Detecting the Instant of Emotion Change From Speech using a Martingale Framework

Zhaocheng (David) Huang, Julien Epps

Never Stand Still | Faculty of Engineering | School of Electrical Engineering and Telecommunication

Outline

- Motivation
- Related Work
- Conventional Martingale Framework for Change detection
- Proposed Martingale for Emotion Change Detection
- Experimental Results
- Conclusions

Motivation behind Emotion Change Detection

- Speech based Emotion Recognition
  - A per-file basis emotional signal processing
  - Increasing popularity in continuously tracking emotion dimensions
  - However, no systematic insights into *Emotion Changes*

- Research question:
  - Can we detect the *time*, at which emotion change occurs?

- Detecting emotion changes in time is important
  - Human Computer Interaction
  - Emotion Regulation
  - Surveillance purpose to detect outburst of emotions
  - Medical purpose to monitor emotion changes of patients

Related Work

- Only a small number of studies for Emotion Change Detection
  - Emotional evolutions are detectable [Böck 2015]
  - Large changes in emotion dimensions using topic model [Lade 2013]
  - Likelihood ratio based methods [Huang 2015]

- Limitations:
  - Speaker Change Detection Methods
    - Phonetic and speaker variability
    - Lack of large emotional databases

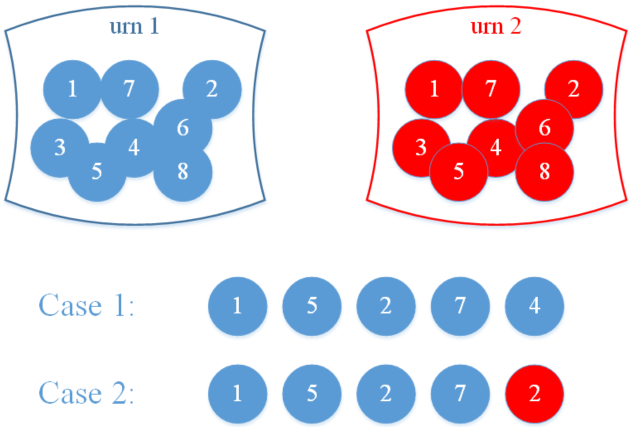
- Generic change-point detection method
  - Martingale framework by exchangeability testing [Vovk 2003]
    - Successfully be used in change-point detection in image processing [Ho 2010]
    - Speech rate change detection [Yasuda 2012]

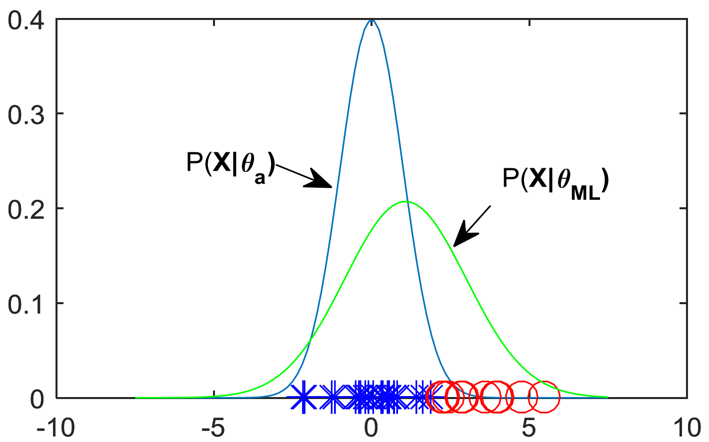
Exchangeability

- Definition:
  - A sequence of random variables $\{x_1, x_2, \dots, x_n\}$ (observed one by one) is exchangeable if their joint distribution remains unchanged regardless of any permutation π of indices $\{1, \dots, n\}$, namely

$$P(x_1, x_2, \dots, x_n) = P(x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(n)})$$

  - Examples





- Lack of exchangeability implies changes in distribution/model
  - Main idea for Exchangeability Testing [Vovk 2003, Ho 2010]

