D. Ververidis and C. Kotropoulos, "A Review of Emotional Speech Databases," in Proc. 9th Panhellenic Conference on informatics (PCI), pp. 560-574, Thessaloniki, Greece, November 2003.

A Review of Emotional Speech Databases

Dimitrios Ververidis and Constantine Kotropoulos

Artificial Intelligence & Information Analysis Laboratory,
Department of Informatics
Aristotle University of Thessaloniki,
Box 451, Thessaloniki 541 24, Greece.
E-mail: costas@zeus.csd.auth.gr
Tel: ++30-2310-996361, Fax: +30-2310-998453.

Abstract. Thirty-two emotional speech databases are reviewed. Each database consists of a corpus of human speech pronounced under different emotional conditions. A basic description of each database and its applications is provided. The conclusion of this study is that automated emotion recognition cannot achieve a correct classification that exceeds 50% for the four basic emotions, i.e., twice as much as random selection. Second, natural emotions cannot be easily classified as simulated ones (i.e., acting) can be. Third, the most common emotions searched for in decreasing frequency of appearance are anger, sadness, happiness, fear, disgust, joy, surprise, and boredom.

Keywords: Speech Recognition, Speech Databases, Emotion Recognition.
Topics: Interfaces (Emotions and personality), Situation Awareness Applications (Training).

1 Introduction

Emotion is an important factor in communication. For example, a simple text dictation that doesn 't reveal any emotion, it does not covey adequately the semantics of the text. An emotion speech synthesizer could solve such a communication problem. Speech emotion recognition systems can be used by disabled people for communication, by actors for emotion speech consistency as well as for interactive TV, for constructing virtual teachers, in the study of human brain malfunctions, and the advanced design of speech coders. Until recently many voice synthesizers could not produce faithfully a human emotional speech. This results to an unnatural and unattractive speech. Nowadays, the major speech processing labs worldwide are trying to develop efficient algorithms for emotion speech synthesis as well as emotion speech recognition. To achieve such ambitious goals, the collection of emotional speech databases is a prerequisite. In this paper, thirty-two emotional speech databases are reviewed. In Section 2, a brief description of each database is provided. In Section 3, a discussion of the reviewed databases features is made. Finally, conclusions are drawn in Section 4.

2 Description of Speech Emotion Databases

This section briefly describes the thirty-two databases included in our comparative study per language used. Table **??** in the Appendix provides a complete listing of the databases and their features.

2.1 English speech emotion databases

Database 1. The database was recorded at the Faculty of Electrical Engineering and Computer Science, University of Maribor, Slovenia [1]. It contains emotional speech in six emotion categories, such as disgust, surprise, joy, fear, anger and sadness. Two neutral emotions were also included: fast loud and low soft. It is seen that the emotion categories are **compliant with MPEG-4** [2]. Four languages (i.e. English, Slovenian, French and Spanish) were used in all speech recordings. The database contains 186 utterances per emotion category. These utterances are divided in isolated words, sentences both affirmative and interrogative, and a passage.

Database 2. R. Cowie and E. Cowie [31], [6], [32] constructed this database at the Queen's University of Belfast. It contains emotional speech in 5 emotional states: anger, sadness, happiness, fear and neutral. The readers are 40 volunteers (20 female, 20 male) aged between 18 to 69 years. The subjects read 5 passages of 7-8 sentences written in an appropriate emotional tone and content for each emotional state. Each passage has strong relationship with the corresponded emotional state.

Database 3 Belfast Natural Database. R. Cowie and M. Schroder constructed this database at Queen's University [36]. The database is designed to sample genuine emotional states and to allow exploration of the emotions through time. Two kinds of recordings took place. One was recorded in studio and the other direct from TV programs. A total of 239 clips (10-60 sec) is included in the database. The clip length is taken to be quite long in order to reveal the development of emotion through time. The studio recordings consist of two parts. The first part contains conversations between students on topics, which provoke strong feelings. The second part contains audio-visual recordings of interviews (one-to-one) involving a researcher with fieldwork experience and a series of friends.

Database 4, Kids' Audio Speech Corpus NSF/ITR Reading Project. R. Cole and his assistants at the University of Colorado recorded database 4 [12]. The aim of the project was to collect sufficient audio and video data from kids in order to enable the development of auditory and visual recognition systems, which enable face-to-face conversational interaction with electronic teachers. The Kids' Audio speech Corpus is not clearly oriented to elicit emotions. Only 1000 out of 45000 utterances are emotion oriented.

Database 5. Emotional Prosody Speech and Transcripts. M. Liberman, Kelly Davis, and Murray Grossman at the University of Pennsylvania constructed database

5 [13]. The database consists of 9 hours of speech data. It contains speech in 15 emotional categories, such as hot anger, cold anger, panic-anxiety, despair, sadness, elation, happiness, interest, boredom, shame, pride, disgust and contempt.

