of affective expressions in speech**', *Tech report*, University of Cambridge, 2009.

Links can be found in: <http://www.bgu.ac.il/~stal>

Acknowledgement

I thank the AAUW- Educational Foundation, Cambridge Overseas Trust and Deutsche Telekom Laboratories at Ben-Gurion University for their partial support of this research.

I thank all the people who contributed to this research and to the recordings of the described datasets.