Bulletin of Electrical Engineering and Informatics
Vol. 8, No. 4, December 2019, pp. 1312~1323
ISSN: 2302-9285, DOI: 10.11591/eei.v8i4.1645

A critical insight into multi-languages speech emotion databases

Syed Asif Ahmad Qadri[1], Teddy Surya Gunawan[2], Muhammad Fahreza Alghifari[3], Hasmah Mansor[4], Mira Kartiwi[5], Zuriati Janin[6]
[1,2,3,4]Department of Electrical and Computer Engineering, International Islamic University Malaysia, Malaysia
[2]Visiting Fellow, School of Electrical Engineering and Telecommunications, UNSW, Australia
[5]Information Systems Department, International Islamic University Malaysia, Malaysia
[6]Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Malaysia

Article Info

Article history:

Received Mar 20, 2019
Revised May 25, 2019
Accepted Jun 23, 2019

Keywords:

Speech corpus
Speech database
Speech emotion
Speech emotion recognition

ABSTRACT

With increased interest of human-computer/human-human interactions, systems deducing and identifying emotional aspects of a speech signal has emerged as a hot research topic. Recent researches are directed towards the development of automated and intelligent analysis of human utterances. Although numerous researches have been put into place for designing systems, algorithms, classifiers in the related field; however the things are far from standardization yet. There still exists considerable amount of uncertainty with regard to aspects such as determining influencing features, better performing algorithms, number of emotion classification etc. Among the influencing factors, the uniqueness between speech databases such as data collection method is accepted to be significant among the research community. Speech emotion database is essentially a repository of varied human speech samples collected and sampled using a specified method. This paper reviews 34 `speech emotion databases for their characteristics and specifications. Furthermore critical insight into the imitational aspects for the same have also been highlighted.

Corresponding Author:

Teddy Surya Gunawan
Department of Electrical and Computer Engineering,
International Islamic University Malaysia,
Jalan Gombak, 51300 Selangor, Malaysia.
Email: tsgunawan@iium.edu.my

1. INTRODUCTION

One of the most important element in communication is the "emotion". Speech emotion recognition (SER) is presently one of the popular topics to be researched. Many systems are currently being developed to recognize different emotions from human speech. The evolution of speech emotion recognition came into existence in 1920 when a celluloid toy, 'Radio Rex' was made. Thistoy was capable of detecting acoustic energyreleased by the vowel 'Rex' i.e. 500Hz [1]. Davisin 1952 developed the first speech recognition system in Bell Laboratory, which was able torecognize digits from 0-9 in male voice. Although emotions cannot solely be understood by facial expressions, however additionally with speech emotion detection is more probable. Every human speech has their own emotion associated with it and they allow individuals to understand the feelings of the speaker. Emotions allow us to understand whether the person is happy, sad, angry etc.

When evaluating a speech emotion recognition system, one of the important factor to be considered is the level of naturalness of the database used to gauge therecognition performance. If a poorly made database is used fallacious conclusions may be established. In this paper, various SER databases are analyzed. Each respective database contains different emotion audio in various conditions.

This paper is divided into six sections. Section 2 outlines the different types of databases. Section 3 outlines the different databases commonly used in speech emotion recognition along with their respective characteristics. Section 4 showcases a table of comparison of the different attributes of the mentioned databases. Section 5 discusses some of the existing problems in emotional databases and Section 6 concludes the paper.

2. TYPES OF DATABASES

During the initial stages in the research of automatic speech recognition, there was the use of functional speech. In the recent time, the focus has been shifted towards more practical data. According to [2] the speech emotion databases are classified into 3 types:

a. **Type 1** is human docketed with acted emotion speech. The acted or simulated speech is demonstrated by a professional. They are acquired by confronting a professional to have a conversation with a predefined emotion. For example, the EMO-DB and DES.
b. **Type 2** is a genuine emotional speech with human labeling. All the emotions are factual in the Natural speech which is unforced.Data collected from call-centers are the examples of such databases.
c. **Type 3** is prompted emotional speech in which the emotions are persuaded with self report as an alternative to labeling where emotions are triggered and self-report is worn over labeling control. The prompted speech is neither simulated nor neutral.

3. SPEECH EMOTION DATABASES IN VARIOUS LANGUAGES

With regard to the speech emotion recognition, the scientific community have created and maintained various gender and linguistic repositories. In this section, 34 of these are selected and illustrated.