Database 6 SUSAS. J. Hansen at the University of Colorado Boulder has constructed in 1999 an emotional speech corpus the so called SUSAS (Speech Under Simulated and Actual Stress). The database contains voice from 32 speakers (13 female and 19 male) with ages ranging from 22 to 76. In addition, four military helicopter pilots were recorded during the flight. Words from a vocabulary of 35 aircraft communication words make up the database. The total number of utterances is 16.000. SUSAS database is distributed by the LDC [13].

Database 7. C. Pereira at Macquire University constructed Database 7 [19]. The database consists of 40 sentences said by two actors in 5 emotional categories. There are 2 repetitions of these 40 utterances, thus creating 80 presentations. In the study, 31 normal hearing subjects (18 men and 13 women) rated all the utterances. The listeners rated each utterance on six Likert intensity scales (Mehrabian and Russel, 1974).

Database 8. M. Edgington at BT Labs, UK collected this emotional speech database for training a voice synthesizer [11]. One professional male actor was employed. The database contains speech in 6 emotional categories, such as anger, fear, sadness, boredom, happiness and neutral. A laryngograph was also recorded. Thirteen raters-judges identified the emotions with 79.3% score rate. The database also includes the signal energy, syllabic duration, and the fundamental frequency of each phoneme.

Database 9. T. S. Polzin and A. H. Waibel at the Carnegie Mellon University constructed this emotional speech database [15]. The database consists of emotional speech in 5 emotional categories. The corpus was comprised of 291 word tokens per emotion per speaker. The sentence length varies from 2 to 12 words. These sentences are comprised of questions, statements, and orders. The database was evaluated from other people. The recognition of each emotion is 70%. The baseline is 25% for random guessing.

Database 10. V. Petrushin at the Center for Strategic Technology Research, Accenture constructed this database in order to train neural networks in speech recognition [25]. It is divided into two studies. The first study deals with a corpus of 700 short utterances expressed by 30 professional actors. The corpus contains emotional speech in 5 emotion categories, such as happiness, anger, sadness, fear, and normal, which were portrayed by thirty non-professional actors. In the second study, 56 telephone messages were recorded. The length of each message is from 15 to 90 seconds.

Database 11. R. Fernandez at MIT labs constructed emotional speech Database 11 [27]. The subjects are drivers who were asked to sum up two numbers while driving a car. The questions produced by a speech synthesizer and the sum of the numbers was less than 100. The two independent variables in this experiment were the driving

speed and the frequency at which the driver had to solve the math questions. Every subject drove in two speed conditions, one at 60 m.p.h. and one at 120 m.p.h. In the low speed, the subject was asked for an answer every 9 seconds. In the high speed, the subject was asked to sum up the numbers every 4 seconds.

2.2 German speech emotion databases

Database 12 Verbmobil. The database was recorded at the University of Hamburg [38], [8]. It is partially emotion oriented, because it contains mainly anger and dissatisfaction. The database contains voice from 58 native German speakers, (29 male, 29 female aged from 19 to 61 years old), while they were speaking to a "pretended" ASR system. From distance, a researcher ("a wizard") controlled the ASR response, in such a way that the speaker believed that he was speaking to a machine. The above dialogues are named "Wisard-Of-Oz" dialogues. Prosodic properties of the emotion, such as syllable lengthening and word emphasis are also annotated. This database can be used for training speech recognition systems.

Database 13 SmartKom Multimodal Corpus. It was constructed at the Institute for Phonetik and Oral Communication in Munich, with national funding [38], [39]. The database consists of Wizard-Of-Oz dialogues (like Verbmobil database) in German and English. The database contains multiple audio channels (which follow speakers' position changes) and two video channels (face, body from side). It is partially oriented to emotion as it includes mainly anger and dissatisfaction. The aim of the project was to build a gesture and voice recognition module for human-computer interfaces.

Database 14. W. F. Sendlmeier et al. at the Technical University of Berlin constructed another emotional speech database [4], [34]. The database consists of emotional speech in seven emotion categories. Each one of the ten professional actors expresses ten words and five sentences in all the emotional categories. The corpus was evaluated by 25 judges who classified each emotion with a score rate of 80%.

Database 15. K. Alter at the Max-Planck-Institute of Cognitive Neuroscience constructed an emotional speech database for medical purposes [20]. Electroencephalogram (EEG) was also recorded. The aim of the project was to relate emotions, which are recognized from speech with a location in the human brain. A trained female fluent speaker was employed. The database contains speech in 3 emotional categories. Twenty subjects judged both the semantic content and the prosodic feature on a five-point scale.

Database 16. K. Scherer at the University of Geneva constructed this database [21]. The purpose of his study was to bring into light the differences in emotional speech perception between people from different countries. Results showed that the Indonesian people understand the emotions in speech in a different way. Four German professional actors were employed. The sentences derived from an artificial language, which was constructed by a professional phonetician.

