MULTIMEDIA DATA ANALISYS FOR EMOTION RECOGNITION

Author: Fernando García Novo
Thesis Advisor: Carmen García Mateo
Departamento de Teoría do Sinal e Comunicación

MOTIVATION OF THE WORK

The motivation that has led to the choice of this area of knowledge for the Doctoral Thesis is threefold:

- The communication of emotions is crucial for social relationships and survival (Ekman, 1992), and the voice is possibly one of the main channels for conveying them. Therefore, it is vital to have systems that allow automatic recognition of the emotions of people through this signal, which also have a wide range of aplications including notably:
  - oHuman-computer interface.
  - o It helps human-human communication.
  - o eHealth, especially in the area of mental health.
  - o Managing voice files.

- It is a field of research much less developed than the automatic speech recognition.

- The rapid development in machine learning, especially with regard to the Deep Neural Networks. The possibility of applying these new techniques in the field of automatic recognition of emotions will open many lines of research with promising results.

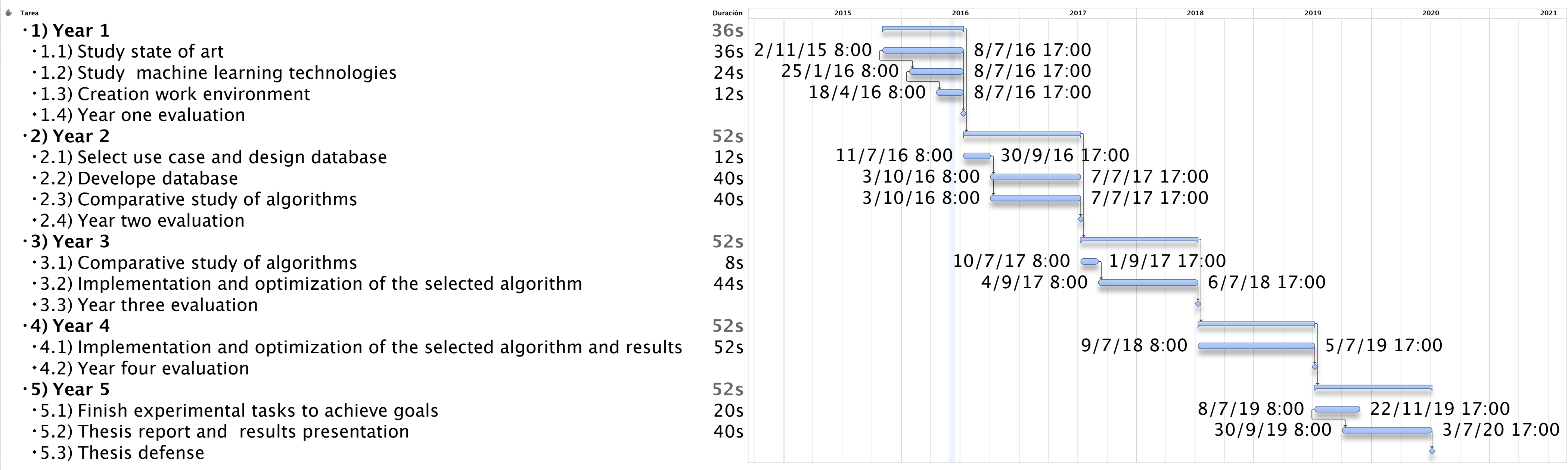
THESIS OBJECTIVES

Develop a methodology that allows to detect the emotions of the people through multimedia data.

