

Real-time vocal emotion recognition in artistic installations and interactive storytelling: Experiences and lessons learnt from CALLAS and IRIS

Thurid Vogt, Elisabeth André, Johannes Wagner
Augsburg University

Steve Gilroy, Fred Charles, Marc Cavazza
Teesside University



What are non-prototypical emotions?



1. Emotions that have not been anticipated by the creator of a system because they are unlikely to occur in a particular setting.

Labeled Emotions in SmartKom: Joy, Anger, Boredom, Neutral, Sadness, Disgust, Fear

What happens if “Surprise” occurs at runtime?

2. Emotions that have been expressed in a non-prototypical manner, for example, because somebody tries to deliberately hide his or her emotional state.



Causes of non-prototypical emotions

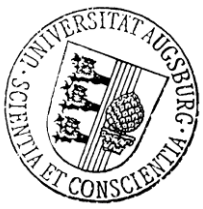


- Non-prototypical emotions may arise due to:
 - Non-prototypical events that cause non-prototypical behaviors
 - Non-prototypical responses to prototypical or non-prototypical events



Non-Prototypicality in CALLAS and IRIS

- CALLAS:
 - Development of interactive art installations that respond to the multimodal emotional input of performers and spectators in real-time
- IRIS:
 - Development of novel technologies for interactive storytelling
- Both projects represent an extreme case of prototypicality because neither the stimuli nor the emotional response to them are prototypical.
 - Prototypical stimuli would be in conflict with the creativity expected from artists.
 - The installations have been presented to a large and heterogeneous user group whose emotional response cannot be anticipated.



Art Installations in the Callas Project

E-tree by University of Teesside:

- Augmented Reality Art installation
- Tree grows and changes colors influenced by user input; among others acoustic emotions



Art Installations in the Callas Project

Galassie by Studio Azzurro

- Galassie creates stylized shapes similar to galaxies for each present user.
- The visual appearance of the galaxies depends on the user's emotional state.



Interactive Storytelling in the IRIS Project



EmoEmma by University of Teesside:

- Users can influence the outcome of the story by acting as one of the characters and their interaction mode is restricted to the emotional tone of their voice





Large Variety of Unpredictable Situations

- Installations are intended to support artists or spectators to express themselves emotionally.
- Large variety of users and situations:
 - Children vs. adults
 - Language (English, Italian)
 - Non-native vs. native speakers
 - Multiple users talking at the same time
 - Background noise
 - Occurring emotions



Large Variety of Emotional States

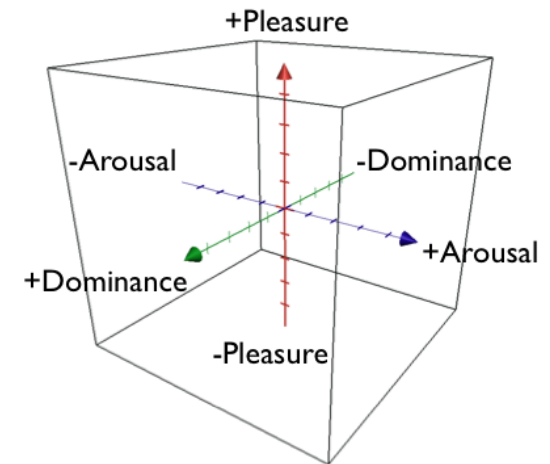


- Aesthetic experiences may be very different and are hard to predict.
- Galassie users reported 14 different emotional states: interest, transport, ludic pleasure, amazement, involvement, creation, serenity, freedom, confusion, irritation, indifference, frustration, boredom, distressed.

Large Variety of Emotional States



- Strategy:
 - train a limited set of emotion classes based on pleasure and arousal in Mehrabian's PAD (Pleasure, Arousal, Dominance) model
 - which should then subsume the actually expressed emotions at runtime
- EmoEmma:
 - five emotion classes (positive/high-arousal, positive/low-arousal, neutral, negative/low-arousal, and negative/high-arousal)





Large Variety of Emotional States

- Strategy: Contextual Interpretation of emotional states

<i>EmoVoice</i> Expectations	Negative-Active	Positive-Active
Negative-Med	Anger (High)	Surprise (High)
Negative-Low	Affinity (Very-Neg) Curiosity (Low)	Surprise (Med) Curiosity (Med)



Unexpected Emotional States and Means of Expression

- Potential Strategy: Introduction of a garbage class
 - concentrate on a few important and specific categories
 - add a “garbage” class for all other occurring emotions
- Requirements for the garbage class
 - Very general
 - Should include very different kinds of emotional expression
- Problems with garbage classes:
 - How to respond to “garbage” emotions?
 - Unexpected emotional states cannot be distinguished from unconventional ways of expressing prototypical emotions.