Database 17 Magdeburger Prosodie Korpus. B. Wendt and H. Scheich at Leibniz Institute of Neurobiology constructed the Magdeburger Prosodie Korpus [9]. The aim was to construct a brain map of emotions. The database contains emotional speech in six emotional categories. The database contains also the word accentuation, word length, speed rate, abstraction/concreteness, categorizations, and phonetic minimal pairs. The total number of utterances is 4200 nouns and pseudowords.

Database 18. M. Schroeder and M. Grice recorded a database of diphones that can be used for emotional speech synthesis [16]. One male speaker of standard German produced a full German diphone set for each of three degrees of vocal effort: "soft", "modal", and "loud". Four experts verified the vocal effort, the pitch constancy, and the phonetic correctness.

Database 19. M. Schroder has constructed an emotional speech database, which consists "affect bursts" [7]. His study shows that affect bursts, presented without context, can convey a clearly identifiable emotional meaning. Professionals selected the affect bursts from the German literature. Altogether, the database comprises about 80 different affect bursts. The database contains speech in 10 emotion categories. In order to define the intended emotions for the recordings, a frame story was constructed for each of the ten emotions.

2.3 Japanese speech emotion databases

Database 20. R. Nakatsu et al. at the ATR Laboratories constructed an emotional speech database in Japanese [37]. The database contains speech in 8 emotion categories. The project employed 100 native speakers (50 male and 50 female) and one professional radio speaker. The professional speaker was told to read 100 neutral words in 8 emotional manners. The 100 ordinary speakers were asked to mimic the manner of the professional actor and say the same amount of words. The total amount of utterances is 80000 words.

Database 21. Y. Niimi et al. at the Kyoto Institute of Technology, Matsugasaki, Japan developed another emotion speech database in Japanese [14]. It consists of VCVs (vowel consonant vowel) segments for each of the three emotion speech categories, such as anger, sadness, and joy. These VCVs can generate any accent pattern of Japanese. These VCVs were collected from a corpus of 400 linguistic unbiased utterances. The utterances were analyzed to derive a guideline for designing VCV databases, and to derive an equation for each phoneme, which can predict its duration based on its surrounding phonemic and linguistic context. Twelve people judged the database and they recognized each emotion with a rate of 84%.

Database 22. A. Iida and N. Campbell at ATR Laboratories constructed a third emotional speech database in Japanese [28]. The aim of the project was emotional speech synthesis for disabled people. The database contains emotional speech in 3 emotion categories, such as joy, anger and sadness. The emotions are simulated but not exaggerated. The database consists of monologue texts collected from newspapers, the

WWW, self-published autobiographies of disabled people, essays and columns. Some expressions typical to each emotion were inserted in appropriate places in order to enhance the expression of each target emotion.

2.4 Dutch emotion speech databases

Database 23 Groningen database. It is constructed at the Psychology School at Groningen University in Netherlands and is distributed by ELRA [8]. It contains 20 hours of Dutch speech. We must state that the database is only partially oriented to emotion. An electroglottograph and an orthographic transcription are also included. The total number of speakers is 238. They are not actors and the emotions are forced rather than natural. The database consists of short texts, short sentences, digits, monosyllabic words, and long vowels.

Database 24. S. Mozziconacci and his assistant collected an emotional speech database in order to study the relationship between speed in speaking and emotion [17], [29]. The database contains emotional speech in seven emotion categories, such as joy, boredom, anger, sadness, fear, indignation, and neutral. The speech material used in the study consists of 315 utterances. Each of the three speakers reads five sentences, which have a semantically neutral content. Twenty-four judges evaluated the utterances and two intonation experts labeled.

2.5 Spanish emotion speech databases

Database 25 Spanish Emotional Speech database (SES). J. M. Montero and his assistants constructed in 1998 a Spanish emotional speech database [10]. It contains emotional speech in 4 emotion categories, such as sadness, happiness, anger, and neutral. Fifteen raters-judges identified each emotion with a score of 85%. The labeling of the database is semi-automatic. The corpus consists of 3 short passages (4-5 sentences), 15 short sentences, and 30 isolated words. All have neutral lexical, syntactical, and semantical meaning.

Database 26. I. Iriondo [26] at the University of R. L. of Barcelona also recorded another emotional speech database. It contains emotional speech in 7 emotion categories. Each of the 8 actors reads 2 texts in 3 emotional intensities. The speech was rated-judged by 1054 students during a perception test. From the 336 discourses, only 34 passed the perception test.

2.6 Danish emotion speech database

Database 27. I. F. Engberg and A.V. Hansen at the Center for Person Kommunication at Aalbord University recorded the Danish Emotional Speech database (DES) [3]. T. Brondsted wrote the phonetical transcription. The construction of DES was a part of Voice Attitudes and Emotions in Speech Synthesis (VAESS) project. The database contains emotional speech in 5 emotion categories, such as surprise, happiness, anger,

sadness and neutral. The database consists of 2 words (yes, no), 9 sentences, and 2 passages. Twenty judges (native speakers from 18 to 58 year old) verified the emotions with a score rate of 67%.

