International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 16 (2017) pp. 5760-5767

Extraction of Emotions from Speech-A Survey

A Pramod Reddy[1] and * V. Vijayarajan[2]

[1]*Research Scholar, School of computer science and engineering, VIT University, Vellore, India.*

[2]*Associate Professor, School of computer science and engineering VIT University, Vellore, India.*

[1]*Orcid: 0000-0002-3912-3302*

Abstract

The study drives an increasing attention towards recognition of emotions spontaneously from speech with thought-provoking task. This paper sketches mainly about the popular datasets available and the classifiers widely used by many authors used for automatic emotion recognition. The databases are reviewed for the purpose of availability, the size of datasets and the number of speakers with the size of dataset. This is an attempt to give a short review about the work on Emotion recognition from speech.

Keywords: Automatic emotion recognition, SVM, HMM.

INTRODUCTION

The extraction of automatic-emotion recognition (AER) through speech is providing a great opportunity in the latest years with emerging field of applications that can benefit from these technologies. A more technical definition is given by Jurafsky[1], where he defines ASR as the building of system for mapping acoustic signals to a string of words[2]. He continues by defining " *automatic speech understanding*(ASU) as *extending the goal to producing some sort of understanding of the sentence"*. Here let us assume a speaker's emotional state system which helps in building a natural and also effective human- interaction machine with interactive interface based on the user behavior. This can be used with a deep learning environment, in which a system can detect different types of emotional states i.e.,(bored, angry, happiness, sadness) users as well as allows a change of style and level of the material provided, or provides a new approach with motive encouragement. This section also improves the possible ways of improving speech emotion recognition. From decades numerous speech recognition systems have been projected where various techniques have been identified and implemented by researches. By applying a proper classifier includes accurate feature extraction and selection which includes prosodic and spectral features like pitch, intonation, duration by means of MFCC, LPCC, PLPCS, RASTA. The different classifiers recycled for emotion recognition Gaussian-Mixture Model (GMM), Hidden-Markov Model(HMM), Artificial-Neural Network(ANN), Support-vector machine(SVM). In a study[3] a novel work so called end-to-end speech emotion recognition with RECOLA database on convolutional Neural Network(CNN) having input features such as MFCC, Perceptual Linear Prediction (PLP) with accuracy of fixed learning rate of 2 .10-3. And in another study[4] the male and female uțterances the average emotion recognition is 61% and 66% respectively. The paper comprises of the following sections: Section two review of database with a short survey which includes problem statement followed by section three briefly describes the classification schemes in speech emotion recognition system. With an idea about feature extraction and feature selection followed by a short review of classifiers used for emotion recognition finally the last section concludes the paper.

Uses of AER

- Medicine: -*Rehabilitation* – It helps in monitoring a patient condition. *Counseling* – identifying the client's emotional state and providing the counseling .*Health and safety*– patents emotions after treatment. *Autism* – usually children of small age groups struggle to project their expressions and with daily treatments it helps in developing their potential. *Music-therapy for depression* – it has a powerful influence on patients with stress and depression. Gives progressive results for those patients who are suffering from Alzheimer.
- E-learning :- Adjust the presentation style of an online tutor by detecting the state of the learner. Creates more interactive and effective environment with positive attitude to maximize learning by providing feedback and guide towards a solution. Monitoring:-*Car driver* – detect the state of driver and gives alert. *Call center system* – detecting voice signal with Anġer and responds to that call with high

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 16 (2017) pp. 5760-5767

priority. In cockpit stressed speech is monitored for better performance.

- Entertainment: -*Music player*:- Recognize the mood and emotion of the user and satisfy the needs .*Law* :- In lie detection

Emotional Speech Databases

A very small speech database SUSE[4]- speech under simulated emotions which is recorded in Telugu language by All India Radio using professional actors situated at Vijayawada for IIT kharagpur is employed in this. The speech corpus consists of 5 Male and 5 Female professionally trained artists in expressing the desired emotions15 sentences ranging 12000 uțterances approx. It consists of a single statement spoken by 30 students 5 times each containing four emotions as Happy, Anġer, Compassion and Neutral.