Training Data for Non-Prototypical Emotions

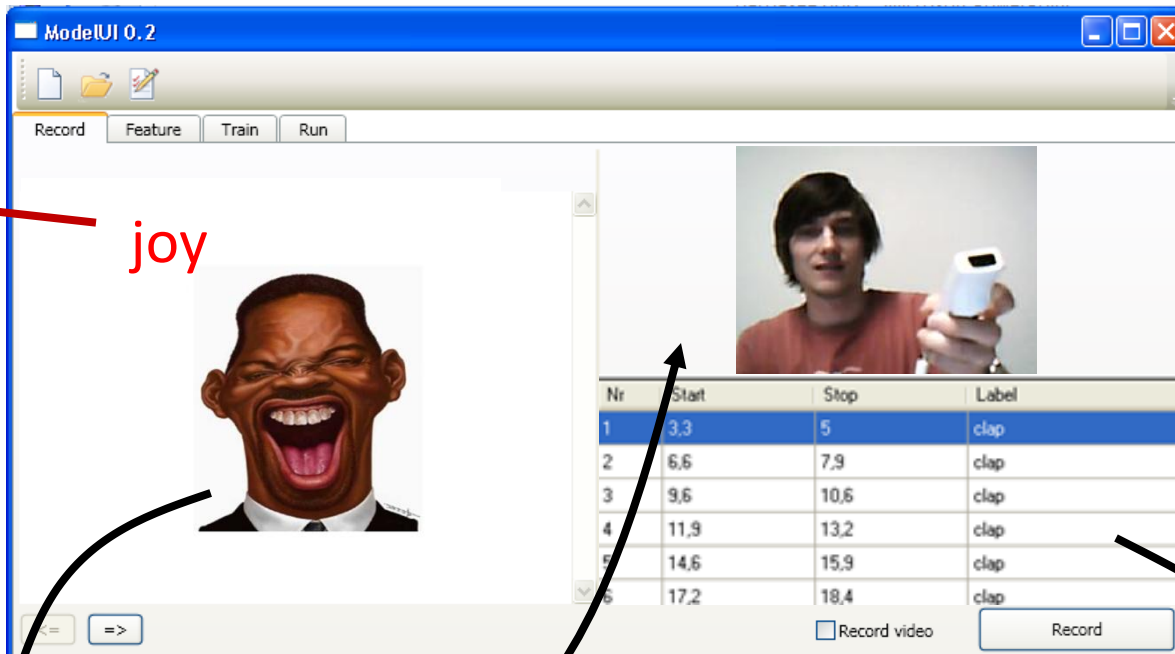


- Performance of a classifier on new test data depends strongly on the quality and similarity of the data used to train the classifier.
- Need data with
 - similar speaker groups,
 - languages (English by native and non-native speakers),
 - native languages of show case developers (Finnish, English, German, French),
 - background noise and
 - occurring emotions



Training Data for Non-Prototypical Emotions

- We designed a work flow for the showcase developers to record their own training database adjusted to their showcase.
- We integrated an easy-to-use interface for recording and training an emotional speech corpus into EmoVoice.
 - The interface offers the possibility to present stimuli that are similar to those occurring in the showcases.
 - The emotion label then results from the stimulus and labeling afterwards is not necessary.