2.7 Hebrew emotion speech database

Database 28. At the faculty of Holon Academic Institute of Technology at Israel, N. Amir et al. recorded emotional multi-modal speech database 28 [5]. It contains emotional speech in 5 emotion categories. The database consists of emotional speech, electromyogram of the curragator (a muscle of the upper face which assists in expressing an emotion), heart rate, and galvanic resistance that is a sweat indicator. The subjects (40 students) were told to recall an emotional situation of their life and speak about that. At his study N. Amir found that there is not an absolute clear way of discovering the true emotion in speech.

2.8 Sweden emotion speech database

Database 29. A. Abelin and his assistants recorded emotional speech database 29 [18]. It contains emotional speech in 9 emotion categories, such as joy, surprise, sadness, fear, shyness, anger, dominance, disgust, and neutral. Different nationality listeners classified the emotional utterances to an emotional state. The listener group consisted of 35 native Swedish speakers, 23 native Spanish speakers, 23 native Finnish speakers and 12 native English speakers. The non-Swedish listeners were Swedish immigrants and all had knowledge of Swedish, of varying proficiency.

2.9 Chinese emotion speech database

Database 30. F. Yu et al. at the Microsoft Research China recorded emotional speech database 30 [23]. It contains speech segments from Chinese teleplays in four emotion categories, such as anger, happiness, sadness, and neutral. Four persons tagged the 2000 utterances. Each person tagged all the utterances. When two or more persons agreed in their tag, the utterance got their tag. Elsewhere the utterance was thrown away. After tagging several times, only 721 utterances remained.

2.10 Russian emotion speech database

Database 31 RUSSian LANguage Affective speech (RUSSLANA). V. Makarova and V.A. Petrushin collected this emotional speech database at the Meikai University in Japan [40]. The total of utterances is 3660 sentences from 61 (12 male) native Russian speakers age from 16 to 28. Features of speech like energy, pitch and formants curves are also included.

2.11 Multilingual emotion speech database

Database 32 Lost Luggage study. K. Scherer has recorded another emotional speech database [24]. The recordings took place in Geneva International Airport. The subjects are 109 airline passengers waiting in vain for their luggage to arrive on the belt.

3 Discussion

3.1 Database purpose

An emotional speech database is collected for a variety of purposes. In the set of thirty databases, we found that most databases were used for automatic speech recognition and speech synthesis.

Purpose	Databases	Total number
Automatic emotion recognition with ASR applications	1, 2, 3, 5, 7, 9, 10, 11, 12, 13, 20 24, 25, 26, 28, 30, 31, 32	17
Emotion speech synthesis	1, 2, 3, 5, 8, 14, 18, 19, 21, 22, 27	10
Medical applications	5, 7, 15, 17	4
Emotion perception by human	7, 9, 13, 16, 21, 29, 32	8
Speech under stress	6, 11	2
Virtual teacher	4	1

Table 1. Database purpose.

For several databases we know the correct recognition rate of emotions by humans. For example, for Database 8 this rate is 79%, whereas for Database 9 is 70%. Slightly higher correct recognition rates were reported for Databases 13 and 20. These rates were 80% and 84%, respectively. To derive the rates, groups of judges (usually 20-30 people) were employed. As indicated, a human can recognize correctly an emotion in speech with an average rate of 80%. Thus, it is difficult to build an automated emotion recognition system, which can classify emotions more accurately than 80%. As indicated by Cowie [36] a successful automatic classifier hits a correct classification rate of 50%, when the classification of four emotions is dealt with. If the emotion categories were four, a random classification would achieve a score of 25%. That is, the highest classification rate is twice as much as a random selection. This observation was restated in Banse-Scherer [35], but for a larger number of emotions.

3.2 Language used in recordings

The English language is dominant, as 11 out of the 30 databases are in English. The German language was found to be the second most popular language. The number of databases recorded in each language is tabulated in Table **??**. The Multi-language database consists of speech recorded at the luggage belt of the Geneva Airport. Database 15 was collected by an expert in intonation who used an artificial language for this purpose. An example of a formula follows:

3.3 Most common emotions

The most common emotions, that can be found in the set of 32 databases reviewed, arranged in decreasing order of their frequency of appearance are summarized in Table

Language	Number of databases	Databases
English	11	1-11
German	7	12, 14, 15, 17-19
Japanese	3	20-22
Spanish	3	1, 25, 26
Dutch	2	23, 24
French	1	1
Sweden	1	29
Hebrew	1	28
Danish	1	27
Slovenian	1	1
Chinese	1	30
Russian	1	31
English and German	1	13
Multi-Language	1	32
Artificial Language	1	16

Table 2. Language used in database recording.

??. The most common recordings are anger, sadness, happiness, fear, disgust, surprise, boredom and joy. This is almost in agreement with Cowie and Cornelius [33], except for boredom. Boredom seems to be more popular than it was originally thought. For situation awareness, we are interested particularly in databases with speech under stress.