The SUSE speech corpus is recorded using SHURE dynamic microphone in a closed room avoiding disturbances. Therefore, 1200 uțterances recorded from one male and one female are used for analysis out-of-which generating a statically model with 960 uțterances and the remaining are used for comparison. Another databases addressed from Table-1, adult-directed emotions while FAUAIBO[5] emotion corpus all about 9 hours of German recordings of 51 children, 20 male and 31 female with a age group of 10 years to 13 years and interacting with Sony pet robot called Aibo recorded spontaneously with a close-talk microphone which has been segmented into small meaningful chunks using semantic prosodic criteria.

Robert Plutchik's theory say's that there are eight basic emotions and here their meanings are considered from the popular dictionaries like merrian-webster, oxford [6] with all possible emotions represented with Figure 1.

- Fear:- emotion comes with an unpleasant situation caused from pain, dAnġer or feeling afraid.
- Anġer:- involves a strong feeling of aggravation, uncomfortable situation stress, displeasure, or hostility
- Sadness:- A feeling caused with disadvantage or loss due to anything.
- Joy :- feeling happy. Other words are *happiness*, *gladness.*
- Disgust:- A feeling with strong disapproval, nasty, dislike
- Surprişe :- occurred with an unexpected event or shock .
- Trust :- belief that someone or something is reliable a positive emotion; admiration is stronger; acceptance is weaker
- Anticipation :- in the sense of looking forward positively to something which is going to happen. Expectation is more neutral.

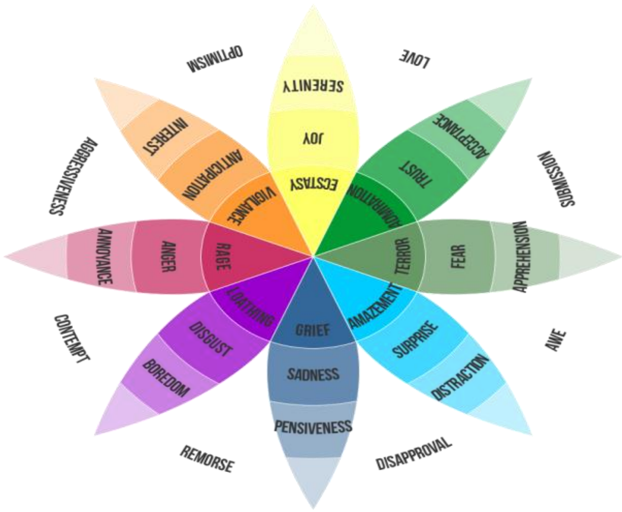

Figure 1: Robert Plutcheks wheel of Emotions [6]

Type of database: Based on the readings the database is categorized into two 1. Natural 2. Acted

1. Natural database: There are some databases which are collected naturally/spontaneously with a microphone where few children are interacting with a pet robot[7]. And collected from call centers[13]. In such databases only few emotions like natural, angry, sad, happy can be extracted.
2. Acted: They are again categorized into two Professional and Non-Professional actors for extracting all possible emotions like Happy, Anġer, sadness, compassion, fear, natural, positive, Borden, disgust, joy and surprişe.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 16 (2017) pp. 5760-5767
© Research India Publications. http://www.ripublication.com