3.1. English speech emotion databases

a. **EDB1:** In 2005, Liscombe et al. [3] created the HMIHY speech database. This database consists of seven emotional speeches such as very frustrated, somewhat angry, very other negative, positive/neutral, somewhat other negative, very angry. The database has 5690 dialogues with 20013 user turns. These dialogues were established by an automated agent interrelated with the recordings of callers concerning calling plans, AT&T rates discussions, bill charge and balance in the account etc.
b. **EDB 2:** In 2006, Ai et al. [4] constructed the ITSPOKE corpus. The database comprises of three emotions such as positive, negative, neutral. The data was recorded from the students interacting with an interactive educating system. The database recorded 20 students with 100 dialogues and 2,252 student turns along with a set of human-human corpus and tutor plans of about 2854lines.
c. **EDB 3:** In 2006, Clavel et al. [5] created the Situation Analysis in a Fictional and Emotional (SAFE) corpus. The database consists of fear, other negative emotions, positive emotions and neutral emotions. The elicitation was carried out from audio-visual excerpts taken from movie DVDs, in abnormal contexts. The database comprises of 400 sequences with the total duration of 7 hours, having 4724 segments of speech (up to 80s).
d. **EDB 4:** In 2006, Devillers et al. [6] constructed the Castaway database. The database has wide range of emotions. The database contains Audio-Visual recordings of a reality TV show, having 10 recordings from 10 speakers, 30 minutes each.
e. **EDB 5:** In 2006, Lee et al. [7] recordeda speech database. The database consists of 4 emotions viz. angry, happy, sad and neutral. The database consists of 80 utterances (four sentences, five repetitions, four emotions). The elicitation was by recording the magnetic resonance images of a male speaker uttering a set of 4 sentences.
f. **EDB 6:** In 2006, Kumar et al. [8] developed an emotional speech corpus database. This database includes 4 emotion speeches such as neutral, lack of clarity, inappropriateness and uncertainty. It consists of 257 utterances. The creation of the database was from the interaction of 17 speakers (7 males,10 females) with the Spoken Language Dialogue Systems (SLDS) based on a customer enquiring about the stores of groceries along with answering the questions.
g. **EDB 7:** In 2006, Neiberg et al. [9] recorded the ISL Meeting corpus. The database consists of 3 speech emotions viz. negative, positive and neutral. The elicitation of the database was roughly 35 minutes complimented by orthographic transcription along with the recordings of 92 individuals in 18 meetings. The database comprises of 12068 utterances (424 negative, 2073 positive and 9571 neutral).
h. **EDB 8:** In 2006, Saratxaga et al. [10] created an emotional speech database for corpus based synthesis. The database consists of 7 emotion speeches such as neutral, happiness, sadness, fear, anger, surprise, digest. The elicitation of this database was by recording 2 participants studying the text. Two speakers

were hired (one male and one female) and 702 sentences were recorded per emotion. The recording duration is 20 hours.

3.2. German speech emotion databases

a. **GDB 1:** In 2008, Grimm et al. [11] formulated the natural database. This database consists of three emotions, for each emotional dimensions are recorded viz. activation (calm-excited), valence (positive-negative) and dominance (weak-strong). For this database 104 native speakers (44 males and 60 females) were recorded. The purpose and aim of this database is to record 12 hour of audio-visual recording using TV talk show Vera am Mittang in German.
b. **GDB 2:** In 2005, Burkhadt et al. [12] developedthe Berlin Database of Emotional speech (EMO-DB). This database consists of more than 800 utterances from 10 people (5 females and 5 males). This database contains 7 emotion speeches,viz. anger, boredom, sadness, fear, joy, disgust and neutral.
c. **GDB 3:** In 2005, Schuller et al. [13] created the EMO-SI. The database consists of 7 types of emotion, viz. anger, joy, fear, disgust, boredom, sadness and neutrality. The database included 39 speakers (3 female and 36 male) with 2730 samples (70 samples per speaker). The elicitation of the database was from spontaneous and recorded emotions in short phrases, taken from dialogues in a car.
d. **GDB 4:** In 2006, Kim and Andre [14] formulated an emotional database. The database consists of two emotions, the first being high arousal which includes negative valence and positive valence and the second being low arousal including negative valence and positive valence. It was recorded from players playing panel games (the dataset includes biosensor data). Three speakers are involved and the length of the recording is 45 min per speaker.

3.3. Japanese speech emotion databases

a. **JDB 1:** In 2004, Hirose et al. [15] developed a prosodic corpus. The database includes 4 emotion speeches viz. anger, calm, joy and sadness. This database consists of recordings of female speakersreading the sentences with every possible emotion. The database consists of approximately 1600 utterances (400 sentences per emotion).
b. **JDB 2:** In 2004,Iwai at al. [16] created an emotional speech database. The database consists of 4 emotion viz. neutral, anger, sadness and joy. The elicitation of this database was from students uttering the word "okaasan" which means "mother" in Japanese, recorded in four emotions. This database has three male speakers with 766 utterances.
c. **JDB 3:** In 2005, Takahash et al. [17] constructedan emotional speech database. This database has 5 types of emotion viz. bright, neutral, angry, excited and rage. The database consists of eight speakers with 1500 meaningful speech sounds, acted by professional actors.
d. **JDB 4:** In 2006, Nisimura et al. [18] developed an emotional speech database. This database has 16 emotion speeches viz. tiredness, surprise, depression, joy, displeasure, pleasure, excitement, anger, sadness, contentment, pressure, tension, contempt, mirth, fear and neutral. The database has recordings of children's voices (2699 utterances) extracted from a general spoken dialogue system.

3.4. Chinese emotion speech database

a. **CDB 1:** In 2006, Tao et al. [19]created anacted speech corpus. This database consists of 5 emotion speeches viz. sadness, fare, anger, happiness and neutral. It contains of 1500 utterances and 3649 phrases. A female professional actress was recorded from a Reader's Digest collection.
b. **CDB 2:** In 2006, Wu et al. [20] formulated an emotional speech corpus. It consists of 5 types of emotion speeches such as anger, fear, sadness, neutral and happiness. It included 150 short passages of about 30-50 seconds and 5000 command phases of 2-10 seconds. 50 speakers (25 females and 25 males) were recorded, where 25 actors uttered command phrases and short passages with emotional content.
c. **CDB 3:** In 2006 Zhang et al. [21] formulated a speech emotion database. The database consists of 5 types of emotion speeches viz. anger, fear, joy, sadness and neutral. The 8 speakers (4 males and 4 females) were recorded in acoustically isolated room. It consists of 2400 sentences where 20 sentences were uttered 3 times each for every emotion.