Emotion	Number of databases
Anger	26
Sadness	22
Happiness	13
Fear	13
Disgust	10
Joy	9
Surprise	6
Boredom	5
Stress	3
Contempt	2
Dissatisfaction	2
Shame, pride, worry, startle, elation, despair, humor,...	1

Table 3. Emotion recorded in databases.

3.4 Databases for situation awareness training

The first speech database recorded under stress was the SUSAS database (Database 6). Databases, which include stressed speech, such as SUSAS, are used for training people at difficult situations. SUSAS for example is used for emotion recognition of pilots during flight. Thus the pilot could have a feedback about his emotional state. The adverse environment could be a building on fire, automobile in traffic, helicopter or aircraft cockpits, and others. SUSAS database contains words, which potentially could be pronounced in a speech-command helicopter. The second database (Database 11) in this category contains speech of drivers who are trying to solve a mathematical problem while driving at various speeds. This study can reveal many aspects of human brain functions. Databases, which include fear as an emotional state, can be used in virtual training. For example, fear recognition from speech can be used in training kids how to deal with an earthquake. So emotion recognition from speech can be a useful tool for assisting virtual reality to cure phobias.

Language	Databases
English	1, 3, 8, 10, 9
German	14, 16, 17
Spanish	1, 26
Sweden	29
French	1
Dutch	24
Japanese	20
Hebrew	28
Slovenian	1

Table 4. Twelve speech emotion databases include fear.

3.5 Simulated or natural emotions

Natural emotions cannot be easily classified by a human [27]. Since a human cannot classify easily natural emotions, it is difficult to expect that machines can offer a higher correct classification. In order to avoid complicated situations, the majority of the databases include forced (simulated) emotional speech. Professional actors, drama students or normal people express these emotional utterances. Extravagation from the actors during speech recording is usually prohibited. Thus, the simulated emotion speech has a more realistic profile. Table **??** indicates the types of speech emotion and their frequency of occurrence.

4 Conclusions

From the comparative study of the thirty-two databases, we conclude that there is a need for establishing a protocol that will address the following issues:

Type of Emotion	Occurrences	Database Index
Simulated	20	1, 5, 7, 8, 9, 14, 15, 16, 17, 18, 31, 29 19, 20, 21, 22, 23, 24, 25, 26, 27,
Natural	7	2, 4, 12, 13, 28, 30, 11, 32
Semi-natural	1	3
Half recordings natural Half recordings simulated	2	6, 32

Table 5. Natural and simulated speech occurrences.

- Collection parameters:
  - Simulated feelings by actors or drama students or
  - natural feelings by ordinary people.
- Types of data
  - Speech
  - Video
  - Laryngograph
  - Myograms of the face, Heart Beat Rate, EEG.
- Multiplicity of data: How many recording sessions exist in the database.
- Data availiability
- Objective methods of measuring the performance of the methods employed.
- Subjective methods of conducting mean opinion scores.

The aforementioned discussion provides a minimal list of specifications for recording any future database to be used for speech emotion recognition. We recommend the data to be distributed by organizations (like LDC or ELRA) under a reasonable fee so that the experiments reported in the could be repeated. This is not the case with the majority of the databases reviewed in this paper, whose terms of distribution are unclear.

Acknowledgments

This work has been partially supported by the research project 01E312 "Use of Virtual Reality for training pupils to deal with earthquakes" financed by the Greek Secretariat of Research and Technology.

References

1. D. C. Ambrus, "Collecting and recording of an emotional speech database", Technical Report, Faculty of Electrical Engineering and Computer Science, Institute of Electronics, University of Maribor.
2. J. Ostermann, "Face animation in MPEG-4", in *MPEG-4 Facial Animation* (I. S. Pandzic and R. Forchheimer, Eds.), pp.17-56, Chichester, U.K.: J. Wiley, 2002.
3. I. S. Engberg, and A. V. Hansen, "Documentation of the Danish Emotional Speech Database (DES)," Internal AAU report, Center for Person Kommunikation, Department of Communication Technology, Institute of Electronic Systems, Aalborg University, Denmark, September 1996.

