Appendix A
Emotional Speech Databases

Basque: Audiovisual Database of Emotional Speech in Basque by Navas et al. (2004a).

Emotions: *Anger, disgust, fear, happiness, sadness, surprise, neutral*

Elicitation: Audio–visual recordings of a professional actress uttering isolated words and digits as well as sentences of different length, both with emotional content and emotion-independent content

Size: 450 utterances with emotional content, 665 utterances with emotion-independent content, 1 female speaker.

Basque: Emotional Speech Database for Corpus Based Synthesis in Basque by Saratxaga et al. (2006).

Emotions: *Anger, disgust, fear, happiness, sadness, surprise, neutral*

Elicitation: Recordings of two speakers reading texts

Size: 702 sentences per emotion (20 h of recordings in total), two speakers (one female, one male).

Burmese: Emotional speech corpus by Nwe et al. (2001).

Emotions: *Anger, dislike, fear, happiness, sadness, surprise*

Elicitation: Recordings of two speakers uttering sentences experienced in daily life in different emotional states including rehearsals

Size: 144 sentences (54 sentences of the first speaker, 90 sentences of the second speaker, 0.6 to 1.6 s per sentence), two speakers.

Chinese: Emotional speech corpus by Yang and Campbell (2001).

Emotions: *Wide range of emotions*

Elicitation: Recordings of Mandarin Chinese speakers, combination of acoustic data from spontaneous conversation and experimental data from perceptual tests

Size: 6 h of recorded speech.

Chinese: Emotional speech database by Yuan et al. (2002).

Emotions: *Anger, fear, joy, sadness*

Elicitation: Speakers were asked to read the first paragraphs of a story evoking certain emotions. Recordings of the last paragraph plus two target sentences uttered emotionally

Size: 288 target sentences, nine female speakers.

Chinese: Drama corpus (Chuang and Wu 2004).

Emotions: *Anger, disgust, fear, happiness, sadness, surprise, neutral*

Elicitation: Recordings of professional impersonators in different emotional states

Size: 2,100 sentences in 440 dialogues by two speakers (1,085 sentences in 227 dialogues from the leading man, 1,015 sentences in 213 dialogues from the leading woman).

Chinese: Emotional speech database by Dan-Ning Jiang (2004).

Emotions: *Anger, fear, happiness, sadness, surprise, neutral*

Elicitation: Recordings of an amateur actress uttering different sentence types (statements, questions)

Size: Approx. 200 utterances per emotion, neutral database with approx. 300 utterances, one female speaker.

Chinese: Mandarin Emotional Speech Corpus I. (Pao et al. 2004).

Emotions: *Anger, boredom, happiness, sadness, neutral*

Elicitation: Recordings of short utterances expressed by native Mandarin speakers in different emotional states

Size: 558 utterances, 12 speakers (seven females, five males).

Chinese: Mandarin Emotional Speech Corpus II used by Pao et al. (2004).

Emotions: *Anger, boredom, happiness, sadness, neutral*

Elicitation: Recordings of utterances expressed by professional Mandarin speakers in different emotional states

Size: 503 utterances, two speakers (one female, one male).

Chinese: Spontaneous speech corpus used by Tao (2004).

Emotions: *Anger, fear, hate, joy, sadness, surprise, neutral*

Elicitation: No further information available

Size: 835 sentences, 5,000 words used by Tao (2004).

Chinese: Acted speech corpus by Tao et al. (2006).

Emotions: *Anger, fear, happiness, sadness, neutral*

Elicitation: A professional actress reading texts from a *Reader's Digest* collection

Size: 1,500 utterances, 3,649 phrases, one speaker.

Chinese: Emotional speech corpus (Wu et al. 2006).

Emotions: *Anger, fear, happiness, sadness, neutral*

Elicitation: Recordings of 25 actors uttering short passages with emotional content and command phrases

Size: 150 short passages (30–50 s), 5,000 command phrases (2–10 s), 50 speakers (25 females, 25 males).