3.5. Hindi emotion speech database

a. **HDB 1:** In 2011, Rao and Koolagudi [22] formulated a native and simulated database. It consists of 8 types of emotion speeches viz. sadness, happy, disgust, neutral, surprise, fear, anger and sarcastic. For the dialect identification, Hindi dialect speech corpus was used for which 5 males and 5 females uttered sentences based on their past memories. For speech emotion recognition IIT KGP-SEHSC corpus was used, it includes 10 professional artists from All India radio Varanasi, India. It consists of 1200 utterances and each emotion has 1500 sentences.

b. **HDB 2:** In 2012, Koolagudi et al. [23] formulated a semi-natural database. It consists of 4 types of emotion speechesviz. sad, angry, happy and neutral. This database proposed a semi-natural database for emotion recognition with Graphic Era University semi-natural speech emotion corpus. The recordings were delivered by Hindi film actors and actresses from where their utterances were taken for the database.

3.6. Dutch emotion speech databases

DDB 1: In 2006, Wilting et al. [24] carried out an experiment at Tilburg University. This database consists of 5 types of emotion speeches such as negative, positive, acted negative, acted positive and neutral. This database recorded 50 speakers (19 males, 31 females) reading 40 sentences each of 20 second. The elicitation of this database was recordings of users reading sentences in various emotional states as per the procedure proposed by Velten of mood induction in the year 1968.

3.7. Korean emotion speech database

KDB 1: In 2005, Kim et al. [25] framed a database at Media & Communication Signal processing laboratory in association with Prof. C.Y Lee of Yonsei University. The database consists of 4 emotions speech such as angry, sad, joyful and neutral. 5400 sentences were recorded from 10 speakers (5 male and 5 female), reading 45 dialogic sentences by expressing native emotions and feasible pronunciation, with 3 repetitions.

3.8. Assamese Emotion speech database

ADB 1: In 2008, Kandali et al. [26] constructed a simulated database. This database has 7 types of emotion speeches namely, surprise, happiness, disgust, sadness, neutral, fear and anger. The elicitation of this database is vocal emotion recognition. A total of 140 simulated utterances for 5 native language of Assam was recorded. For the recording 30 participants (3 female and 3 male per language) specifically faculty members and students from different institutes were chosen

3.9. Sweden emotion speech database

SDB 1: In 2006,VP, Neiberg et al. [9] framed the Voice Provider Material database. The database consists of 3 types of emotion speeches viz. emphatic, negative and neutral. It includes 7619 utterances (160 emphatic, 335 negative and 7124 neutral). The purpose this database was to record voice controlled services taking in account traffic information, postal assistance etc.

3.10. Persian emotion speech database

PDB 1: In 2014, Esmaileyan and Marvi [27] formulated a simulated database. This database consists of 8 emotion speeches viz. anger, boredom, disgust, fear, neutral, sadness, surprise and happiness. It was aimed with Automatic Persian speech emotion recognition database's design in mind. 33 native speakers (15 females and 18 males) of Persian language recorded 748 utterances. Persian Drama Radio Emotional Corpus (PDREC) contains emotional utterances derives from programs on radio.

3.11. Oriya emotion speech database

ODB 1: In 2010, Mohanty and Swain [28] constructed an elicited database, consisting of six types of emotion speeches such astonish, happiness, sadness, anger, fear and neutral. The elicitation of this database was for the development of Oriya database a**n**d emotion recognition from Oriya speech. Oriya drama scripts were used from where the text fragments were recorded from 35 speakers (23 male and 12 female).

3.12. Italian emotion speech database

IDB 1: In 2014, Mencattini et al. [29] developed a simulated database. The database consists of 7 types of emotion speeches viz. sadness, joy, surprise, disgust, anger, fear and neutral. The decoction of database was to introduce PLS regression model and also enhanced speech features related to speech amplitude modulation attributes. EMOVO-the Italian speech corpus, contains 588 recordings. 6 professional actors (3 female and 3 male) were recorded in 14 Italian sentences.

3.13. Multilingual emotion speech database

a. **MLDB 1:** In 2014, Ooi et al. [30] formulated a simulated multilingual database consisting 7 different languages viz. English, Italian, German, Punjabi, Urdu, Mandarin and Persian . The database consists of 3 different categories of emotion speeches. First emotion speech consists happiness, fear, neutral, sadness, disgust and boredom; Second consists of sad, disgust, fear, surprise, angry and happy and third

consists of happiness, sadness, surprise, anger, neutral, fear and disgust. The elicitation of this database was analysis of different prosodic and spectral features was carried out. The introduction of a new architecture of intelligent audio emotion recognition is carried out. RML (audio-visual emotional database) includes8 subjects with 720 videos. eNTERFACE'05 audio-visual database consists of 42 subjects (8 female and 34 male) chosen from various nations recorded 1170 utterances. A total of 840 utterances were recorded from 10 speakers (5 female and 5 male) was included in EMO-DB (Berlin emotional database).