4. F. Burkhardt, and W. F. Sendlmeier, "Verification of acoustical correlates of emotional speech usingformant-synthesis", in *Proc. ISCA Workshop (ITRW) Speech and Emotion: A conceptual framework for research*, pp. 29-33, Belfast, 2000.
5. N. Amir, S. Ron, and N. Laor, "Analysis of an emotional speech corpus in Hebrew based on objective criteria", in *Proc. ISCA Workshop (ITRW) Speech and Emotion: A conceptual framework for research*, pp. 29-33, Belfast, 2000.
6. R. Cowie, E. Douglas-Cowie, S. Savvidou, E. McMahon, M. Sawey, and M. Schroder, "Feeltrace: An instrument for recording perceived emotion in real time", in *Proc. ISCA Workshop (ITRW) on Speech and Emotion: A conceptual framework for research*, pp. 19-24, Belfast, 2000.
7. M. Schroder, "Experimental study of affect bursts", in *Proc. ISCA Workshop (ITRW) Speech and Emotion: A conceptual framework for research*, pp. 132-137, Belfast. 2000.
8. European Language Resources Association, (ELRA), www.elra.info.
9. B. Wendt and H. Scheich, "The Magdeburger Prosodie-Korpus", in *Proc. Speech Prosody Conf. 2002*, pp. 699-701, Aix-en-Provence, France, 2002.
10. J. M. Montero, J. Gutierrez-Arriola, J. Colas, E. Enriquez, and J. M. Pardo, "Analysis and modelling of emotional speech in Spanish", in *Proc. ICPhS'99*, pp. 957-960, San Francisco 1999.
11. M. Edgington. "Investigating the limitations of concatenative synthesis", in *Proc. Eurospeech 97*, pp 593-596, Rhodes, Greece, September 1997.
12. The Center for Spoken Language Research (CSLR), CU Kids' speech corpus, http://cslr.colorado.edu/beginweb/reading/data_collection.html.
13. Linguistic Data Consurtium (LDC), http://www.ldc.upenn.edu/.
14. Y. Niimi, M.L. Kasamatu, T. Nishimoto, and M. Araki, "Synthesis of emotional speech using prosodically balanced VCV Segments", in *Proc. 4th ISCA tutorial and Workshop on research synthesis, Scotland*, August 2001.
15. T. S. Polzin and A. H. Waibel, "Detecting emotions in speech", in *Proc. CMC 1998.*
16. M. Schroder and M. Grice. "Expressing vocal effort in concatenative synthesis", in *Proc. 15th Int. Conf. Phonetic Sciences*, Barcelona, Spain 2003.
17. S. J. L. Mozziconacci and D. J. Hermes, "Expression of emotion and attitude through temporal speech variations", in *Proc. 2000 Int. Conf. Spoken Language Processing (ICSLP 2000)*, vol. 2, pp. 373-378, Beijing, China, 2000.
18. A. Abelin and J. Allwood, "Cross linguistic interpretation of emotional prosody", in *Proc. ISCA Workshop (ITRW) on Speech and Emotion: A conceptual framework for research*, Belfast, 2000.
19. C. Pereira, "Dimensions of emotional meaning in speech", in *Proc. ISCA Workshop on Speech and Emotion: A conceptual framework for research*, pp. 25-28, Belfast, 2000.
20. K. Alter, E. Rank, and S. A. Kotz, "Accentuation and emotions - Two different systems? ", in *Proc. ISCA Workshop on Speech and Emotion: A conceptual framework for research*, Belfast, 2000.
21. K. Scherer, "A cross-cultural investigation of emotion inferences from voice and speech: Implications for speech technology", in *Proc. 2000 Int. Conf. Spoken Language Processing (ICSLP 2000)*, Beijing, China, 2000.
22. H. Chung, "Duration models and the perceptual evaluation of spoken Korean", in *Proc. Speech Prosody 2002*, pp. 219-222, Aix-en-Provence, France, April 2002.
23. F. Yu, E. Chang, Y.Q. Xu, and H.Y. Shum, "Emotion detection from speech to enrich multimedia content", in *Proc. 2nd IEEE Pacific-Rim Conference on Multimedia 2001*, pp. 550-557, Beijing, China, October 2001.
24. K. R. Scherer. "Emotion effects on voice and speech: Paradigms and approaches to evaluation", in *Proc. ISCA Workshop (ITRW) on Speech and Emotion: A conceptual framework for research*, Belfast, 2000.