Exchangeability Testing using Martingale

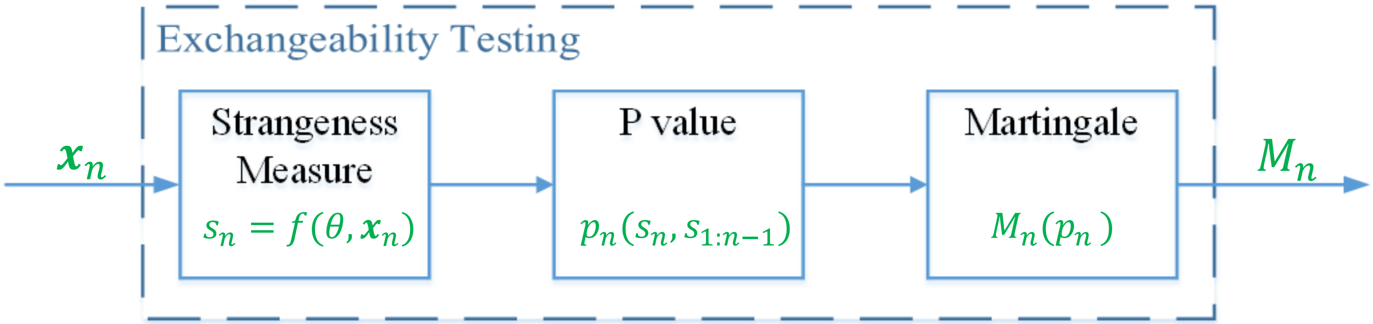
- Exchangeability Testing

$$Exchangeable \rightarrow H_0 : \theta_1 = \theta_2 = \ldots = \theta_N = \theta_a$$

$$Non-exchangeable \rightarrow H_1 : \exists n \text{ with } \theta_1 = \ldots = \theta_n \neq \theta_{n+1} = \ldots = \theta_N$$

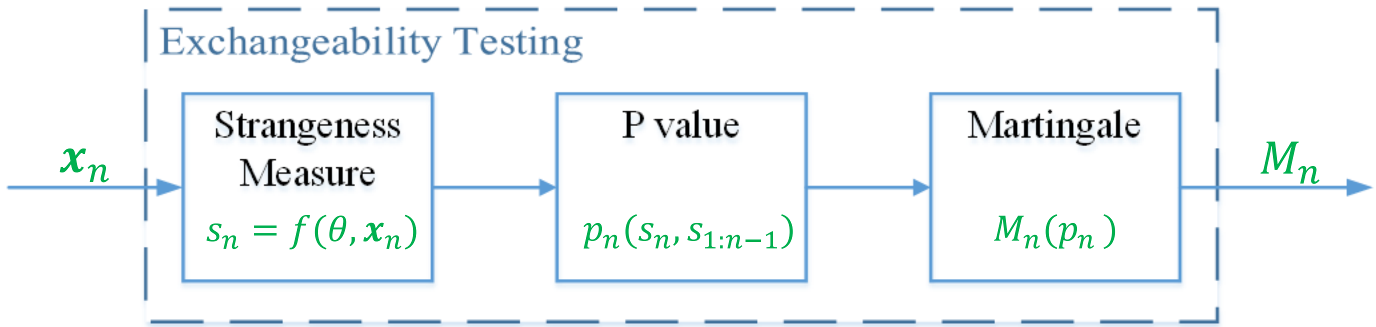
- Definition of *Martingale* [Ho 2010]
  - Given a sequence of random variables $\boldsymbol{X}_i : \{x_1, x_2, \ldots x_i : 1 \leq i \leq \infty\}$, M_i is a measurable function of $\boldsymbol{X}_i$ and $E(|M_i|) < \infty$, then M_n is a
    - *Martingale* if $E(M_{n+1}|\boldsymbol{X}_{n+1}) = M_n$
    - *Super-martingale* if $E(M_{n+1}|\boldsymbol{X}_{n+1}) < M_n$
    - *Sub-martingale* if $E(M_{n+1}|\boldsymbol{X}_{n+1}) > M_n$
  - Given $\boldsymbol{X}_{n+1}$, the expected M_{n+1} remains unchanged/decrease/increase
  - M_n measures the confidence of rejecting H_0
    - Reject H_0, when $M_n > \lambda$, where λ is a predefined threshold