b. **MLDB 2:** In 2015, Kadiri et al.[31] framed a semi natural and simulated multilingual database, consisting of Telugu and German language. The database consists of 4 types of emotion speeches viz. angry, happy, neutral and sad. The decoction of the database was to analyze speech emotion recognition via excitation source feature. A total of 535 utterances were recorded from which 339 utterances were recorded for final experiment. EMO-DB Berlin emotion database consists of 10 professional native German actors (5 male, 5 female), recorded 10 sentences in 10 different emotions. IIT-H Telugu emotion database consists of 7 students (5 males, 2 females), were supposed to record the utterances based on past experiences. A total of 200 utterances were recorded for the evaluation.
c. **MLDB 3:** In 2016, Song et al. [32] developed a simulated multilingual database consisting of German and English. It consists of two categories of emotion speeches. The one characterized by disgust, fear, sadness, anger, happiness, neutral and boredom. The other by happiness, surprise, disgust, sadness, fear and anger. Two experiments are carried for this database. The approach of this database is a novel transfer non-negative matrix factorization (TNFM) presented for cross-corpus speech emotion recognition. A total of 1170 video samples were collected from allotting 42 speakers (34 males and 8 females) for recording and was included eNTERFACE (audio-visual database). Berlin dataset: a total of 494 utterances were used, recorded from 10 actors (5 female and 5 male) in German language.
d. **MLDB 4:** In 2016,Brester et al. [33] framed a simulated and natural multilingual database consisting of three languages viz. German, English and Japanese. The database consists of 4 categories of emotion speech. The first being joy, fear, anger, disgust, boredom, neutral and sadness; The second includes disgust, fear, sadness, neutral, anger, happiness and surprise; Third comprises of very angry, slightly angry, neutral, non-speech (critically noisy recordings and even silence), friendly, angry; And the forth comprises of relaxed-serene, happy-exciting, sad-bored and angry-anxious. The elicitation of this database was to select processes for evolutionary feature depending on the criterion optimization model. It consists of 4 emotional databases viz. EMO-DB (German database) recorded at Berlin Institute of Technology. It contains labelled emotional German utterances recorded from 10 professionals; SAVEE (Surrey audio-visual expressed emotion) corpus. It consists of 4 male native English speakers; LEGO emotion database. It consists of non-acted American English utterances excerpted from an automated bus information system of the Carnegie Mellon University at Pittsburgh USA; UUDB (The Utsunomiya University spoken dialogue database for paralinguistic information studies) (Japanese) consists of spontaneous human-human speech.
e. **MLDB 5:** In 2017, Pravena and Govind [34] formulatedmultilingual simulated database, consisting of Indian English, Tamil and Malayalam. The database contains of three emotion speeches such as sad, angry and happy. The elicitation of the database is the establishment of the simulated emotion database for excitation source analysis. Emotionally biased utterances were recorded from 10 speakers.

3.14. No specific language database

DB 1: In 2006, Matsunaga et al. [35]developed the database known as Corpus of infants cries. Anger, sleepiness, hunger, sadness, pampered are 5 types of emotion speeches of this database. In this database, 402 cries were recorded from infants' mothers at home making the usage of digital recorder. The facial expressions behaviour were judged by mothers of 23 infants (12 females and 11 males; 8-13months old). The raking was done from 0-4 on the basis of emotional intensity (containing no emotion to fully emotion).

4. COMPARISON BETWEEN DATABASES

The prime intension of the paper in-hand is to demonstrate the importance of the emotional speech repositories, for the overall purpose of speech signal synthesis and emotion recognition. The collected 34 databases have been described briefly and compared with regard to parameters such as cumulative size, language, emotion attributes and types of situations as shown in Table 1.

It can be seen that there is a performance gap between machine speech recognition and humans. Although speech recognition technology has made drastic progress over 30 years, despite this progress

various fundamental and practical limitations in technology have curbed its predominant use.Some of the loopholes of emotional speech databases are briefly mentioned [36, 37]:

a. It is not comprehensible which features of speech are most substantial in differentiating between emotions. The existence of different speaking styles, speakers, various sentences and rates of speaking introduces acoustic variability which adds more hurdles as these things have a direct effect on some of the speech features like tone, expression and energy.
b. It wholly depends on the speaker, his/her environment and culture, that how he expresses certain emotions, making an supposition that cultural difference does not exist in between speakers. For this reason most of the work is focused on classification of monolingual emotion.
c. Another challenging problem is one may experience some emotional state as disgust for some days or even weeks. In this case, other emotions can be brief and would last for few minutes. Due t which it is not appropriately clear for the automatic emotion recognizer to detect any specific emotion i.e. the long term emotion or the short term emotion.
d. Speech recognition by machine is a very complex problem, as speech is distorted by a background noise and echoes, electrical characteristics.
e. Due to language constraints, continuous or discontinuous isolated speeches and vocabulary size, the accuracy of speech recognition may be varied.
f. Security is another big concern in SER. Speech recovery can become a means of theft or attack. Attackers may be able to gain across of personal information, private messages, documents etc. They may also be able to impersonate the user to send messages.

Tabel 1. SER database comparison

No.	Database	Language	Emotion	Purpose And Approach	Size
1.	2005 Liscombe et al.[3]	English	Very frustrated, somewhat frustrated, very other negative, somewhat other negative, positive-neutral, very angry, somewhat angry.	Recordings of customers interacting with a automated agent concerning about the balance, enquiry of calling plans, bill charges and AT&T rates.	20013 user turns, 5690 dialogues.
2.	2006 Ai et al.[4]	English	Negative, neutral, positive.	Interaction of students with inter relative education system.	20 students, 2252 students turns, 2854 tutor turns plus human-human corpus, 100 dialogues.
3.	2006 Clavel et al.[5]	English	Positive emotions, neutral , other negative emotions, fear	Audio-visual fragments extracted abnormal contexts and movie DVDs.	4724 segments of speech (up to 80 sec), 400 sequences, total length 7 hours.
4.	2006 Devillers et al.[6]	English	Vast range of emotions.	Audio and video recordings from reality TV show.	10 speakers, 30 recordings, 30 min each.
5.	2006 Lee et al.[7]	English	Happy, angry, sad neutral.	A speaker (male) uttering a set of four sentences from recordings and magnetic resource images.	80 utterances (4 sentences, 5 repetitions and 4 emotions) and 1 male speaker.
6.	2006 Kumar et al.[8]	English	Lack of clarity, uncertainty, inappropriateness, neutral.	Recordings of the speakers interacting with SLDS enquiring about the stores of groceries and answering questions in terms of customer survey.	17 participants (10 females and 7 males), 257 utterances.
7.	2006 Neiberg et al.[9]	English	Positive, neutral, negative.	duration of 35 minute followed by orthographic transcription and recordings of 18 meetings.	12068 utterances(424 negative, 2073 positive and 9571 neutral), 92 speakers.
8.	2006 Saratxaga et al.[10]	Basque	Fear, anger, sadness, surprise, neutral, disgust, happiness.	Recording of 2 speakers uttering the text.	702 sentences(20 hr long recording).
9.	2008 Grimm et al.[11]	German	Three categories of emotion:-Dominance (weak-strong);Activation (calm-activation);Valence (positive-negative);	12 hour of audio-visual recording is done using TV show VERA am Mittang in German.	104 for native speakers (44 males and 60 females).
10.	2005 Burkhardt et al.[12]	German	Disgust, joy, anger, boredom, neutral, sadness, fear.	Recordings of speakers expressing sentences with no emotion content in every emotion.	10 speakers (5 male and 5 female), more than 800 utterances.
11.	2005 Schuller et al.[13]	German	Joy, fear, surprise, disgust, neutrality, anger, sadness.	Simultaneous and framed emotions in short phrases of car inter relation dialogues.	39 speakers (3 female and 36 male), 2730 samples (70 samples per speaker).