Chinese: Speech database by Zhang et al. (2006).

Emotions: *Anger, fear, joy, sadness, neutral*

Elicitation: Recordings of eight speakers (acoustically isolated room)

Size: 2,400 sentences (20 sentences, uttered three times each, for every emotion), eight speakers (four females, four males).

Chinese, English, Japanese: JST/CREST database by Campbell (2002).

Emotions: *Wide range of emotional states and emotion-related attitudes*

Elicitation: Natural emotions of volunteers recording their domestic and social spoken interactions for extended periods throughout the day

Size: Recordings still ongoing, target: 1,000 h over 5 years.

Chinese, English, Japanese: Speech corpus by Jiang et al. (2005).
Emotions: *Angry, calm, happy, sad, surprise*
Elicitation: Recordings of a speaker uttering a sentence in three languages and in different emotional states
Size: 750 utterances (50 utterances per language and emotion), one speaker.

Danish: Danish Emotional Speech Database (Engberg et al. 1997).
Emotions: *Anger, happiness, sadness, surprise, neutral*
Elicitation: Recordings of actors familiar with radio theater uttering single words, sentences and passages of fluent speech in different emotional states
Size: Approx. 10 min of speech in total (30 s per emotion per speaker) plus neutral recordings (2 speakers) plus extra recordings (three speakers), four speakers in total (two females, two males).

Dutch: Emotional database by Van Bezooijen (1984).
Emotions: *Anger, contempt, disgust, fear, interest, joy, sadness, shame, surprise, neutrality*
Elicitation: Recordings of speakers reading semantically neutral phrases
Size: four phrases, eight speakers (four females, four males).

Dutch: Groningen corpus S0020 ELRA (1996) (see `http://www.elra.info/`).
Emotions: *Mostly neutral, few emotions*
Elicitation: Recordings of speakers reading short texts, sentences, numbers, monosyllabic words, long vowels and an extended word list
Size: 20 h of speech, 238 speakers.

Dutch: Emotional database by Mozziconacci and Hermes (1999).
Emotions: *Anger, boredom, fear, indignation, joy, sadness, neutrality*
Elicitation: Recordings of actors expressing semantically neutral sentences in different emotional states after the respective emotion has been elicited by reading a semantically emotional sentence
Size: three speakers (one female, two males), 315 utterances, five sentences.

Dutch: Experiment at Tilburg University (Wilting et al. 2006).
Emotions: *Acted negative, acted positive, negative, positive, neutral*
Elicitation: Recordings of participants reading sentences in different (partly acted) emotional states according to the mood induction procedure proposed by Velten (1968)
Size: 50 participants (31 females, 19 males), each reading 40 sentences of 20 s length.

English: Database produced by Cowie and Douglas-Cowie (1996).
Emotions: *Anger, fear, happiness, sadness, neutral*
Elicitation: Recorded passages of speakers from the Belfast area
Size: 40 speakers (20 females, 20 males), passages of 25–30 s length.

English: SUSAS database (Hansen et al. 1998).
Emotions: *Talking styles (angry, clear, fast, loud, question, slow, soft), single tracking task (high stress, Lombard effect, moderate), dual tracking task (high stress, moderate), actual speech under stress (anxiety, fear, G-force, Lombard effect, noise), psychiatric analysis (angry, anxiety, depression, fear)*

Elicitation: Recordings of isolated-word utterances under simulated or actual stress in several scenarios, e.g., amusement park roller-coaster, helicopter cockpit, patient interviews

Size: Approx. 16,000 utterances of 36 speakers (13 females, 23 males) in total.

English: Emotional speech database by Li and Zhao (1998).

Emotions: *Anger, fear, happy, sad, surprised, neutral*

Elicitation: Recordings of actors uttering 20 sentences with emotional content and three sentences without emotional content in different emotional states

Size: 5 untrained speakers (two females, three males), 23 sentences per speaker.

English: Emotional speech database by Whiteside (1998).