Conventional Martingale for Change-Point Detection

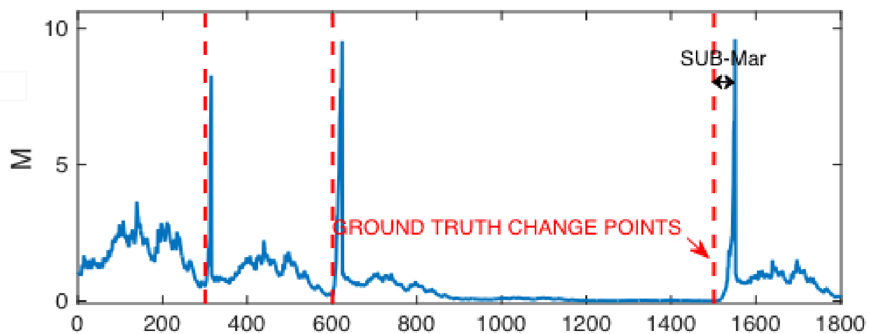


- Calculate M_n for each feature vector x_n, based on model θ
  1. Calculate s_n: how unlikely x_n comes from θ
  2. Calculate p_n: compare s_n to all previous $s_{1:n-1}$
  3. Calculate M_n: increase/decrease/unchanged
- Compare M_n with a predefined threshold λ
  - if $0 < M_n < \lambda$, continue detection
  - if $M_n \geq \lambda$, reject H_0 and restart detection with $M_{n+1} = 1$
- If there is a change point
  - Large $s_n \rightarrow$ small $p_n \rightarrow$ increasing M_n

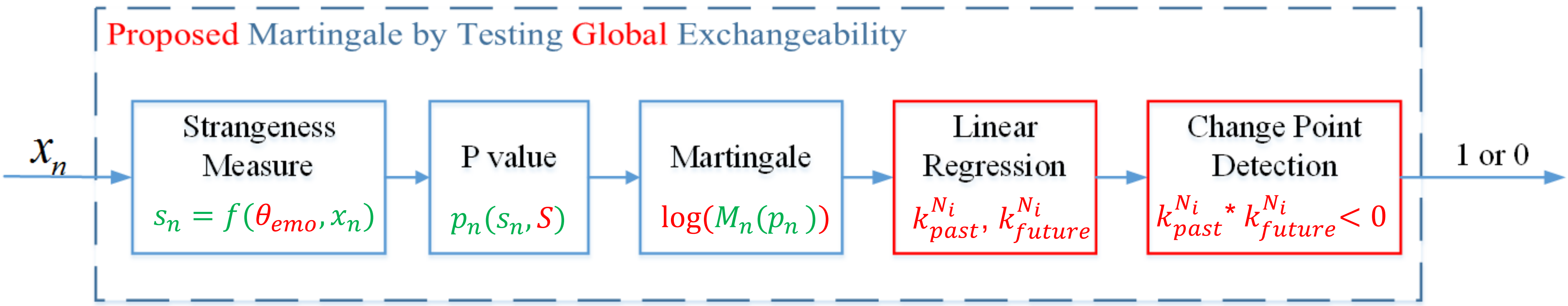
Conventional Martingale for Change-Point Detection



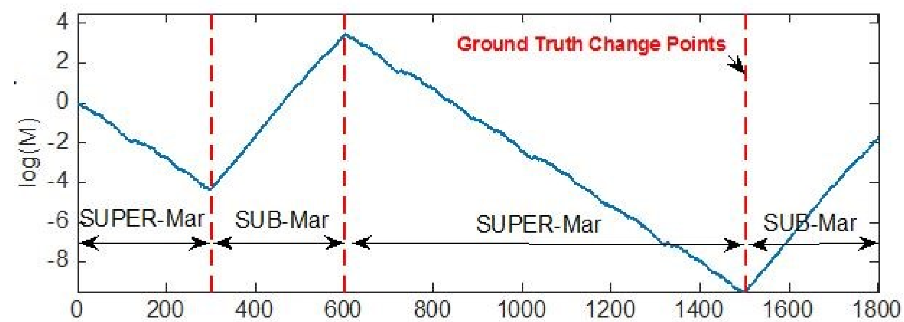
- Limitations:
  - **Phonetic variability**
  - Incapability of handling long term no-change situation
  - Large delay time