No.	Database	Language	Emotion	Purpose And Approach	Size
12.	2006 Kim and Andre[14]	German	Low arousal (negative valence, positive valence), High arousal (negative valence, positive valence)	Recordings of participants playing a panel game (with biosensor data).	3 speakers, 45 min per speaker.
13.	2004 Hirose et al.[15]	Japanese	Joy, anger, sadness, calm.	A female participant uttering sentences with emotional content.	1600 utterances (400 sentences per emotion), 1 female speaker.
14.	2004 Awai et al.[16]	Japanese	Joy, neutral, anger, sadness.	Recordings of the students reading the word "Okaasan"(Japanese Mother) using 4 emotions.	3 male speakers, 766 utterances.
15.	2005 Takahash et al.[17]	Japanese	Excited, neutral, bright, angry, raging.	Professional actors recorded expressed speech sounds.	8 speakers, 1500 expressive speech sounds.
16.	2006 . Nisimura et al.[18]	Japanese	Pleasure, sadness, neutral, contempt, anger, excitement, fear, tiredness, pressure, mirth, depression, joy, surprise, displeasure, contentment.	Recordings of children's taken from general spoken dialogue system.	2699 utterances.
17.	2006 Tao et al.[19]	Chinese	Happiness, neutral, fear, sadness, anger.	An actress reading stanza from a Readers Digest collection.	3649 phrases, 1 speaker, 1500 utterances.
18.	2006 Wu et al.[20]	Chinese	Neutral, happiness, anger, sadness, fear.	25 actors were recorded uttering short passages with emotional content and command phrase.	50 speakers (25 males and 25 females), 150 short passages (30-50 sec), 5000 command phrases.
19.	2006 Zhang et al.[21]	Chinese	Sadness, joy, neutral, fear, anger.	8 speakers are recorded in acoustically isolated room.	2400 sentences (20 sentences, uttered 3 times each for every emotion), 8 speakers (4 females and 4 males).
20.	2011 Rao and Koolagudi[22]	Hindi	Sadness, happy, surprise, neutral, disgust, fear, sarcastic, anger.	Feature analysis, emotion recognition, dialect identification.	For Hindi dialect speech corpus for the identification of dialect consists of 5 females and d5 males, uttered the sentences based on past experiences. IITGP-SEHSC corpus contains 10 professional artists used for speech emotion recognition from All India Radio Varanasi, India. Every emotion has 1500 sentences with 12000 recorded utterances.
21.	2012 Koolagudi et al.[23]	Hindi	Sad, anger, happy, neutral.	In Graphic Era University semi-natural speech emotion corpus, semi-natural database was proposed for emotion recognition.	Hindi film actor and actress, recorded Hindi dialogues from which utterances were taken for the database.
22.	2006 Wilting et al.[24]	Dutch	Positive, acted positive, negative, acted negative, neutral.	As per the mood induction procedure proposed in 1968 by "Velten", different emotional speeches were used by the speakers in recording sentences.	50 participants (31 female and 19 males). Each reading 40 sentences of 20 sec length.
23.	2005 Kim et al.[25]	Korean	Sad, angry, neutral, joyful.	10 speakers recording dialogic sentences by demonstrating natural emotions with easy pronunciations and subjective emotion recognition by volunteers for verification.	10 speakers (5 female and 5 male). Three repetitions, 5400 sentences, 4 emotions, 45 dialogic sentences.
24.	2008 Kandali et al.[26]	Assamese	Sadness, fear, happiness, surprise, disgust, neutral, anger.	Vocal emotion recognition.	5 native language of Assam were used for 140 simulated utterances 30 students and faculty members (3 female and 3 male per language) were chosen for recording.

