EMOTION AND SPEECH OF CHILDREN
WITH AUTISM SPECTRUM CONDITIONS

Erik Marchi* (Advisor: Björn Schuller)

Machine Intelligence & Signal Processing Group (MISP), MMK, Technische Universität München, GERMANY

erik.marchi@tum.de

* Credits: Florian Eyben, Jürgen Geiger, Felix Weninger, Fabien Ringeval (TUM)

ASC-Inclusion: Interactive Emotion Games

- Children with Autism Spectrum Conditions (ASC) show **difficulties** in communication, social interaction and understanding

- ASC-Inclusion develops interactive **games** to **assist** children with ASC (high-functioning) to recognise, understand and express **emotions**

- It assists children with ASC to improve their **socio-emotional communication skills**

- It combines **voice**, face, and body gesture analysis and provides **corrective feedback**

http://www.asc-inclusion.eu
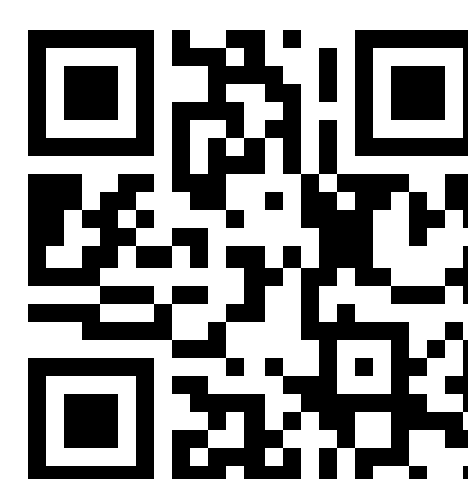
B. Schuller, E. Marchi, S. Baron-Cohen et al., "The state of play of ASC-Inclusion: Integrated Internet-Based Environment for Social Inclusion of Children with Autism Spectrum Conditions", in Proc. 2nd International Workshop on Digital Games for Empowerment and Inclusion (IDGEI 2014) held in conjunction with IUI 2014, ACM, Haifa, Israel, 24.02.2014.

Emotions in the voice of children with ASC

- Novelty: Automatic recognition of **emotions in atypical children's voice**
- Task: Arousal, Valence and Emotion (9-classes)

- ASC-Inclusion Database (ASC-I DB)
  - Languages: English, Hebrew, Swedish
  - 7-10 children with ASC, aged 6-10 (Focus group)
  - 10 typically developing children, aged 5-9 (Control group)
  - **9 emotion** elicited with 9 stories
    (happy, sad, angry, surprised, afraid, ashamed, calm, proud, neutral)

- **Audio features:** INTERSPEECH 2012 Speaker Trait Challenge set
  - **Spectral**, **voice quality** and **prosodic** features

- **Classifier:** linear SVMs with Sequential Minimal Optimisation (SMO)

opensmile.sourceforge.net

% UAR	#Classes	Focus group	Control group
Arousal	2	84.9	89.0
Valence	2	82.1	81.8
Emotion	9	42.6	55.9

E. Marchi, B. Schuller, A. Batliner, S. Fridenzon, S. Tal, O. Golan: "Emotion in the Speech of Children with Autism Spectrum Conditions: Prosody and Everything Else", in Proc. 3rd Workshop on Child, Computer and Interaction (WOCCI 2012), Satellite Event of INTERSPEECH 2012, Portland, OR, USA, 14.09.2012.

Atypical Speech Classification

- Recognition of speaker with Autism Spectrum Conditions (ASC) by their acoustics
- Few studies exists on automatic diagnosis classification from speech
- Application: health-care, socially assistive robotics, clinical tools, …

- **Audio features:**
  - INTERSPEECH 2013 Computational Paralinguistics Challenge set
  - INTERSPEECH 2012 Speaker Trait Challenge set

- **Classifier:** linear SVMs trained with SMO

- **Databases:**
  - ASC-Inclusion Database (ASC-I DB)
    529 instances, 20 children (11 TYP, 5 HF, 4 AS)
  - Child Pathological Speech Database (CPSD)
    2.5k instances, 99 children (64 TYP, 12 ASC, 13 DYS, 10 NOS)

% UAR	ASC-I DB	CPSD
Typicality	80.0	90.7
Diagnosis	82.6[1]	67.1

[1] In the ASC-I DB it is a two-class task while in CPSD it is a 4-class task

E. Marchi, A. Batliner, B. Schuller, S. Fridenzon, S. Tal, O. Golan: "Speech, Emotion, Age, Language, Task, and Typicality: Trying to Disentangle Performance and Feature Relevance", in Proc. First International Workshop on Wide Spectrum Social Signal Processing (WS³P 2012), held in conjunction with SocialCom 2012, IEEE, Amsterdam, The Netherlands, 03.-06.09.2012.
B. Schuller et al. : "The INTERSPEECH 2013 Computational Paralinguistics Challenge: Social Signals, Conflict, Emotion, Autism", in Proc. INTERSPEECH 2013, ISCA, Lyon, France, 2013.