LITERATURE SURVEY

Features of collective emotional speech corpus

Table 1: Some features of available emotional speech corpus

Corpus Name	Accessibility	Language	Size & duration	Type of nature	Emotions
RECOLA data base[3]	Open access with license fee	French	5 min for each train 16 subjects validation 15 sub test 15 sub	Natural Non-Professional	All possible emotions
IITKGP-SESC: Speech Database [4]	Open access on request	Telugu	~ 7 hours of duration with 12000 uțterances	Professional actors from AIR	Anġer, Compassion, Happy, Neutral
FAU AIBO corpus[9][5]	Exclusive	German	9 hours of 51 children between 10-13 ages	Spontaneous recordings with micro-phone	Anġer, Empathetic, Neutral, Positive & Rest
IE MOCAP[10]	Exclusive	English	12 hours of audio visual data	acted	Anġer, happiness, sadness, neutrality
Berlin emotional database[11]	Open access and free	German	800 uțterances(10actors -7 emotions -10 uțterances) ¼ 800 uțterances	Professional actors	Anġer, joy, sadness, fear, disgust, boredom, neutral
Smartkom mobile db[12]	Exclusive	German		Natural - Recorded in mobile handset	Anġer,Happyness,Natural,Sadness.
Danish emotional database[15]	Open access with license fee	Danish	4actors of 5 emotions(2words * 9sentences * 2passages)	Non-professional actors	Anġer, joy, sadness, surprişe, neutral
Natural[16]	Exclusive	Mandarin	388 uțterances, 11 speakers, 2 emotions	Call centers	Anġer, neutral
ESMBS[17]	Exclusive	Mandarin	720 uțterances, 12 speakers, 6 emotions	Nonprofessional actors	Anġer, joy, sadness, disgust, fear, surprişe
INTERFACE[18]	Commercially available	English, Slovenian, Spanish, French	English (186 uțterances), Slovenian (190 uțterances), Spanish (184 uțterances), French (175 uțterances)	Actors	Anġer, disgust, fear, joy, surprişe, sadness, slow neutral, fast neutral
KISMET[19]	Exclusive	American English	1002 uțterances, 3 female speakers, 5 emotions	Nonprofessional actors	Approval, attention, prohibition, soothing, neutral
Baby Ears[20]	Exclusive	English	509 uțterances, 12 actors (6 males + 6 females), 3 emotions	Mothers and fathers	Approval, attention, prohibition
SUSAS[21]	Open access	English	16,000 uțterances, 32 actors (13 females + 19 males)	Speech under simulated and actual stress	Stress, Tracking Task, Amusement Park Roller-oaster,
MPEG-4[22][23]	Exclusive	English	2440 uțterances, 35 speakers	U.S. American movies	Joy, Anġer, disgust, fear, sadness, surprişe, neutral
Beihang University	Exclusive	Mandarin	7actors -5 emotions -20 uțterances	Non-professional actors	Anġer, joy, sadness, disgust, surprişe

FERMUS- III	Open access with license fee	English& German	2829 uțterances, 7 emotions, 13 actors	Automotive environment	Anġer, disgust, joy, neutral, sadness, surprişe
KES	Exclusive	Korean	5400 uțterances, 10 actors	Non-professional actors	Neutral, joy, sadness, Anġer
CLDC[24]	Exclusive	Chinese	1200 uțterances,30 actors	Non-professional actors	Joy, Anġer, surprişe, fear, neutral, sadness
Hao Hu et al	Exclusive	Chinese	actors -5 emotions -40 uțterances	Non-professional actors	Anġer, fear, joy, sadness, neutral
Amir et al.	Exclusive	Hebrew	60 Hebrew and 1 Russian actors		Anġer, disgust, fear, joy, neutral, sadness
Pereira	Exclusive	English	actors -5 emotions -8 uțterances	Non-professional actors	0 Anġer, Anġer, joy, neutral, sadness
LDC Speech Transcripts[25]	Commercially available	English	7actors -15 emotions - 10 uțterances	Professional actors	Neutral, panic, anxiety, hot Anġer, cold Anġer, despair, sadness, elation, joy, interest, boredom, shame, pride, contempt

RECOLA: Remote Collaborative and Affective Interactions
IITKGP-SESC Simulated Emotion Speech Corpus

We propose a methodology rooted in the above studies for a confusion problem. Usually there are many such statements which will come with different emotions Let us assume a funny sentence a conversation made between two persons X and Y. where X states this statement looking towards Y "The chicken is ready to eat" the statement has an acute confusion problem that states with two meanings the first states as the chicken is really hungry and it is about to eat and the other states as the chicken is cooked and we are ready to eat it for what state Y has to react for the first or second assuming Y has been invited for lunch by X at his villa. In an unsupervised learning identifying different sources of ambiguity of different classes over regional languages within the state.