No.	Database	Language	Emotion	Purpose And Approach	Size
					MESDNEI (multilingual emotional speech database of North-East India) database contains short sentences of 6 basic emotions with neutral.
25.	2014 Esmaileyan and Marvi[27]	Persian	Boredom, neutral, happiness, anger, sadness, surprise, fear, disgust.	For automatic Persian speech emotion recognition, design of database was formulated.	748 utterances recorded from 33 native speakers of Persian language (15 females and 18 males). Emotional utterances are contained in Persian Drama Radio Emotional Corpus (PDREC) taken from various radio programs.
26.	2010 Mohanty and Swain[28]	Orissa language	Astonish, happiness, fare, neutral, anger, sadness.	Emotion recognition and development of Odiya speech and database respectivily.	35 speakers (23 males and 12 females) reading the fragmented text from various Oriya drama scripts.
27.	2014 Mencattini et al.[29]	Italian	Surprise, joy, disgust, anger, fear, sadness, neutral.	PLS regression and enhanced speech features related to speech amplitude modulation parameters were introduced and discussed.	EMOVO Italian speech corpus, 588 recordings, 6 professional actors (3 females and 3 males) recorded 14 Italian sentences.
28.	2014 Ooi et al.[30]	English, Punjabi, Italian, German, Mandarin, Persian, Urdu.	3 categories of emotion: Happiness, sadness, boredom, disgust, fear, anger, neutral. Surprise, angry, happy, fear, sad, disgust. Anger, disgust, sad, neutral, surprise, fear.	A new architecture of intelligent audio emotion recognition is introduced and analyzation of various prosodic and spectral features were done.	RML (audio-visual emotion database) recorded 8 subjects for 270 videos. EMO-DB (Berlin emotion database) consists of 10 speakers (5 male and 5 female), total 840 recorded utterances were used and 10 sentences were chosen for recording. eNTERFACE'05 audio-visual emotion database, consists of 42 speakers (34 male and 8 female) chosen from different nations with 1170 utterances.
29.	2015 Kadiri et al.[31]	German and Telugu.	Anger, happy, neutral, sad.	For the analysis of speech emotion recognition excitation source feature is used.	For the recording process, 7 students were involved (2 females and 5 males) and utterances were based on past memories.200 utterances were recorded for this experiment (IIT-H Telugu emotion database) EMO-DB Berlin emotion database contains 10 professional native German actors (5 males and 5 females) were recorded in different emotions in 10 sentences. Total 535 utterances were recorded out of which 339 utterances were used for final experiment.
30.	2016 Song et al.[32]	English and German.	Two categories of emotions in this database. 1. Happiness, disgust, sad, fear, boredom, anger, neutral. 2. Surprise, disgust, anger, sad, fear, happiness.	For the presentation of cross-corpus speech emotion recognition, a novel transfer non-negative matrix factorization (TNFM) is used.	Berlin dataset: 10 actors (5 males and 5 females) in German language recorded emotional utterances with total 494 utterances being used. eNTERFACE (audio-

No.	Database	Language	Emotion	Purpose And Approach	Size
					visual database): total 1170 samples were collected, 42 speakers were recorded (34 male and 8 female).
31.	2016 Brester et al.[33]	Japanese, German, English.	4 categories of emotion: Angry, very angry, neutral, slightly angry, friendly, non-speech (critical nosy recordings or silence) Boredom or disgust, joy, sad, neutral, fear, anger. Relaxed-serene, angry-anxious, sad-bored, happy-exciting. Neutral, disgust, anger, surprise, sadness, happy, fear.	On two criterion, optimization model evolutionary feature selection technique came into account.	Four emotional databases: 1. EMO-DB (GERMAN database) recorded from 10 actors at Technical University of Berlin, consisting of emotional German utterances. 2. SAVEE (Surrey audio-visual expressed emotion corpus, consisting of 4 native English male speakers. 3. LEGO emotion database (English, comprises of Non-acted American English utterances taken from the automated bus information system of the Carnegie Mellon University of Pittsberg USA. 4. UUDB (The Utsunomiya University Spoken Dialouge database for paralinguistic information studies) (Japanese) consists of spontaneous human-human speech.
32.	2017 Pravena and Govind[34]	Indian English, Tamil, Malayalam.	Angry, happy, sad.	For excitation source analysis, simulated emotion database was constructed.	Emotionally biased utterances recorded from 10 speakers.
33.	2006 VP, Neiberg et al.[9]	Swedish	Neutral, emphatic, negative.	Recordings of voice controlled telephone service by postal assistance traffic information etc.	7619 utterances (160 emphatic,335 negative,7124 neutral).
34.	2006 Matsunaga et al.[35]	No specific language	Sleepiness, anger, hunger, pampered, sadness.	Using digital recorder, infants cried were recorded in their homes. The remarks on emotion was done by mothers keeping in account the facial expression and behaviour.	23 infants (11 males and 12 females; 8-13 months), 402 cries.

5. CONCLUSION

In recent years, researchers have made amends efforts in the field of speech emotion recognition. In this study, a good amount of publications were surveyed barely on the parameter-Database. This paper recapitulates the work done by various researchers between 2004 to 2017 speech emotion recognition. It is notable from the analysis of data, that the classification is a challenging task, in order to identify the correct emotion. The comparison of various databases remarks that two or three emotion classification with a broad classification of emotion based on speaking rate at the first stage and finer classification at the broad group emotions at the second stage has improved the performance, compared to single stage emotion classification. However, the drawback or problem with speech motion recognition is that the majority of databases are not capable of evaluation of speech emotion recognition. It is noticeable from the study that there is a lot of scope in the field of speech emotion recognition.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to the Malaysian Ministry of Education (MOE), through the Fundamental Research Grant Scheme, FRGS19-076-0684 and Universiti Teknologi MARA (UiTM) Shah Alam which have provided funding and facilities for the research.