Emotions: *Cold Anger, elation, happiness, hot anger, interest, sadness, neutral*

Elicitation: Recordings of actors uttering sentences in different emotional states

Size: 70 utterances, two speakers (one female, one male), five different short sentences.

English: Emotional corpus by Cowie et al. (1999b).

Emotions: *Wide range of emotions as defined in the FEELTRACE tool*

Elicitation: Video tape recordings of groups of three friends each discussing about issues they strongly felt about

Size: Recordings of 1 h per group, nine speakers (three groups of three friends).

English: Emotional speech database by Robson and Mackenzie-Beck (1999).

Emotions: *Smiling, neutral*

Elicitation: Recordings of speakers uttering sentences in a neutral state and while smiling

Size: 66 utterances, 11 speakers, three sentences.

English: Reading/Leeds Emotional Speech Corpus (Greasley et al. 2000).

Emotions: *Anger, disgust, fear, happiness, sadness, neutral*

Elicitation: Recordings of interviews on radio/television, speakers asked by interviewers to relive emotionally intense experiences

Size: Approx. 5 h of samples of emotional speech.

English: Emotional speech database by McGilloway et al. (2000).

Emotions: *Anger, fear, happiness, sadness, neutral*

Elicitation: Recordings of speakers reading emotional texts in appropriate style

Size: 40 speakers, five texts (100 syllables each).

English: Emotional speech database by Pereira (2000).

Emotions: *Cold anger, happiness, hot anger, sadness, neutral*

Elicitation: Recordings of actors uttering two sentences in different emotional states

Size: 80 utterances (two repetitions of 40 utterances), two speakers.

English: Database produced by Polzin and Waibel (2000).

Emotions: *Anger, sadness, neutrality (other emotions as well, but in insufficient numbers to be used)*

Elicitation: Audio–visual data, i.e., sentence-length segments taken from acted movies

Size: 1,586 angry segments, 1,076 sad segments, 2,991 neutral segments.

English: Belfast Naturalistic Database (Douglas-Cowie et al. 2000).

Emotions: *Wide range of emotions*

Elicitation: Audio–visual recordings of people discussing emotive subjects with each other/the research team plus recordings of extracts from television programs, i.e., members of the public interacting in a way that appears essentially spontaneous

Size: 239 clips (209 from TV Recordings, 30 from interview recordings, clip durations: 10–60 s), 125 speakers (94 females, 31 males).

English: Database by France et al. (2000).

Emotions: *Depression, suicidal state, neutrality*

Elicitation: Recordings of spontaneous dialogues between patients and therapists in therapy sessions, phone conversations and post therapy evaluation sessions

Size: 115 speakers (48 females, 67 males).

English: DARPA Communicator corpus (Walker et al. 2001).

Emotions: *Annoyance, frustration*

Elicitation: Users making air travel arrangements over the phone

Size: Recordings of simulated interactions with a call center, 13,187 utterances in total (1,750 emotional utterances).

English: Capital Bank Service and Stock Exchange Customer Service (Devillers et al. 2002).

Emotions: *Anger, excuse, fear, satisfaction, neutral*

Elicitation: Human–human interaction in a stock exchange customer service (call) center

Size: 100 dialogues, 5,229 speaker turns.

English: Emotional speech database by Fernandez and Picard (2003).

Emotions: *Stress*

Elicitation: Recordings of speakers solving mathematical problems while driving in a car simulator

Size: four speakers, four situations, 598 utterances, length varying from 0.5 to 6 s.

English: Speech database by Lee and Narayanan (2003).

Emotions: *Negative, non-negative*

Elicitation: Users interacting with a machine agent in a call center

Size: 1,367 utterances (776 utterances of female speakers, 591 utterances of male speakers).

English: LDC Emotional Prosody Speech and Transcription used by Liscombe et al. (2003) and Yacoub et al. (2003).