Speech Dereverberation and ASR

- Reverberation severely degrades ASR performance

- Robust recognition of reverberated speech:
  - Multi-channel dereverberation with **Correlation Shaping** (CS)
  - Bidirectional Long Short-Term Memory (LSTM) RNNs for phoneme prediction
  - GMM-LSTM double-stream decoding

- **Audio features:** MFCCs
- **Database** 2014 REVERB Challenge:
  - **Sim:** 7.8k (train), 1.4k (devel), 2.1k (test) utterances (WSJCAM0 corpus)
  - **Real:** 179 (devel), 372 (test) utterances (MC-WSJ-AV corpus)
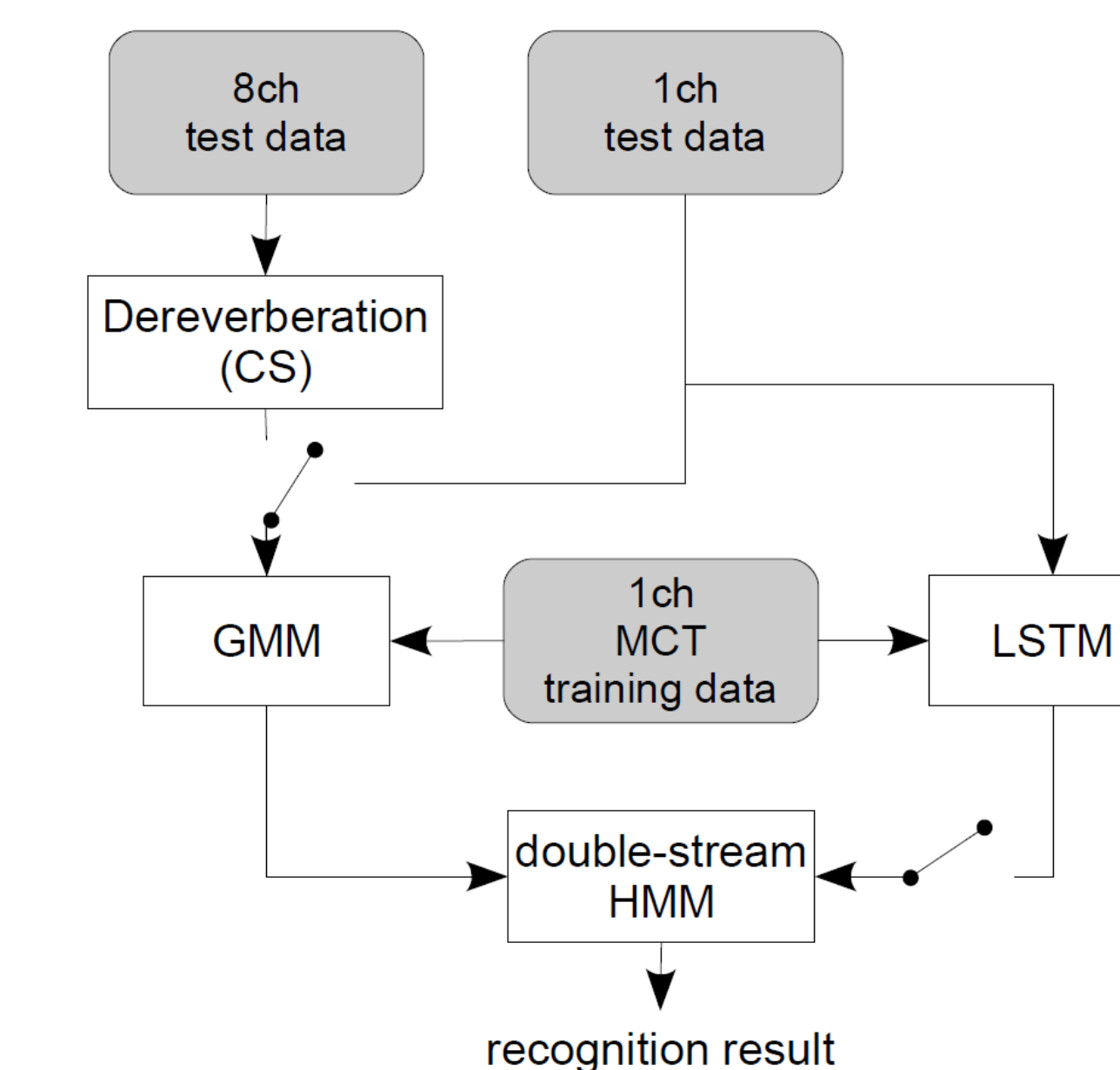
% WER	Sim Data	Real Data
Baseline	25.25	48.85
CS + Baseline	20.62	38.83
GMM	17.26	39.76
CS + GMM	13.20	28.15
GMM + LSTM	13.75	36.78
CS + GMM + LSTM	11.19	28.13

J. T. Geiger, E. Marchi, B. Schuller, G. Rigoll: "The TUM System for the REVERB Challenge: Recognition of Reverberated Speech Using Multi-Channel Correlation Shaping Dereverberation and BLSTM Recurrent Neural Networks", to appear in Proc. REVERB Workshop 2014 (REVERB 2014), ICASSP 2014, Florence, Italy, 10.05.2014

sourceforge.net/projects/currennt/