Sadness Detection in Emotional Acted Speech

Francisco Solís[1], Sergio Suárez[1], Cornelio Yáñez[1], Antonio García[1],
Erick Zúñiga[2],

[1] Computing Center Research, National Polytechnic Institute, Juan de Dios Batiz street, Nueva Industrial Vallejo, D. F., México
[2] Superior Technologic Institute of Huichapan, known adress, Huichapan, Hidalgo, México
fx05xiuhcoatl@gmail.com, {ssuarez, cyanez}@cic.ipn.mx gamejoan@gmail.com, zusea28@hotmail.com

Abstract. Emotional speech recognition has been studied using different approaches, which some works use real emotions and other uses acted ones, usually real emotional speech databases include like two or three emotions and acted ones have five or more, for this work Berlin Emotional Speech Database [1] was selected due to its availability, which has 535 sentences expressed in seven emotions (anger, boredom, disgust, happiness, sadness, fear and neutral), recorded by 10 professional actors (five men and five women), 10 different sentences in German language were considered, the contribution of this paper, is a 99.2% recognition rate for sadness emotion, in literature [3] 86.35% has been report, using same database.

Keywords: Emotional speech recognition, stratified cross validation, feature selection, emotional database.

1 Introduction

Speech interaction in human-machine communication has relevant information besides verbal, such as emotive data; there are some applications like measuring immersion in virtual reality environments [8], automatic search in television or films [3], recognition and synthesis of emotions in agents systems [9], etc.

Several databases has been created with emotional states classification purposes, most of them handle anger, sadness, happiness and fear mainly, most databases were generated using simulated emotions, due to real emotions are hard to classify for humans or computers [7].

2 Corpus

Berlin Emotional Database [1] was chosen, due to its availability, this corpus contains 535 utterances recorded by 10 professional actors, 5 men and 5 women. The database includes 7 emotions (anger, sadness, happiness, fear, boredom, disgust and neutral) and 10 sentences, for example ‘Der lappen liegt auf dem eisschrank’ (The tablecloth is lying on the fridge).

This database has 127 instances of anger, 81 of boredom, 46 of disgust, 69 of fear, 71 of happiness, 62 of sadness and 79 of neutral. In order to see the database as a two class problem, all no sadness records were labeled as ‘x’ thus the subset ‘x’ has 473 instances.

3 Features

All features extracted are statistical measures: maximum, minimum, mean, median, mode and standard deviation of amplitudes energy peaks, silence durations, pitch, pitch durations and sonority; maximum, mean, median, mode and standard deviation of energy; finally maximum, minimum, mean, median mode and standard deviation of maximums, minimums, means, modes and standard deviations of 13 MFCC’s coefficients.

4 Classification

For classification and feature selection processes, weka open source software was used, available in [10]. 16 features were selected using Sequential Backward Elimination: median of amplitude energy peaks and pitch; mean and standard deviation of silence durations; standard deviation of pitch durations; max, min, mean, median and standard deviation of sonority; min and median of MFCC’s maximums; min and standard deviations MFCC’s minimums; min of MFCC’s medians and mean of MFCC´s standard deviations.

Simple-Logistic was experimentally chosen, the model can be seen in Table 1.

Sadness	Neutral
28.25+ [Median of amplitude energy peaks]*-0.13+ [Mean silence durations]*-0.17+ [Standard deviation of silence durations]*0.11+ [Standard deviation of pitch durations]*0.11+ [Median of pitch]*-0.03+ [Maximum of sonority]*1.28+ [Minimum of sonority]*-25.73+ [Mean of sonority]*-20.02+ [Standard deviation of sonority]*-65.9+ [Median of MFCC´s maximums]*3.44+ [Standard deviation of MFCC´s minimums]*-4.65+ [Minimum of MFCC´s medians]*-0.64+ [Mean of MFCC´s standard deviations]*6.75	-28.25+ [Median of amplitude energy peaks]*0.13+ [Mean silence durations]*0.17+ [Standard deviation of silence durations]*-0.11+ [Standard deviation of pitch durations]*-0.11+ [Median of pitch]*0.03+ [Maximum of sonority]*-1.28+ [Minimum of sonority]*25.73+ [Mean of sonority]*20.02+ [Standard deviation of sonority]*65.9+ [Median of MFCC´s maximums]*-3.44+ [Standard deviation of MFCC´s minimums]*4.65+ [Minimum of MFCC´s medians]*0.64+ [Mean of MFCC´s standard deviations]*-6.75

Table 1. Simple-Logistic model obtain by weka open source software [10].

5 Results and discussion

Using the k-fold cross validation, with k=10, 99.25 % of recognition rate, confusion matrix can be seen in Table 2. Two false positives and 2 false negatives are the errors; a detailed accuracy by class is represented in Table 3.

Sadness	X	
60	2	Sadness
2	471	x

Table 2. Confusion matrix of sadness detection in Berlin emotional database.

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.968	0.004	0.968	0.968	0.968	Sadness
0.996	0.032	0.996	0.966	0.966	x

Table 3. Detailed accuracy by class.

In literature, we found 82 % for sadness recognition [11], never the less, best recognition rate of this emotion was found in [3] with a 86.35 %. In this paper when the problem of 7 emotions is considered as a 2 class problem, recognizing sadness can be improved.

Acknowledgments

The authors would like to thank the following institutions for their economical support to develop this work: CONACyT, SNI, and National Polytechnic Institute, COFAA (Grant 20090807, "Herramientas para el reconocimiento y traducción de voz y clasificación de emociones"), SIP and CIC.

References

1. Berlin emotional speech database, http://www.expressive-speech.net/
2. Ringeval, F., Chetouani, M.: Exploiting a Vowel Based Approach for Acted Emotion Recognition. Lecture Notes in Computer Science, vol. 5042/2008, pp. 243-254. Springer, Heidelberg (2008)
3. Xiao, Z., Dellandréa, E., Dou, W., Chen, L.: Hierarchical Classification of Emotional Speech, Technical Report RR-LIRIS-2007-006, LIRIS UMR 5205 CNRS, (2007)
4. Truong, K., Van Leeuwen, D.: An 'open-set' detection evaluation methodology for automatic emotion recognition in speech. In Workshop on Paralinguistic Speech – between models and data, pp. 5-10 (2007)
5. Cichosz, J., Slot, K.: Emotion recognition in speech signal using emotion-extracting binary decision trees. In Proceedings of Affective Computing and Intelligent Interaction, Lisbon, Portugal (2007)
7. Ververidis, D., Kotropoulos, C.: A State of the Art Review on Emotional Speech Database. In Proceedings of the 1st Richmedia Conference, pp. 109-119 (2003)

8. Ververidis, D., Kotropoulos, C.: Automatic speech classification to five emotional states based on gender information. In Proceedings Eusipco 2004, Vienna, Austria, pp. 341-344 (2004)
9. Vladimir, H., Zdravco, K.: Context-Independent Multilingual Emotion Recognition from Speech Signals. In International Journal of Speech Technology, volume 6, number 3, pp. 311-320 (2003)
10. http://www.cs.waikato.ac.nz/ml/weka/
11. Mansoorizadeh, M., Charkari, N. M.: Speech Emotion Recognition: Comparison of Speech Segmentation Approaches. In IKT'07 Mashad, Iran, (2007)