

# The INTERSPEECH 2009 Emotion Challenge

September 7, 2009



Björn Schuller<sup>1</sup>, **Stefan Steidl**<sup>2</sup>,  
Anton Batliner<sup>2</sup>

<sup>1</sup>Technische Universität München, Germany

<sup>2</sup>University Erlangen-Nuremberg, Germany



**1** FAU Aibo Emotion Corpus

**2** Emotion Challenge



# FAU Aibo Emotion Corpus

Scenario: Children interacting with Sony's pet robot Aibo



# FAU Aibo Emotion Corpus (cont.)



- 51 children (30 f, 21 m) at the age of 10 to 13
- two schools: OHM and MONT
- 8.9 hours of spontaneous speech (mainly short commands)
- 48 401 words
- 18 216 manually defined chunks based on syntactic-prosodic criteria
- emotion labelling:
  - 5 labellers (advanced students of linguistics)
  - 11 emotion categories (emotion-related user states)
  - labelled on word level



Now available for **scientific, non-commercial use**

- <http://www5.cs.fau.de/FAUAiboEmotionCorpus>
- fee: 150 €
- data:
  - audio files (wav format, mono, 16 bit, 16 kHz) for the chunks
  - transliteration of the chunks
  - corresponding lexicon
  - information about the age and the gender of the children
  - original emotion labels of the five human labellers for each word
  - single emotion label for the whole chunk



# Emotion Challenge

## Challenges

- realistic data:
  - spontaneous, naturally occurring emotions/emotion-related states
  - non-prompted, non-acted
  - low emotional intensity
- handling *all* data
  - sparse data problem (highly unbalanced classes)
  - no selection of prototypical data
  - merging rare categories, rest class



# Emotion Challenge (cont.)

## Classification

- chunk level: heuristic mapping of emotion labels
- 2-class problem:

#	NEG	IDL	$\Sigma$
train (OHM)	3,358	6,601	9,959
test (MONT)	2,465	5,792	8,257
$\Sigma$	5,823	12,393	<b>18,216</b>

- 5-class problem:

#	A	E	N	P	R	$\Sigma$
train (OHM)	881	2,093	5,590	674	721	9,959
test (MONT)	611	1,508	5,377	215	546	8,257
$\Sigma$	1,492	3,601	10,967	889	1,267	<b>18,216</b>

- speaker independent classification



# Emotion Challenge (cont.)

## Sub-Challenges

- 1 Feature Sub-Challenge**
  - optimisation of feature extraction/selection;  
classifier settings fixed
- 2 Classifier Sub-Challenge**
  - optimisation of classification techniques;  
feature set given
- 3 Open Performance Sub-Challenge**
  - optimisation of feature extraction/selection and  
classification techniques





# Baseline Results

## Standard set of acoustic features

- 384 features using the free software openSMILE  
<http://sourceforge.net/projects/opensmile/>

LLD (16 · 2)	functionals (12)
( $\Delta$ ) ZCR	mean
( $\Delta$ ) RMS Energy	standard deviation
( $\Delta$ ) F0	kurtosis, skewness
( $\Delta$ ) HNR	extremes: value, rel. position, range
( $\Delta$ ) MFCC 1-12	linear regression: offset, slope, MSE

## Classification

- dynamic modelling with HMMs (HTK)
- static modelling with SVMs (WEKA 3)



# Baseline Results (cont.)

## Results by dynamic modelling:

		recall [%]		precision [%]	
		<b>UA</b>	WA	UA	WA
<i>2-class</i>	# HMM states				
	1	<b>62.3</b>	71.7	65.2	69.8
	3	<b>62.9</b>	57.5	61.1	70.0
	5	<b>66.1</b>	65.3	63.6	71.3
<i>5-class</i>	1	<b>35.5</b>	50.8	29.6	57.1
	3	<b>35.2</b>	34.7	27.5	57.2
	5	<b>35.9</b>	37.2	29.3	59.0

UA: unweighted average recall/precision

WA: weighted average recall/precision



# Baseline Results (cont.)

## Results by static modelling:

	process		recall [%]		precision [%]	
	1	2	UA	WA	UA	WA
<i>2-class</i>	-	-	<b>62.7</b>	72.6	66.4	70.6
	S	B	<b>67.6</b>	68.3	65.2	72.3
	B	S	<b>67.7</b>	65.5	64.8	72.7
<i>5-class</i>	-	-	<b>28.9</b>	65.6	35.5	57.0
	S	B	<b>38.2</b>	39.2	30.0	59.7
	B	S	<b>38.0</b>	32.2	29.4	59.8

S: standardisation

B: training balancing by SMOTE

UA: unweighted average recall/precision

WA: weighted average recall/precision

Thank you for your attention!