Emotions: *Anxiety, boredom, cold anger, contempt, despair, disgust, elation, happy, hot anger, interest, panic, pride, sadness, shame, neutral*

Elicitation: Professional actors reading short (4-syllables each) dates and numbers

Size: eight actors (five females, thre males), 44 utterances used by Liscombe et al. (2003), 2,433 utterances used by Yacoub et al. (2003) .

English: ORESTEIA database by McMahon et al. (2003).

Emotions: *Irritation, shock, stress*

Elicitation: Audio-physiological (and partly visual) recordings of driving persons encountering various problems (deliberately positioned obstructions, dangers, annoyances "on the road")

Size: 90 min per session and speaker, 29 speakers.

English: Sensitive Artificial Listener (SAL) database by Cowie et al. (2004).

Emotions: *Wide range of emotions or emotion related states*

Elicitation: Audio–visual recordings of speakers interacting with an artificial listener with different personalities

Size: Recordings of approx. 10 h, 20 speakers.

English: Speech database by Lee et al. (2004) and Yildirim et al. (2004).

Emotions: *Angry, happy, sad, neutral*

Elicitation: Recordings of a semi-professional actress uttering 112 unique sentences in four emotions

Size: 880 utterances, one female speaker.

English: Emotional speech synthesis database (Tsuzuki et al., 2004).

Emotions: *Anger, happiness, sadness, neutral*

Elicitation: Recordings of a non-professional male speaker uttering short declarative sentences with emotional content

Size: 363 utterances, one male speaker.

English: Modified LDC CallFriend corpus prepared by Yu et al. (2004).

Emotions: *Boredom, happy, hot anger, interest, panic, sadness, no emotion plus numerical values (on a discretized scale from 1 to 5) for each of arousal, valence and engagement*

Elicitation: Recordings of social telephone conversations between friends

Size: 1,888 utterances (1,011 utterances from female speakers, 877 utterances from male speakers, eight speakers (four females, four males).

English: WOZ data corpus (Zhang et al., 2004).

Emotions: *Confidence, puzzle, hesitation*

Elicitation: Audio–visual recordings of children interacting with an intelligent tutoring system for learning basic concepts of Mathematics and Physics

Size: 714 students' utterances (approx. 50 min of clean speech), 4.2 s of speech and 8.1 words per utterance, 17 speakers.

English: Speech database by Lee et al. (2005).

Emotions: *Angry, happy, sad, neutral*

Elicitation: Recordings of a male speaker producing sentences with non-emotional content in the respective emotional states (including biosensor data)

Size: 280 utterances, 14 sentences, one male speaker.

English: HMIHY speech database (Liscombe et al. 2005).

Emotions: *Positive/neutral, somewhat angry, somewhat frustrated, somewhat other negative, very angry, very frustrated, very other negative*

Elicitation: Recordings of callers interacting with an automated agent concerning account balance, explanation of bill charges, AT&T rates and calling plans, etc.

Size: 5,690 dialogues, 20,013 user turns.

English: Expressive spoken corpus of children's stories modified by Alm and Sproat (2005).

Emotions: *Angry, disgusted, fearful, happy, sad, surprised, neutral*

Elicitation: Recordings of a semi-professional female speaker reading two children's stories in an extremely expressive mode

Size: Approx. 10 min of speech, 128 sentences, one female speaker.

English: ITSPOKE corpus (Ai et al. 2006).

Emotions: *Positive, negative, neutral*

Elicitation: Students interacting with an interactive tutoring system

Size: 100 dialogues, 20 students, 2,252 student turns, 2,854 tutor turns plus a set of human–human corpus.

English: Situation Analysis in a Fictional and Emotional (SAFE) corpus (Clavel et al., 2006).

Emotions: *Fear, other negative emotions, positive emotions, neutral*

Elicitation: Audio–visual excerpts taken from movie DVDs, abnormal contexts

Size: 400 sequences, total length 7 h, 4,724 segments of speech (up to 80 s).

English: Castaway database (Devillers et al. 2006).

Emotions: *Wide range of emotions*

Elicitation: Audio–visual recordings of a reality TV show

Size: 10 recordings (30 min each), 10 speakers.