SECUNDARY OBJECTIVES

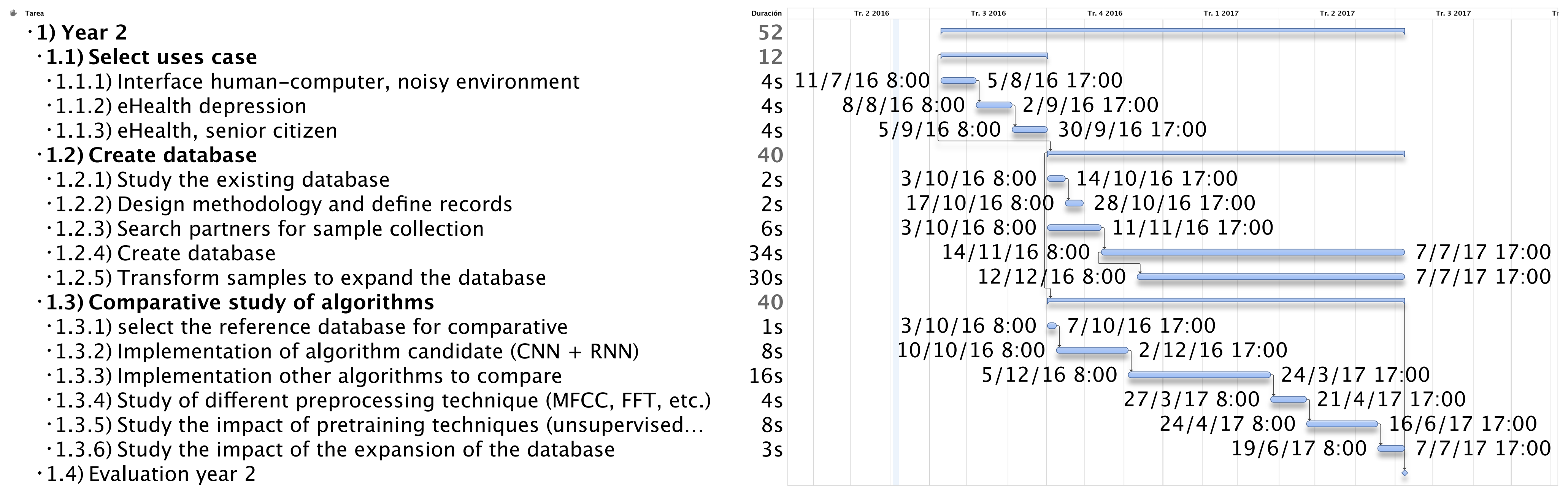
- Detect cases of use in which emotion recognition have an application of interest.
- Select multimedia records that they are interesting for the resolution of the problem: voice, image, movement records, etc.
- If it is necessary, we will create a new database enabling us to innovative results.
- Detect those algorithms and/or methodologies that they have the better behavior to solving the problem.
- Improve, if it is possible, the behaviour of the selected algorithms.
- Define an architecture that enables in real-time to solve these problems.

RESEARCH PLAN



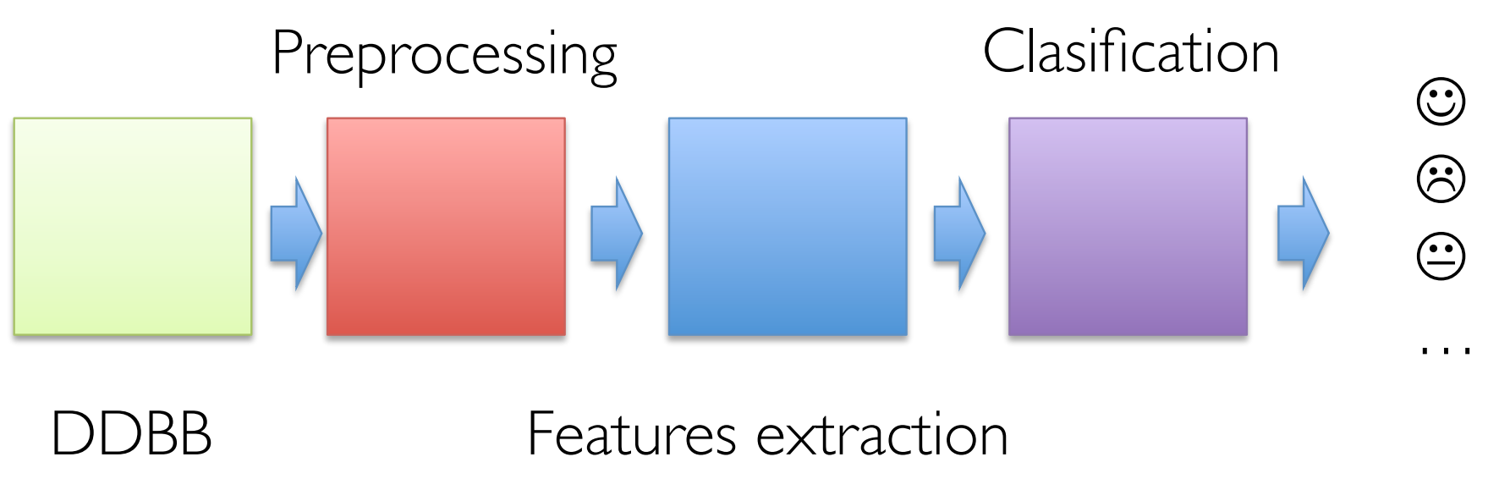
Tarea	Duración		
1) Year 1	36s		
1.1) Study state of art	36s	2/11/15 8:00	8/7/16 17:00
1.2) Study machine learning technologies	24s	25/1/16 8:00	8/7/16 17:00
1.3) Creation work environment	12s	18/4/16 8:00	8/7/16 17:00
1.4) Year one evaluation			
2) Year 2	52s		
2.1) Select use case and design database	12s	11/7/16 8:00	30/9/16 17:00
2.2) Develope database	40s	3/10/16 8:00	7/7/17 17:00
2.3) Comparative study of algorithms	40s	3/10/16 8:00	7/7/17 17:00
2.4) Year two evaluation			
3) Year 3	52s		
3.1) Comparative study of algorithms	8s	10/7/17 8:00	1/9/17 17:00
3.2) Implementation and optimization of the selected algorithm	44s	4/9/17 8:00	6/7/18 17:00
3.3) Year three evaluation			
4) Year 4	52s		
4.1) Implementation and optimization of the selected algorithm and results	52s	9/7/18 8:00	5/7/19 17:00
4.2) Year four evaluation			
5) Year 5	52s		
5.1) Finish experimental tasks to achieve goals	20s	8/7/19 8:00	22/11/19 17:00
5.2) Thesis report and results presentation	40s	30/9/19 8:00	3/7/20 17:00
5.3) Thesis defense			