Proposed Martingale for Emotion Change Detection

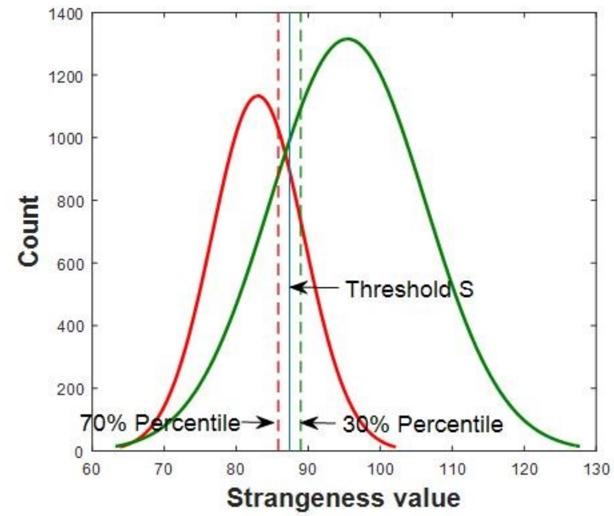


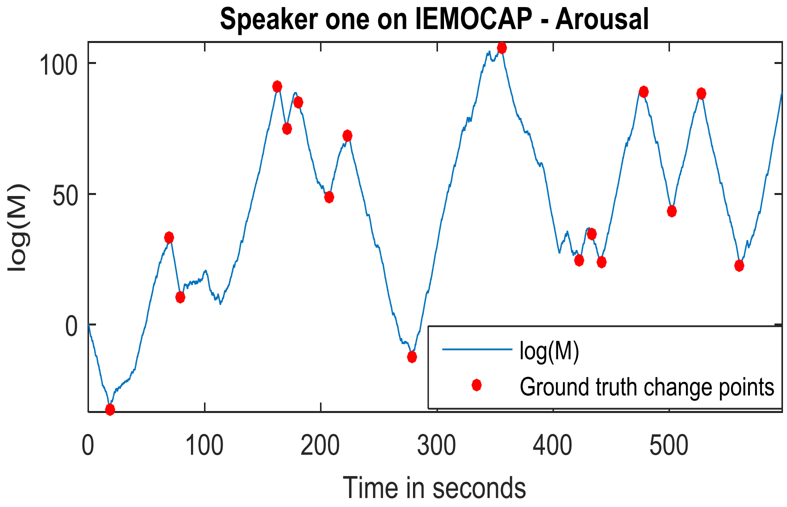
- Solution:
  - Global emotional model θ_{emo}
    - Global S from θ_{emo}
  - Fix p_n based on S
    - $s_n < S \rightarrow \log(M_n) \downarrow$
    - $s_n \geq S \rightarrow \log(M_n) \uparrow$
  - Peaks and troughs
    - Linear regression (slope)



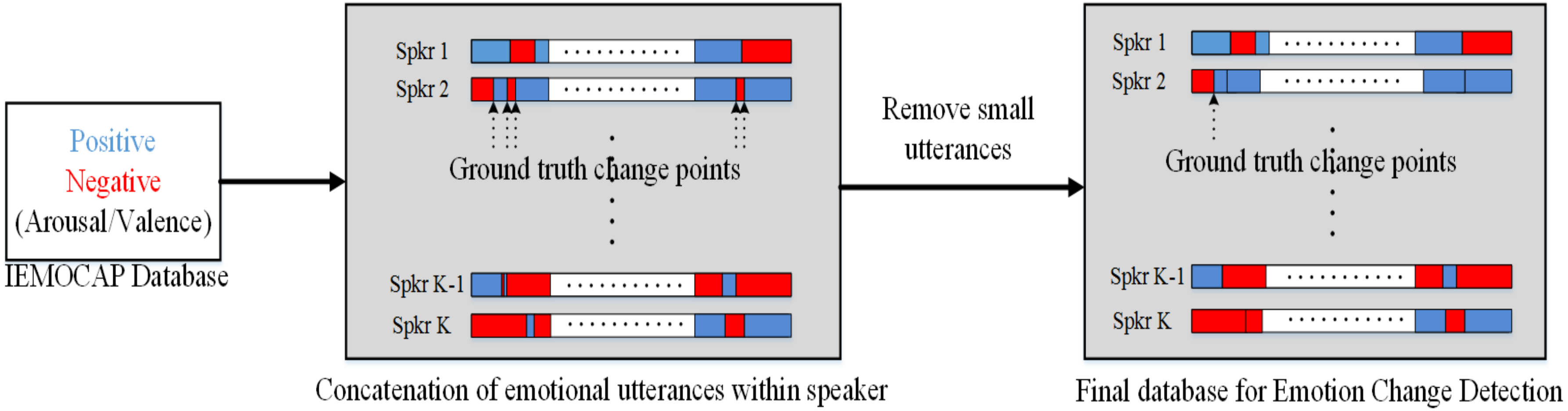
Proposed Martingale for Emotion Change Detection