English: Speech database by Lee et al. (2006).

Emotions: *Angry, happy, sad, neutral*

Elicitation: Recordings and magnetic resonance images of a male speaker uttering a set of four sentences

Size: 80 utterances (four sentences, five repetitions, four emotions), one male speaker.

English: Emotional speech corpus by Kumar et al. (2006).

Emotions: *Inappropriateness, lack of clarity, uncertainty, neutral*

Elicitation: Recordings of participants interacting with an SLDS in terms of a customer survey about grocery stores plus answering of a questionnaire

Size: 257 utterances, 17 participants (10 females, 7 males).

English: ISL Meeting corpus (Neiberg et al. 2006).

Emotions: *Negative, positive, neutral*

Elicitation: Recordings of 18 meetings with a total of 92 persons and an average duration of 35 min accompanied by orthographic transcription

Size: 12,068 utterances, thereof 424 negative, 2,073 positive and 9,571 neutral.

English: "Yeah right" corpus by Tepperman et al. (2006).

Emotions: *Sarcastic, neutral*

Elicitation: "Yeah right" utterances taken from the Switchboard and Fisher corpora of spontaneous telephone dialogues

Size: 131 utterances.

English, French, German: Speech database by Klasmeyer et al. (2000).

Emotions: *Emotional, neutral*

Elicitation: Recordings of English, French and German speakers reading sentences and uttering passages of spontaneous speech

Size: Approx. 13,000 utterances, 120 speakers (25 English, 65 French, 30 German).

English, French, German, Italian: Geneva Airport Lost Luggage Speech Database by Scherer and Ceschi (1997).
Emotions: *Anger, good humor, indifference, resignation, worry*
Elicitation: Audio–visual recordings of passengers at the lost luggage desk at Geneva Airport plus interviews
Size: 112 passengers, 12 airline employees, 10–20 min of interaction per passenger at the desk.

English, French, Slovenian, Spanish: INTERFACE Emotional Speech Synthesis Database (IESSDB, Nogueiras et al. 2001).
Emotions: *Anger, disgust, fear, joy, sadness, surprise, neutral*
Elicitation: Six different kinds of sentences (affirmative, exclamatory, interrogative, paragraphs of approx. five sentences, isolated words and digits) spoken by professional actors in each language and each emotion
Size: Two actors (one female, one male), 150–190 utterances for each of the six emotional styles in four languages.

English, German: AIBO (Erlangen database, Batliner et al. 2004a).
Emotions: *Angry, bored, emphatic, helpless, joyful, motherese, reprimanding, surprised, touchy (irritated), neutral and rest*
Elicitation: Children interacting with a WOZ robot
Size: 51 German children (30 females, 21 males, 51,393 words, 9.2 h of speech), 30 English children (5,822 words, 1.5 h of speech).

English, German: Corpus in the framework of the FERMUS-III project (Rigoll et al. 2005).
Emotions: *Anger, disgust, fear, sadness, surprise, neutral*
Elicitation: First set: Actors uttering sentences in different emotional states, second set: utterances of automotive infotainment speech interaction dialogues
Size: 3,529 utterances (first set: 2,829 utterances, second set: 700 utterances), 13 speakers (one female, 12 males).

English, German, Japanese: Material taken from the TV series *Ally McBeal* (Braun and Katerbow 2005).
Emotions: *Cold anger, fear, hot anger, joy, sadness, neutral*
Elicitation: Audio–visual data taken from a DVD
Size: six speakers (three females, three males), 135 utterances in total (45 utterances per language).

English, Japanese: Emotional database by Shigeno (1998).
Emotions: *Anger, disgust, fear, happiness, sadness, surprise*
Elicitation: Audio–visual recordings of actors uttering short sentences and words in English and Japanese
Size: two speakers (one American, one male), 36 audio–visual stimuli.