REFERENCES

[1] G. N. Peerzade, R. Deshmukh, and S. D. Waghmare, *A Review Speech Emotion Recognition*. 2018, pp. 400-402.
[2] S. Ramakrishnan, "Recognition of emotion from speech: A review," in *Speech Enhancement, Modeling and recognition-algorithms and Applications*: InTech, 2012.
[3] J. Liscombe, G. Riccardi, and D. Hakkani-Tur, "Using context to improve emotion detection in spoken dialog systems," *Proceedings of Eurospeech'05*. 2005.
[4] H. Ai, D. J. Litman, K. Forbes-Riley, M. Rotaru, J. Tetreault, and A. Purandare, "Using system and user performance features to improve emotion detection in spoken tutoring dialogs," in *Ninth International Conference on Spoken Language Processing*, 2006.
[5] C. Clavel *et al.*, "Fear-type emotions of the SAFE Corpus: annotation issues," in *LREC*, 2006, pp. 1099-1104.
[6] L. Devillers, R. Cowie, J.-C. Martin, E. Douglas-Cowie, S. Abrilian, and M. McRorie, "Real life emotions in French and English TV video clips: an integrated annotation protocol combining continuous and discrete approaches," in *LREC*, 2006, pp. 1105-1110.
[7] S. Lee, E. Bresch, J. Adams, A. Kazemzadeh, and S. Narayanan, "A study of emotional speech articulation using a fast magnetic resonance imaging technique," in *Ninth International Conference on Spoken Language Processing*, 2006.
[8] R. Kumar, C. P. Rosé, and D. J. Litman, "Identification of confusion and surprise in spoken dialog using prosodic features," in *Ninth International Conference on Spoken Language Processing*, 2006.
[9] D. Neiberg, K. Elenius, and K. Laskowski, "Emotion recognition in spontaneous speech using GMMs," in *Ninth International Conference on Spoken Language Processing*, 2006.
[10] I. Saratxaga, E. Navas, I. Hernáez, and I. Luengo, "Designing and Recording an Emotional Speech Database for Corpus Based Synthesis in Basque," in *LREC*, 2006, pp. 2126-2129.
[11] M. Grimm, K. Kroschel and S. Narayanan, "The Vera am Mittag German audio-visual emotional speech database," *2008 IEEE International Conference on Multimedia and Expo*, Hannover, 2008, pp. 865-868.
[12] F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, and B. Weiss, "A database of German emotional speech," in *Ninth European Conference on Speech Communication and Technology*, 2005.
[13] B. Schuller, R. Müller, M. Lang, and G. Rigoll, "Speaker independent emotion recognition by early fusion of acoustic and linguistic features within ensembles," in *Ninth European Conference on Speech Communication and Technology*, 2005.
[14] J. Kim and E. André, "Emotion recognition using physiological and speech signal in short-term observation," in *International Tutorial and Research Workshop on Perception and Interactive Technologies for Speech-Based Systems*, 2006, pp. 53-64: Springer.
[15] K. Hirose, "Improvement in corpus-based generation of F0 contours using generation process model for emotional speech synthesis," in *Eighth International Conference on Spoken Language Processing*, 2004.
[16] A. Iwai, Y. Yano, and S. Okuma, "Complex emotion recognition system for a specific user using SOM based on prosodic features," in *Eighth International Conference on Spoken Language Processing*, 2004.
[17] T. Takahashi, T. Fujii, M. Nishi, H. Banno, T. Irino, and H. Kawahara, "Voice and emotional expression transformation based on statistics of vowel parameters in an emotional speech database," in *Ninth European Conference on Speech Communication and Technology*, 2005.
[18] R. Nisimura, S. Omae, H. Kawahara, and T. Irino, "Analyzing dialogue data for real-world emotional speech classification," in *Ninth International Conference on Spoken Language Processing*, 2006.
[19] J. Tao, Y. Kang, A. J. I. T. o. A. Li, Speech,, and L. Processing, "Prosody conversion from neutral speech to emotional speech," vol. 14, no. 4, pp. 1145-1154, 2006.
[20] W. Wu, T. F. Zheng, M.-X. Xu, and H.-J. Bao, "Study on speaker verification on emotional speech," in *Ninth International Conference on Spoken Language Processing*, 2006.
[21] S. Zhang, P. Ching, and F. Kong, "Automatic Emotion Recognition of Speech Signal in Mandarin," in *Ninth International Conference on Spoken Language Processing*, 2006.
[22] K. S. Rao, S. G. J. I. I. J. o. S. Koolagudi, "Identification of Hindi dialects and emotions using spectral and prosodic features of speech," *Cybernetics, and Informatics*, vol. 9, no. 4, pp. 24-33, 2011.
[23] S. G. Koolagudi and S. R. J. I. J. o. S. T. Krothapalli, "Emotion recognition from speech using sub-syllabic and pitch synchronous spectral features," *International Journal of Speech Technology*. vol. 15, no. 4, pp. 495-511, 2012.
[24] J. Wilting, E. Krahmer, and M. Swerts, "Real vs. acted emotional speech," in *Ninth International Conference on Spoken Language Processing*, 2006.
[25] Eun Ho Kim, Kyung Hak Hyun and Yoon Keun Kwak, "Robust emotion recognition feature, frequency range of meaningful signal," *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, Nashville, TN, USA, 2005, pp. 667-671.
[26] A. B. Kandali, A. Routray and T. K. Basu, "Emotion recognition from Assamese speeches using MFCC features and GMM classifier," *TENCON 2008 - 2008 IEEE Region 10 Conference*, Hyderabad, 2008, pp. 1-5.