NEXT YEAR PLANNING



Tarea	Duración		
1) Year 2	52		
1.1) Select uses case	12		
1.1.1) Interface human–computer, noisy environment	4s	11/7/16 8:00	5/8/16 17:00
1.1.2) eHealth depression	4s	8/8/16 8:00	2/9/16 17:00
1.1.3) eHealth, senior citizen	4s	5/9/16 8:00	30/9/16 17:00
1.2) Create database	40		
1.2.1) Study the existing database	2s	3/10/16 8:00	14/10/16 17:00
1.2.2) Design methodology and define records	2s	17/10/16 8:00	28/10/16 17:00
1.2.3) Search partners for sample collection	6s	3/10/16 8:00	11/11/16 17:00
1.2.4) Create database	34s	14/11/16 8:00	7/7/17 17:00
1.2.5) Transform samples to expand the database	30s	12/12/16 8:00	7/7/17 17:00
1.3) Comparative study of algorithms	40		
1.3.1) select the reference database for comparative	1s	3/10/16 8:00	7/10/16 17:00
1.3.2) Implementation of algorithm candidate (CNN + RNN)	8s	10/10/16 8:00	2/12/16 17:00
1.3.3) Implementation other algorithms to compare	16s	5/12/16 8:00	24/3/17 17:00
1.3.4) Study of different preprocessing technique (MFCC, FFT, etc.)	4s	27/3/17 8:00	21/4/17 17:00
1.3.5) Study the impact of pretraining techniques (unsupervised…	8s	24/4/17 8:00	16/6/17 17:00
1.3.6) Study the impact of the expansion of the database	3s	19/6/17 8:00	7/7/17 17:00
1.4) Evaluation year 2			