25. V. A. Petrushin. "Emotion in speech recognition and application to call centers",in *Proc. ANNIE 1999*, pp. 7-10, 1999.
26. I. Iriondo, R. Guaus, and A. Rodriguez, "Validation of an acoustical modeling of emotional expression in Spanish using speech synthesis techniques", in *Proc. ISCA Workshop (ITRW) Speech and Emotion: A conceptual framework for research*, pp. 161-166, Belfast, 2002.
27. R. Fernandez and R. W. Picard. "Modeling drivers' speech under stress", in *Proc. ISCA Workshop (ITRW) on Speech and Emotion: A conceptual framework for research*, Belfast 2002.
28. A. Iida, N. Campbell, S. Iga, F. Higuchi, and M. Yasumura, "A speech synthesis system with emotion for assisting communication", in *Proc. ISCA Workshop (ITRW) on Speech and Emotion: A conceptual framework for research*, pp. 167-172, Belfast 2002.
29. S. J. L. Mozziconacci and D. J. Hermes, "A study of intonation patterns in speech expressing emotion or attitude: production and perception", IPO Annual Progress Report 32, pp. 154-160, IPO, Eindhoven, The Netherlands, 1997.
30. R. Cowie, R. R. Cornelius, "Describing the emotional states that are expressed in speech", *Speech Communication*, vol. 40, pp. 5-32, 2003.
31. M. Schroder, R. Cowie, E. Douglas-Cowie, M.J.D. Westerdijk, and C.C.A.M. Gielen, "Acoustic correlates of emotion dimensions in view of speech synthesis", in *Proc. Eurospeech 2001*, vol. 1, pp. 87-90, Aalborg, Denmark, 2001
32. S. McGilloway, R. Cowie, E. Douglas-Cowie, and C.C.A.M. Gielen, and M.J.D. Westerdijk, and S. H. Stroeve, "Approaching automatic recognition of emotion from voice: A rough benchmark" in *Proc. ISCA Workshop Speech and Emotion*, pp. 207-212, Newcastle, 2000.
33. M. Schroder, "Emotional speech synthesis: A review", in *Proc. Eurospeech 2001*, vol. 1 pp. 561-564, Aalborg, Denmark.
34. M. Kienast and W. F. Sendlmeier, "Acoustical analysis of spectral and temporal changes in emotional speech", in *Proc. ISCA (ITWR) Workshop Speech and Emotion: A conceptual framework for research*, Belfast 2000.
35. R. Banse, and K. Scherer, "Acoustic profiles in vocal emotion expression",in *Journal of Personality and Social Psychology*, vol. 70, no. 3, pp. 614-636, 1996.
36. E. Douglas-Cowie, R. Cowie, and M. Schroder, "A New Emotion Database: Considerations, Sources and Scope", in *Proc. ISCA (ITWR) Workshop Speech and Emotion: A conceptual framework for research*, pp. 39-44, Belfast 2000.
37. R. Nakatsu, A. Solomides, and N. Tosa, "Emotion recognition and its application to computer agents with spontaneous interactive capabilities", in *Proc. IEEE Int. Conf. Multimedia Computing and Systems*, vol. 2, pp. 804-808, Florence, Italy, July 1999.
38. Bavarian Archive for Speech Signals, http://www.bas.uni-muenchen.de/Bas/.
39. F. Schiel, Silke Steininger, Ulrich Turk, "The Smartkom Multimodal Corpus at BAS", in *Proc. Language Resources and Evaluation*, Canary Islands, Spain, May 2002.
40. V. Makarova and V. A. Petrushin, "RUSLANA: A database of Russian Emotional Utterances", in *Proc. 2002 Int. Conf. Spoken Language Processing (ICSLP 2002)*, pp. 2041-2044, Colorado, USA, September 2002.

Appendix

Features of the thirty-two emotion speech databases reviewed.

#	Language	Author	Collector/ Distributor	Subjects	Other	Tr/on	Fe/gs
1	Slovenian, French, English, Spanish	D. Ambrus	Univ. of Maribor, Slovenia	2 actors	LG	Yes	Sim/ted
2	English	R & E/.Cowie	Belfast Queen's Univ.	40 volunteers	-	Yes	Natural
3	English	R. & E/.Cowie	Belfast Queen's Univ.	125 from TV	V	Yes	Semi-Nat
4	English,	R. Cole	Univ. Colorado, CSLR	780 kids	V	Yes	Natural
5	English	M. Liberman	Univ. Pensylv., LDC	Actors	-	Yes	Sim/ted
6	American English	J. Hansen	Univ. Colorado, LDC	32 Soldiers and students	-	Yes	Half-Half
7	English	C. Pereira	Univ. Macquire, Australia	2 actors	-	Yes	Sim/ted
8	English	M. Edgington	BT Labs, U.K.	1 male actor	LG	Yes	Sim/ted
9	English	T. Polzin, A.Waibel	Carnegie Mellon Univ.	5 drama students	LG	Yes	Sim/ted
10	English	V. Petrushin	Univ. of Indiana	30 native	-	Yes	Half-Half
11	English	R. Fernandez	MIT	4 drivers	-	Yes	Natural
12	German	K. Fisher	Univ. of Hamburg	58 native	-	70%	Natural
13	German	F. Schiel	Inst. for Phonetik and Oral Comm.,/Munich	45 native	V	Yes	Natural
14	German	W. Sendlemeier	Univ. of Berlin	10 actors	V, LG	Yes	Sim/ted
15	German	K. Alter	Max-Planck Institute	1 female	EEG	Yes	Sim/ted
16	German	K. R. Scherer	Univ. of Geneva	4 actors	-	Yes	Sim/ted
17	German	B. Wendt	Leibniz Institute	2 actors	-	Yes	Sim/ted
18	German	M. Schroeder	Univ. of Saarland	1 male	-	Yes	Sim/ted
19	German	M. Schroeder	Univ. of Saarland	6 native	-	Yes	Sim/ted
20	Japanese	R. Nakatsu	ATR Japan	100 native, 1 actor	-	Yes	Sim/ted
21	Japanese	Y. Niimi	Matsugasaki Univ., Kioto	1 male	-	Yes	Sim/ted
22	Japanese	N.Campbell	ATR, Japan	2 native	-	Yes	Sim/ted
23	Dutch	A.M. Sulter	Univ. of Groningen, ELRA	238 native	LG	Yes	Sim/ted
24	Dutch	S. Mozziconacci	Univ. of Amsterdam	3 native	-	Yes	Sim/ted
25	Spanish	J. M. Montero	Univ. of Madrid	1 male actor	-	Yes	Sim/ted
26	Spanish	I. Iriondo	Univ. Barcelona	8 actors	-	Yes	Sim/ted
27	Danish	I. Engberg	Univ. Aalborg	4 actors	-	Yes	Sim/ted
28	Hebrew	N. Amir	Holon, Israel	40 students	LG,M,G,H	Yes	Natural
29	Sweden	A. Abelin	Univ. of Indiana	1 native	-	Yes	Sim/ted
30	Chinese	F. Yu	Microsoft, China	Native from TV	-	Yes	Sim/ted
31	Russian	V. Makarova	Meikai Univ. /Japan	61 native	-	Yes	Sim/ted
32	Various	K. Scherer	Univ. of Geneva	109 passengers	V	No	Natural