expected user state

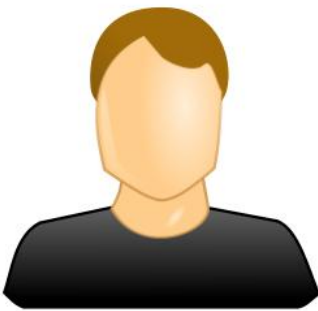
joy

pre-label

stimuli

video for later analysis

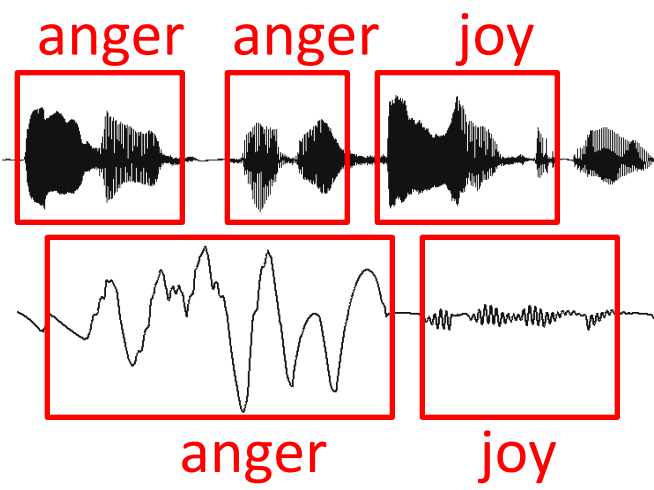
pre-segmentation

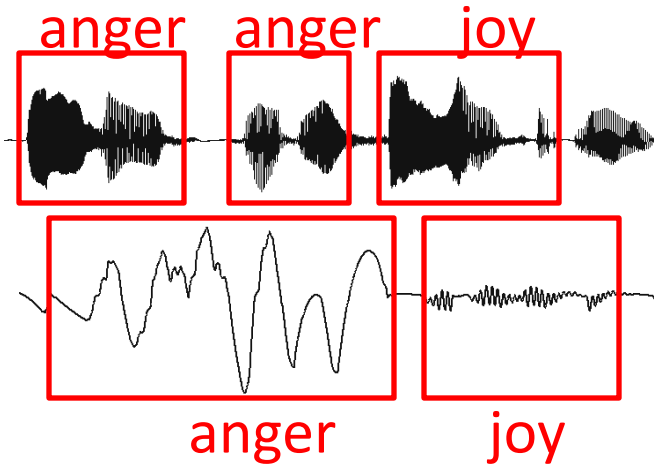


user



record data





ModelUI 0.2

Record Feature Train Run

User: stiastr, tanman, tasos, teo, tpar

Date: Apr 10 2009 16:47:11, Apr 10 2009 16:57:00, Apr 10 2009 17:06:52

Signal: audio.data, glove.data, wii01.data, wii02.data

Annotation: audio.anno, stimuli.anno

Nr	Label	Start	Stop	Duration
46	positive	245.7	250.7	5
47	positive	250.7	253.4	2.7
48	positive	255.1	256.7	1.6
49	positive	262.1	264.2	2.1
50	positive	267	267.7	0.7
51	positive	271.7	273.8	2.1
52	positive	277.7	279	1.3
53	positive	282.4	283.8	1.4
54	positive	288	289.7	1.7
55	positive	295	296.3	1.3
56	positive	300.1	302.6	2.5
57	positive	308.5	309.3	0.8
58	positive	314.8	317.6	2.8
59	positive	318.7	319.9	1.2
60	positive	324.5	325.6	1.1
61	positive	326.2	327.1	0.9
62	positive	328	328.9	0.9
63	positive	335.6	336.4	0.8

audio1

Value

Time: 195, 196, 197, 198, 199, 200, 201

Calculate

load data,
segmentation
and video

adjust boundaries
and labels

SSI Tool

extract features

anger 0.60 0.57 0.55 0.52 0.49
 anger 0.54 0.23 0.05 0.04 0.14
 anger 0.02 0.98 0.72 0.81 0.34
 joy 0.54 0.12 0.14 0.22 0.03
 joy 0.03 0.11 0.21 0.43 0.12

anger 0.60 0.57 0.55 0.52 0.49

anger 0.54 0.23 0.05 0.04 0.14

anger 0.02 0.98 0.72 0.81 0.34

joy 0.54 0.12 0.14 0.22 0.03

joy 0.03 0.11 0.21 0.43 0.12

load features

evaluation

anger:	49	12	23	->	58.33%
joy:	18	26	17	->	42.62%
neutral:	25	16	26	->	38.81%
				=>	46.59%

SSI Tool

train model



anger joy

- We need to distinguish between
 - the non-prototypicality of stimuli
 - and the non-prototypicality of behaviors
 - Weakness of the current approach:
 - In our approach, classifiers are trained based on a limited set of standard stimuli validated by psychological research.
 - At runtime, the users were exposed, however, to a larger and more diverse set of stimuli.
- Provide stimuli that are similar to those occurring in the final scenario



- To distinguish between non-prototypical emotional states and non-prototypical ways of expressing prototypical or non-prototypical emotional states.
- To decide how to respond to non-prototypical emotions
 - Ignore them?
 - Same response as to similar prototypical emotions?
- To record data to train classifiers for non-prototypical emotions



Potential Solutions

- Provide support for presentation of stimuli and recording data
 - Provide recording tools that help elicit emotions during training that are as similar as possible to the stimuli at runtime.
- Iterative design
 - Additional corpora should be collected at runtime as a basis for training new classifiers that are adapted in a better way to the showcases in which they are actually being used.