[27] Z. Esmaileyan and H. J. I. J. o. E. Marvi, "A database for automatic Persian speech emotion recognition: collection, processing and evaluation," *International Journal of Engineering, Transactions B: Applications*. vol. 27, no. 2013, pp. 79-90, 2014.
[28] S. Mohanty and B. K. Swain, "Emotion recognition using fuzzy K-means from Oriya speech," in *2010 for International Conference [ACCTA-2010]*, 2010, pp. 3-5.
[29] A. Mencattini *et al.*, "Speech emotion recognition using amplitude modulation parameters and a combined feature selection procedure," *Knowledge-Based Systems*. vol. 63, pp. 68-81, 2014.
[30] C. S. Ooi, K. P. Seng, L.-M. Ang, and L. W. J. E. s. w. a. Chew, "A new approach of audio emotion recognition," *Expert Systems with Applications*. vol. 41, no. 13, pp. 5858-5869, 2014.
[31] S. R. Kadiri, P. Gangamohan, S. V. Gangashetty, and B. Yegnanarayana, "Analysis of excitation source features of speech for emotion recognition," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.
[32] P. Song, S. Ou, W. Zheng, Y. Jin and L. Zhao, "Speech emotion recognition using transfer non-negative matrix factorization," *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, 2016, pp. 5180-5184.
[33] C. Brester, E. Semenkin, M. J. J. o. A. I. Sidorov, and S. C. Research, "Multi-objective heuristic feature selection for speech-based multilingual emotion recognition," *Journal of Artificial Intelligence and Soft Computing Research.* vol. 6, no. 4, pp. 243-253, 2016.
[34] D. Pravena and D. J. I. J. o. S. T. Govind, "Development of simulated emotion speech database for excitation source analysis," *International Journal of Speech Technology.* vol. 20, no. 2, pp. 327-338, 2017.
[35] S. Matsunaga, S. Sakaguchi, M. Yamashita, S. Miyahara, S. Nishitani, and K. Shinohara, "Emotion detection in infants' cries based on a maximum likelihood approach," in *Ninth International Conference on Spoken Language Processing*, 2006.
[36] N. R. Kanth and S. Saraswathi, "A Survey on Speech Emotion Recognition". *Advances in Computer Science and Information Technology (ACSIT).* Volume 1, Number 3, pp. 135-139, November, 2014
[37] M. F. Alghifari, T.S. Gunawan, and M. Kartiwi, "Speech emotion recognition using deep feedforward neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, 10 (2). pp. 554-561, 2018.

BIOGRAPHIES OF AUTHORS

Syed Asif Ahmad Qadri completed his Bachelor of Technology in Computer Science and Engineering, from Baba Ghulam Shah Badshah University, Kashmir. Currently, he is working as a Research Assistant with the Department of Elec. & Comp. Engineering in International Islamic University Malaysia IIUM. His research interests include Speech processing, Artificial Intelligence, Information security, IoT etc. Currently, he is working on the application of speech emotion recognition using Deep Neural Networks.

Teddy Surya Gunawan received his BEng degree in Electrical Engineering with cum laude award from Institut Teknologi Bandung (ITB), Indonesia in 1998. He obtained his M.Eng degree in 2001 from the School of Computer Engineering at Nanyang Technological University, Singapore, and PhD degree in 2007 from the School of Electrical Engineering and Telecommunications, The University of New South Wales, Australia. His research interests are in speech and audio processing, biomedical signal processing and instrumentation, image and video processing, and parallel computing. He is currently an IEEE Senior Member (since 2012), was chairman of IEEE Instrumentation and Measurement Society – Malaysia Section (2013 and 2014), Associate Professor (since 2012), Head of Department (2015-2016) at Department of Electrical and Computer Engineering, and Head of Programme Accreditation and Quality Assurance for Faculty of Engineering (2017-2018), International Islamic University Malaysia. He is a Chartered Engineer (IET, UK) and Insinyur Profesional Madya (PII, Indonesia) since 2016, and registered ASEAN engineer since 2018.

Muhammad Fahreza Alghifari has completed his B.Eng. (Hons) degree in Electronics: Computer Information Engineering from International Islamic University Malaysia (IIUM) in 2018 and is currently pursuing his Masters in Computer Engineering while working as a research assistant. His research interests are in signal processing, artificial intelligence and affective computing. He received a best FYP award from IEEE Signal Processing – Malaysia chapter and achieved recognition in several national level competitions such as Alliance Bank EcoBiz Competition and IMDC2018.

Hasmah Mansorgraduated from the University of Salford, United Kingdom in Electronic & Electrical Engineering, started her career as a Research & Development Engineer at Digital Aura (M) Sdn. Bhd. in January 2000. She was one of the pioneers who was directly involved in the design and development of the first Malaysian Programmable Logic Controller (PLC) named DA2000. She is currently an Associate Professor and Deputy Dean of Student Affairs at Kulliyyah of Engineering, International Islamic University Malaysia. She is a Chartered Engineer (IET, UK) since 2018 and an IEEE Senior Member.

Mira Kartiwi completed her studies at the University of Wollongong, Australia resulting in the following degrees being conferred: Bachelor of Commerce in Business Information Systems, Master in Information Systems in 2001 and her Doctor of Philosophy in 2009. She is currently an Associate Professor in Department of Information Systems, Kulliyyah of Information and Communication Technology, and Deputy Director of e-learning at Centre for Professional Development, International Islamic University Malaysia. Her research interests include electronic commerce, data mining, e-health and mobile applications development.

Zuriati Janin received her B.Eng in Electrical Engineering from the Universiti Teknologi Mara, Malaysia in 1996 and MSc. in Remote Sensing & GIS from the Universiti Putra Malaysia (UPM) in 2001. In 2007, she began her study towards a Ph.D in Instrumentation and Control System at the Universiti Teknologi Mara, Malaysia. She has served as a lecturer at Universiti Teknologi Mara for more than 20 years and currently she is a Senior Lecturer at Faculty of Electrical Engineering, UiTM, Shah Alam. She has been involved with IEEE since 2012 and been mainly working with IEEE Instrumentation & Measurement Chapter (IM09), Malaysia Section since 2013. The IM09 acknowledged her role as a founder Treasurer in initiating and promoting ICSIMA as a series of annual chapter's flagship conferences since its inception in 2013. She also has more than 10 years experiences in organizing the International Conferences, Workshops and Seminars. Her role as a conference treasurer started since 2005.