Speech Emotion Recognition System

The basic ER system from speech comprises of following steps as shown:

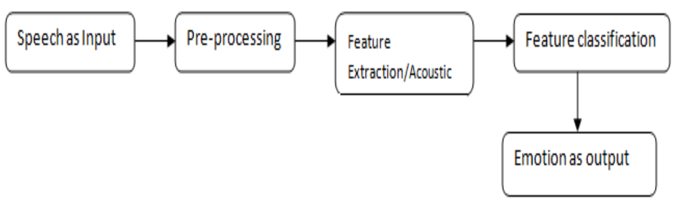

Figure 2: Speech Emotion Recognition System

The speech samples are taken as input in first phase , if no standard database used then those samples are pre-processed for removing the noise from samples using many tradeoffs like PRATT, Audacity, ocenaudio, wavepad, sony creative noise reduction by considering the spectrogram of speech sample. The main purpose of doing this is to obtain high frequency characteristics of signals.

Feature Extraction and classification:

Another important task in emotion recognition system is Feature Selection and Feature Extraction. Mainly they are classified into two Prosodic Features and Spectral Features.

The prosodic features include pitch, Intonation, Duration, The mean Standard deviation, minimum, maximum, range and variance of pitch. The features are selected various feature selection algorithms and they are processed. Spectral features are extracted using LPA[26],PLP,PLPCS, RASTA,FT, FFT[27], MFCC[28] with average rate of recognition of 87.5% EMODB and CASIA dataset and extracting emotions like sadness, happiness, natural, Borden, anxiety and surprise.

Feature Selection

The intuition behind using the acoustic and prosodic features is to summarize the intentional variations observed in humans.

The acoustic features are

1. Maximum & Minimum counter ascent energy.
2. Mean and Median values of energy.
3. Mean and Median of energy decline in values.
4. Maximum of pitch frequency
5. Mean and Median of pitch frequency.
6. Maximum duration of pitch in terms of frequency.
7. Mean and Median of first format
8. Rate of change in formats.
9. Speed in voice frames.

Feature Analysis

In the first discussed database the baseline is obtained by considering the basic features such as duration of the sample, pitch, standard deviation and energy. The entire speech dataset is captured with a rate of 16kHz and the other is recorded using a DAT-recorder with 48kHz and later corrected to 16kHz using a high quality wireless set. The major databases uses the parameter classification mentioned in table-2 and the fundamental frequency feq0 where the order of frequencies are mentioned as feq1, feq2 thereof.

The basic features are

Table 2: The basic features

parameter	units	features
Duration	Seconds(sec)	Length of the speech sample
Energy	Decibel(Db)	Mean and standard deviation
Rate	kHz	Frequency of sample
HNR(Harmonic Noise Ratio)	Decibel(Db)	Mean,SD,range,duration

CLASSIFICATION SCHEMES

By applying a proper classifier includes accurate feature extraction and selection which includes prosodic and spectral features like pitch, intonation, duration by means of MFCC, LPCC, PLPCS, RASTA for low level frequencies. The different classifiers described in table-5 for emotion recognition where the traditional classifiers include Gaussian-Mixture Model (GMM), Hidden-Markov Model (HMM), Artificial-Neural Network (ANN), Support-vector machine (SVM) and the advanced features like General Regression NN (GRNN), Deep Neural Network (DNN).

In Machine Learning algorithms SVM[11]- Support Vector Machine algorithms is widely used pattern recognition algorithm among all others because it is efficient and simple with a recognition ratio ranges from 75 – 81.29% good efficiency rate with improving quality. The GRNN method has a recognition rate of 77 -80.24% with a less training time but with a high computational difficulty therefore it is not suitable for analyzing new training set in such cases back propagation algorithm is used for classification using Neural Network.