1. Strangeness
   - θ_{emo} is a GMM model for one emotion
     - $s_n = -\log(P(\boldsymbol{x}_n|\theta_{emo}))$
   - $S = \frac{(\boldsymbol{s}_Q^1 + \boldsymbol{s}_{100-Q}^2)}{2}$, where Q means Q percentile of all strangeness values belongs to θ_{emo}
2. P value
   - $p_n = \begin{cases} p^{sub}, & s_n \geq S \\ p^{super}, & s_n < S \end{cases}$
3. Log martingale value: $\log(M_n(p_n))$
4. Two-pass linear regression
   - $k_{past}^{N_1} * k_{future}^{N_1} < 0$
   - $k_{past}^{N_2} * k_{future}^{N_2} < 0$





Database: IEMOCAP [Busso 2008]



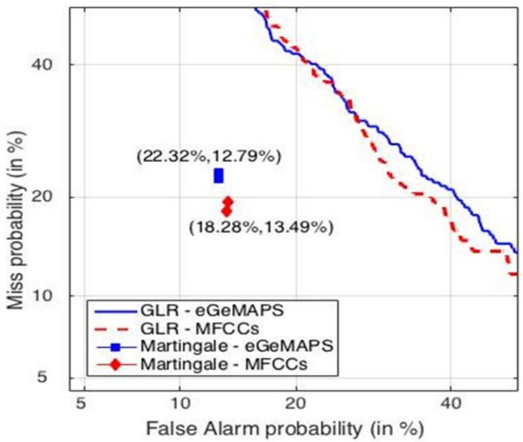
Experimental Settings

- Features
  - 13 MFCCs and their first order derivatives, VoiceProb = 0.55
  - 28 eGeMAPS low level descriptors [Eyben 2015], VoiceProb = 0.7
- Emotional Model - Gaussian Mixture Models (GMMs)
  - Leave-one-speaker-out 16-Mixtures GMM for Neutral, Negative Arousal and Negative Valence
- Baseline
  - Generalized Likelihood Ratio, no prior emotion information [Huang 2015]
- Parameters for Martingale
  - $p^{sub} = 0.25, p^{super} = 0.5, N_1 = 10, N_2 = 60$
  - $Q = 50\%\ and\ 70\%$

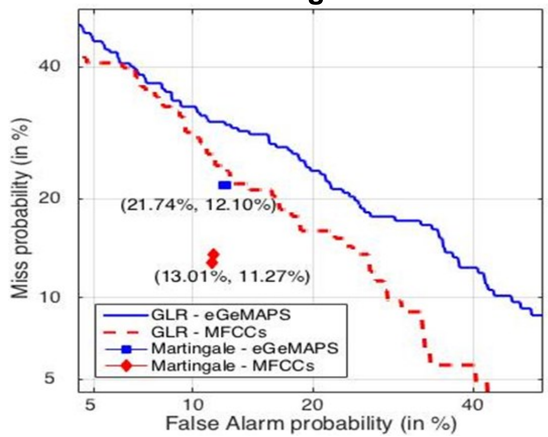
Experimental Results

- Baseline: Generalized Likelihood Ratio, No prior emotion information [Huang 2015]