RESULTS



Preprocessing

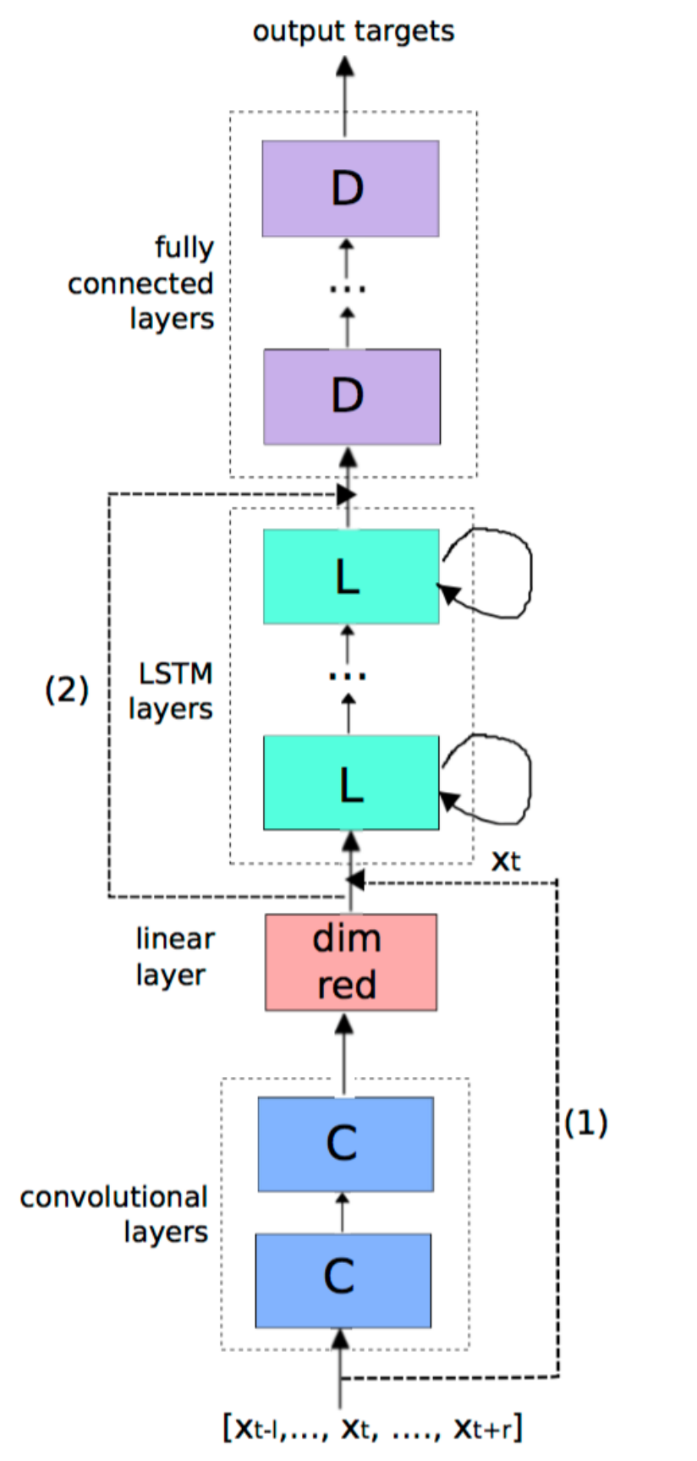
Clasification

☺
☹
😐
…

DDBB

Features extraction

DDBB: Extend with transformations and adding noise
Preprocessing: FFT, Mel scale, log transformation, multiwindow.
Features extraction: Pretrainning (autoencoder), choose pooling, activation, architecture, Inception, optimization algorithm,…
Clasification: choose architecture, optimization algorithm,….



CLDNN Architecture [12]

REFERENCES

Emotions

[1] P.Ekman,W.V.Friesen,M.O'Sullivan,A.Chan,I.Diacoyanni- Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti et al., "Universals and cultural differences in the judg- ments of facial expressions of emotion." Journal of personality and social psychology, vol. 53, no. 4, p. 712, 1987.

Emotions and Vocal source

[2] R.Banse, K.R.Scherer, "Acoustic profiles in vocal emotion expression", Journal of Personality and Social Psychology, Vol.70, 614-636, 1996.

[3] Patrik N. Juslin and Petri Laukka "Communication of Emotions in Vocal Expression and Music Perfomance: Different Channels, Same Code?", Psychologicla Bulletin, Vol. 129, No. 5, 770-814

[4] R. Horwitz, T. Quatieri, B. Helfer, B. Yu, J. Williamson, and J. Mundt, "On the relative importance of vocal source, system, and prosody in human depression," in IEEE International Conference on Body Sensor Networks (BSN), 2013, pp. 1–6.

Deep Neuronal Networks

[5] T.Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolu- tional, Long Short-Term Memory, Fully Connected Deep Neural Networks," in to appear in Proc. ICASSP, 2015.

[6] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich;"Going Deeper with Convolutions", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1-9

Emotion Recognition

[7] Improved Strategies for Speaker Segmentation and Emotional State Detection. Ph D. Thesis, Paula López Otero. Universidade de Vigo, 1995.

[8] Sainath, Tara N. et al. "Learning the speech front-end with raw waveform CLDNNs." INTERSPEECH (2015).

[9] Kun Han, Dong Yu, Ivan Tashev, "Speech Emotion Recognition Using Deep Neuronal Network and Extreme Learning Machine", INTERSPEECH (2014).

[10] James R. Williamson, Thomas F. Quatieri, Brian S. Helfer, Gregory Ciccarelli, and Daryush D. Mehta. 2014. Vocal and Facial Biomarkers of Depression based on Motor Incoordination and Timing. In Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge (AVEC '14).

[11] W. Q. Zheng, J. S. Yu and Y. X. Zou, "An experimental study of speech emotion recognition based on deep convolutional neural networks," Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on, Xi'an, 2015, pp. 827-831.doi: 10.1109/ACII.2015.7344669.

[12] Sainath, T. N., Vinyals, O., Senior, A., & Sak, H. (2015, April). Convolutional, long short-term memory, fully connected deep neural networks. In Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on (pp. 4580-4584). IEEE.

fgnovo@gmail.com