Table 3: Table describes the classifier accuracy

Author	Algorithm	Classification	Speaker	Database	Accuracy
BJORN SHULLER [29]	Viterbi-beam search algorithm with token passing with MFCC	SVM	Both	EMODB	84.3%
GEORGE TRIGEIRGS[3]	Machine Learning based back propagation algorithm	CNN(Convolutional Neural Network) with LSTM Network	Both	RECOLA	2.10^{-3} learning accuracy
Longfei Li[14]	Machine Learning based back propagation algorithm	DNN-HMM	Speaker Recognition in unsupervised learning	eNTERFACE'05 and Berlin	77.92%

According to the investigations made by the author Bjorn Shuller[3] the result carried out for independent speaker recognition using Leave-one-Speaker-out(LOSO) manner on EMODB by extracting chunks. This chunk analysis is realized by one-pass Viterbi beam algorithm with MFCC for fast pre segmentation and Brute _Force for subsequent subset collection. Low Level Descriptors for first order delta coefficients depicted as pitch, energy, shimmer, jitter, noise ratio and realized by function F, Scalar Feature X and independent speaker x represented as

$$F: X \rightarrow x \in \boldsymbol{R}$$

DNN-HMM:-

The investigations carried out by author Longfei Li[14] on an unsupervised Learning where Restricted Boltzmann Machine (RBM), a graphical model constructed from binary stochastic hidden units method is used for pre training for initializing Deep Neural Network by using traditional logistic cost function depicted as

$$y = \frac{1}{1 + e - (b + xw)}$$

Where input feature is denoted by *x,* weights between connections *w,* bias *b,* on super-fix to *e* with output denoted as *y*. and the likelihood probability *p*(*ot*|*qt*) as

$$P(o|q) = \frac{P(q|o)P(o)}{P(q)}$$

In this P(o) is derived as words segments from the sentence and P(q) as the whole statement of the set. Using Viterbi algorithm the likelihood performance is calculated with probability λ as P(o|λ).

The results noticed with the comparison Berlin corpus has higher accuracy when it is implemented on DNN-HMM unsupervised learning

Table 4: Comparison between databases

Method	ENTERFACE'05 DB	BERLIN DB
GMM-HMM	42.22	76.18
Shallow-NN-HMM (2 hidden layer)	36.67	57.86
DNN-HMM (discriminative pre-training, 6 hidden layers)	53.89	77.92
DNN-HMM(unsupervised pre-training, 6 hidden layers)	41.67	74.28
MLP-HMM (back-propagation, 6 hidden layers)	43.33	69.94

Analysis of some available active classifiers

Table 5: Classification schemes employed in ESR systems

Classifier	Recognition rate	Average training time	Sensitivity
HMM	75.5–78.5%	less	Sensitive
GMM	74.83–81.94%	Much less	Sensitive
ANN	51.19–52.82% 63–70%	Back-propagation: large	Sensitive
SVM	75.45–81.29%	Large	Insensitive
GRNN[30]	77 -80.24%	Much less	Efficient computing

S V M Classification

SVM – a supervised learning methods widely used for classification and regression with early practical implementation since 90's with high performance, simple and efficient computation of machine learning algorithms. The independent speaker identification[28]experiments using Beihang speech corpora the HMMs system with SVM system or without SVM are 76.1% and 57.8% respectively. According to[31] optimization of parameters for SVM for independent speaker "who is talking" and recognition of hand-written letters for classification problem by optimizing Euclidean distance to 7 k-Nearest Neighbors but it can be used for regression SVR sophisticatedly and still it can be optimized with weighted order classes – NN technique or ranking can also be implemented and from the discussions of[32], while retrieving facial expressions ambiguities are not discussed as it is a major confusion in human face recognition.

Hidden Markov Models

The Hidden Markov model, widely used in the literature but its classification property is not up to the mark and they have a serious shortcoming that they are statically inefficient for modeling data and when compared with DNN-HMM has improved efficiency of 83.33% when compared with GMM-HMM model using ENTERFACE'05 dataset[14].