Table 6. Tr/on: Transcription, Fe/gs: Feelings, LG:Laryngograph, M: Myogram of face, V: Video, G: Galvanic resistance, EEG: ElectroEngephaloGram, H: Heart Beat Rate, Sim/ted: Simulated.

#	Emotions	Material
1	Disgust, surprise, fear, anger sad, joy, 2 neutral.	Words, numbers, sentences (affirmative, interrogative), paragraph.
2	Anger, happy, sad, fear, neutral.	5 passages written in appropriate emotional tone and content (x emotion x speaker).
3	Various.	239 Clips (10-60 sec) from television, and real interviews.
4	Not- exactly.	Computer commands, Mathematics, words, Full digit sequences, Monologue.
5	Hot/cold anger, panic, shame anxiety, despair, sad, happy ...	Numbers, months, dates.
6	Stress, angry, question, fast, Lombard eff., soft, loud, slow.	Special words (brake, left, slow, fast, right...) that used in army; 16.000 utter.
7	Happy, sad, 2 anger, neutral.	80 sentences in each emotion per actor.
8	Anger, fear, sad, boredom, happy, neutral.	4 emotionally neutral sentences and 1 phrase in six emotional styles.
9	Fear, happy, anger, sad.	50 sentences x emotion x sp/er.
10	Fear, sad, anger, happy, neutral.	700 short utterances and 56 telephone messages of 15-90 sec.
11	Stress, Natural.	Numbers.
12	Mainly anger, dissatisfaction.	58 dialogues last from 18 to 33 minutes.
13	Mainly anger, dissatisfaction.	90 dialogues of 4.5 minutes.
14	Fear, anger, disgust, boredom, sad, joy, neutral.	10 words and 5 sentences per emotion; total 1050 utterances.
15	Happy, anger, neutral.	148 sentences x emotion; 1776 sentences.
16	Fear, anger, joy, sad, disgust.	Sentences from artificial language.
17	Fear, happy, sad, anger, disgust, neutral.	3000 nouns and 1200 pseudo words.
18	Soft, modal, loud.	All Diphones for all emotions.
19	Admiration, anger, disgust, boredom, contempt, worry,...	80 affect bursts for all emotions.
20	Anger, sad, happy, fear, surprise, disgust, playfulness.	100 neutral words x emotion, total 80000 words.
21	Anger, sad, joy.	Vowel-Consonant-Vowel sequences from 400 utterances.
22	Joy, anger, sad.	72 monologue texts of 450 sentences each.
23	Not exactly.	Subjects read 2 short texts with many quoted sentences to elicit emotional speech.
24	Fear, neutral, joy, boredom, anger, sad, indignation.	Total 315 sentences of semantically neutral content.
25	Sad, happy, anger, neutral.	3 passages, 15 sent., 30 words per emotion.
26	Fear, joy, desire, fury, surprise, sad, disgust.	2 texts x 3 emotion intensities per emotion; total 336 discourses.
27	Anger, happy, sad, surprise.	2 words, 9 sentences, 2 passages per emotion.
28	Anger, fear, joy, sad, disgust.	The subjects were told to recall an emotional situation of their life and speak about that.
29	Fear, anger, joy, shy, disgust, sad, surprise, dominance.	Sentences spoken in all emotional styles.
30	Anger, happy, sad, neutral.	Short audio clips from TV, 721 utter.
31	Surprise, happy, anger, sad, fear and neutral.	10 sent. x emotion x sp/er; total 3660 sentences.
32	Anger, humor, indifference, stress, sad.	Unobtrusive videotaping of passengers at lost luggage counter followed up by interviews.

Table 6 continued.