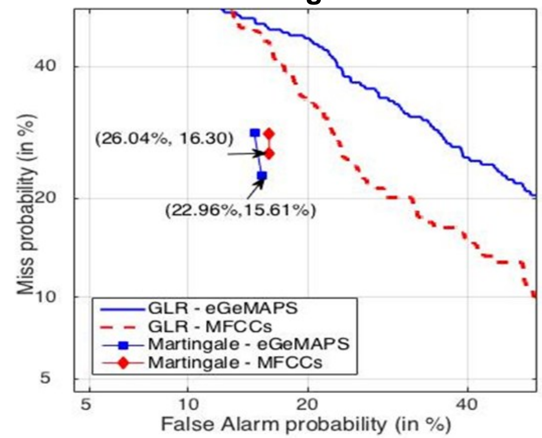
Neutral vs Emotional

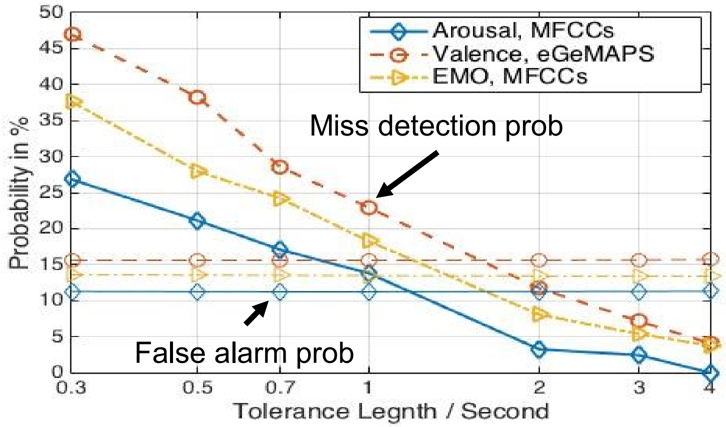


Pos vs Neg Arousal



Pos vs Neg Valence





Conclusions & Limitations

- Conclusions
  - We proposed a modified Martingale framework for detecting emotion change points in time from speech based on exchangeability testing
    - Robustness to potential variability using one emotion model
    - A lower delay and able to handle non-change over a long time
  - A general framework for emotion change detection
    - Requires only one model
    - Functionals may improve the detection performances

- Limitations
  - Only handle two classes.
    - However, can use multiple martingales
  - Related to performances of emotion recognition algorithms
  - Database is artificially concatenated

References

- HUANG, Z., EPPS, J. & AMBIKAIRAJAH, E. An Investigation of Emotion Change Detection from Speech. INTERSPEECH, 2015.
- LADE, P., BALASUBRAMANIAN, V. N., VENKATESWARA, H. & PANCHANATHAN, S. Detection of changes in human affect dimensions using an Adaptive Temporal Topic model. Multimedia and Expo (ICME), 2013 IEEE International Conference on, 2013. IEEE, 1-6.
- VOVK, V., NOURETDINOV, I. & GAMMERMAN, A. Testing exchangeability on-line. ICML, 2003. 768-775.
- HO, S.-S. & WECHSLER, H. 2010. A martingale framework for detecting changes in data streams by testing exchangeability. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 32**,** 2113-2127.
- BUSSO, C., BULUT, M., LEE, C.-C., KAZEMZADEH, A., MOWER, E., KIM, S., CHANG, J. N., LEE, S. & NARAYANAN, S. S. 2008. IEMOCAP: Interactive emotional dyadic motion capture database. *Language resources and evaluation,* 42**,** 335-359.
- EYBEN, F., WENINGER, F., GROSS, F. & SCHULLER, B. Recent developments in opensmile, the munich open-source multimedia feature extractor. Proceedings of the 21st ACM international conference on Multimedia, 2013. ACM, 835-838.
- BÖCK, R. & SIEGERT, I. Recognising Emotional Evolution from Speech. ERM4CT'15, 2015. 13-18.

THANK YOU

zhaocheng.huang@unsw.edu.au

UNSW AUSTRALIA

NICTA