CONCLUSIONS

From the above readings we come to know about the successful classifier and reliable -database to be selected for the accurate findings of emotions. Twenty six papers were reviewed and tabulated with the features as corpus names, languages, size in terms of length of the sentence collected as male, female and children voices. Most common emotions searched and extracted are Happiness', Sadness, Disgust, Neutral along with other features such as joy, Borden, fear and surprise. The extraction rate depends on the classifier used. There are some disadvantages in the databases due to low quality recordings of samples the recognition ration has been decreased and the accuracy in DBN[10] networks is ranging from 56 to 57%. With new level of contributions I will conclude for creation of new paradigm in Machine learning and Deep Learning with increasing efforts in education, medicine, psychology and agriculture.

REFERENCES

[1] D. Jurafsky, "Speech and language processing: An introduction to natural language processing," *Comput. Linguist. speech Recognit.*, 2000.

[2] D. Jurafsky and J. H. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition." MIT Press.

[3] G. Trigeorgis *et al.*, "Adieu Features ? End-To-End Speech Emotion Recognition Using A Deep Convolutional Recurrent Network," *Icassp 2016*, pp. 5200–5204, 2016.

[4] S. G. Koolagudi, S. Maity, V. A. Kumar, S. Chakrabarti, and K. S. Rao, "IITKGP-SESC : Speech

Database for Emotion Analysis," pp. 485–492, 2009.

[5] B. Schuller, S. Steidl, and A. Batliner, "The INTERSPEECH 2009 emotion challenge," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, pp. 312–315, 2009.

[6] R. Plutchik, *Emotion: A psychoevolutionary synthesis*. Harpercollins College Division, 1980.

[7] H. Cao, R. Verma, and A. Nenkova, "Speaker-sensitive emotion recognition via ranking: Studies on acted and spontaneous speech," *Comput. Speech Lang.*, vol. 29, no. 1, pp. 186–202, 2015.

[8] W. Li, Y. Zhang, and Y. Fu, "Speech emotion recognition in E-learning system based on affective computing," *Proc. - Third Int. Conf. Nat. Comput. ICNC 2007*, vol. 5, no. Icnc, pp. 809–813, 2007.

[9] Y. Attabi and P. Dumouchel, "Anchor models for emotion recognition from speech," *IEEE Trans. Affect. Comput.*, vol. 4, no. 3, pp. 280–290, 2013.

[10] Y. Kim, "DEEP LEARNING FOR ROBUST FEATURE GENERATION IN AUDIOVISUAL EMOTION RECOGNITION Yelin Kim , Honglak Lee , and Emily Mower Provost ∗ University of Michigan Electrical Engineering and Computer Science , Ann Arbor , Michigan , USA," pp. 3687–3691, 2013.

[11] P. Shen, Z. Changjun, and X. Chen, "Automatic Speech Emotion Recognition using Support Vector Machine," *Electron. Mech. Eng. Inf. Technol. (EMEIT), 2011 Int. Conf.*, vol. 2, pp. 621–625, 2011.

[12] T. Vogt and E. André, "Improving Automatic Emotion Recognition from Speech via Gender Differentiation," *Proc. Lang. Resour. Eval. Conf.*, no. January, pp. 1123–1126, 2006.

[13] M. S. Hossain, G. Muhammad, M. F. Alhamid, B. Song, and K. Al-Mutib, "Audio-Visual Emotion Recognition Using Big Data Towards 5G," *Mob. Networks Appl.*, vol. 21, no. 5, pp. 753–763, 2016.

[14] L. Li *et al.*, "Hybrid Deep Neural Network--Hidden Markov Model (DNN-HMM) Based Speech Emotion Recognition," *2013 Hum. Assoc. Conf. Affect. Comput. Intell. Interact.*, pp. 312–317, 2013.

[15] 22. Engberg, I. S., & Hansen, A. V. (1996). Documentation of the danish emotional speech database des. Internal AAU report, Center for Person Kommunikation, Denmark, "Documentation of the danish emotional speech database des.," *Intern. AAU report, Cent. Pers. Kommun. Denmark*, vol. 22., 1996.

[16] and L. C. D. S. Morrison, Donn, Ruili Wang, "Ensemble methods for spoken emotion recognition in call-centres.," *VMorrison, Donn, Ruili Wang, Liyanage C.De Silva.*, p. 98–112., 2007.

[17] L. D. T.Nwe,S.Foo, "Speechemotion recognition using hidden Markov models," *SpeechCommun.*, p. 603–623., 2003.

[18] A. B. V.Hozjan,Z.Moreno, "Nogueiras, Interfacedatabases: design and collectiono famultiling ualemotional speech database," *Proceedingsof the 3rdInternationalConferenceon Language Resources and Evaluation (LREC'02) Las Palmasde Gran Canaria, Spain*, p. pp.2019–2023., 2002.

[19] C. Breazeal and L. Aryananda, "Recognition of affective communicative intent in robot-directed speech," *Auton. Robots*, vol. 12, no. 1, pp. 83–104, 2002.

[20] G. M. M.Slaney, "Babyears:are cognition system for affective vocalizations," *SpeechCommun.*, vol. 39, pp. 367–384, 2003.

[21] Z. Zhang, E. Coutinho, J. Deng, and B. Schuller, "Cooperative Learning and its Application to Emotion Recognition from Speech," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 1, pp. 1–1, 2014.

[22] C. N. Anagnostopoulos, T. Iliou, and I. Giannoukos, "Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011," *Artif. Intell. Rev.*, vol. 43, no. 2, pp. 155–177, 2012.

[23] "SPEAKER INDEPENDENT SPEECH EMOTION RECOGNITION BY ENSEMBLE CLASSIFICATION Institute for Human-Machine Communication Technische Universität München," *Communication*, pp. 0–3, 2005.

[24] H. Hong, Z. Zhao, X. Wang, and Z. Tao, "Detection of dynamic structures of speech fundamental frequency in tonal languages," *IEEE Signal Process. Lett.*, vol. 17, no. 10, pp. 843–846, 2010.

[25] V. Sethu, E. Ambikairajah, and J. Epps, "Speaker normalisation for speech-based emotion detection," *2007 15th Int. Conf. Digit. Signal Process. DSP 2007*, pp. 611–614, 2007.

[26] J. Yadav, A. Kumari, and K. S. Rao, "Emotion recognition using LP residual at sub-segmental, segmental and supra-segmental levels," *Proc. - 2015 Int. Conf. Commun. Inf. Comput. Technol. ICCICT 2015*, vol. v, 2015.

[27] K. Wang, N. An, B. N. Li, Y. Zhang, and L. Li, "Speech emotion recognition using Fourier parameters," *IEEE Trans. Affect. Comput.*, vol. 6, no. 1, pp. 69–75, 2015.

[28] L. Fu, X. Mao, and L. Chen, "Speaker independent emotion recognition based on SVM/HMMs fusion

system," *ICALIP 2008 - 2008 Int. Conf. Audio, Lang. Image Process. Proc.*, pp. 61–65, 2008.

[29] B. Schuller, B. Vlasenko, R. Minguez, G. Rigoll, and A. Wendemuth, "COMPARING ONE AND TWO-STAGE ACOUSTIC MODELING IN THE RECOGNITION OF EMOTION IN SPEECH Institute for Human-Machine Communication , Technische Universität München , Germany Cognitive Systems , IESK , Otto-von-Guericke University , Magdeburg , Germany," pp. 596–600, 2007.

[30] M. W. Bhatti, Y. Wang, and L. Guan, "A neural network approach for human emotion recognition in speech," *Proc. Int. Symp. Circuits Syst. 2004*, vol. 2, pp. 181–184, 2004.

[31] W. M. Campbell, D. E. Sturim, and D. A. Reynolds, "Support vector machines using GMM supervectors for speaker verification," *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 308–311, 2006.

[32] Q. Yao, "Multi-sensory emotion recognition with speech and facial expression, 2014. Copyright-ProQuest." UMI Dissertations Publishing, 2014.