English, Korean: Emotional database by Chung (2000).
Emotions: *Joy, sadness, neutral*
Elicitation: Audio–visual recordings of female speakers in Korean and American television shows talking about problems in their lives, expressing sadness and joy

Size: one Korean speaker (eight utterances, each lasting 1 min), five American speakers (one neutral and one emotional utterance, each lasting 1 min, per speaker).

English, Malay: Emotional speech database by Razak et al. (2005).
Emotions: *Anger, disgust, fear, happy, sad, surprise*
Elicitation: Recordings of learning actors uttering sentences frequently used in everyday communication
Size: 1,200 utterances, four different sentences, female and male speakers.

English, Swedish: Speech database by Laukka (2004).
Emotions: *Anger, disgust, fear, happiness, sadness, neutral*
Elicitation: Recordings of British and Swedish actors uttering short sentences
Size: 176 utterances, eight speakers (four females, four males, four British, four Swedish).

Farsi/Persian: Farsi emotional speech corpus (Gharavian and Ahadi 2005).
Emotions: *Angry, sad, neutral*
Elicitation: Sentences of a non-emotional corpus reuttered angrily, sadly and neutrally
Size: 1,518 utterances, one male speaker.

Finnish: MediaTeam emotional speech corpus (Väyrynen et al. 2003).
Emotions: *Anger, happiness/joy, sadness, neutral*
Elicitation: Recordings of professional actors reading out a phonetically rich Finnish passage with non-emotional content
Size: 56 monologues, 14 speakers (six females, eight males).

French: Emotional database by Johnstone and Scherer (1999).
Emotions: *Anxious, bored, depressed, happy, irritated, tense, neutral*
Elicitation: Recordings of students playing a manipulated computer space game and making statements to their emotions, furthermore recordings of biosignals
Size: 36 males speakers.

French: Messages corpus by Chateau et al. (2004).
Emotions: *Positive emotion, negative emotion, no particular emotion (neutral)*
Elicitation: Recordings of France Telecom customers describing their opinions about the customer care service
Size: 103 messages split into 478 emotional utterances, 103 speakers.

French: EmoTV corpus (Abrilian et al. 2005).
Emotions: *Anger, despair, disgust, doubt, exaltation, fear, irritation, joy, pain, sadness, serenity, surprise, worry, neutral plus 176 fine-grain categories*
Elicitation: Audio–visual recordings of TV news interviews
Size: 51 recordings, 48 speakers, 14 s per recording, 2,500 words.

French: Talkapillar corpus by Beller and Marty (2006).
Emotions: *Anger, boredom, disgust, happiness, indignation, sadness, surprise (negative and positive), neutral and neutral question*
Elicitation: Recordings of one actor reading semantically neutral sentences in different emotional states
Size: 539 utterances, one speaker, 26 sentences, 2 h of speech.

French: CEMO corpus (Devillers et al. 2006).

> Emotions: *8 coarse-grained classes: anger, fear, hurt, other positive, relief, sadness, surprise, neutral, additionally 21 fine-grained classes*
> Elicitation: Real agent-client recordings obtained from a medical emergency call center offering medical advice
> Size: 688 agent-client dialogues of around 20 h, seven agents (four females, three males), 784 clients (513 females, 271 males), 48 turns per dialogue in average, 262,000 words thereof 92,000 distinct words.

German: Emotional speech database by Tolkmitt and Scherer (1986).

> Emotions: *Cognitive and emotional stress*
> Elicitation: Recordings of speakers who were shown slides either containing logical problems which they had to solve verbally or photos of injured people which they were asked to comment on
> Size: 60 speakers (27 females, 33 males), 20 slides, max. 40 s speech per slide.

German: Geneva Vocal Emotion Expression Stimulus Set (GVEESS, Banse and Scherer, 1996).

> Emotions: *Anxiety, boredom, cold anger, contempt, disgust, elation, happiness, hot anger, interest, panic fear, pride, sadness, shame*
> Elicitation: Audio–visual recordings of actors acting scripted emotion-eliciting scenarios for each emotion
> Size: 12 actors (six females, 6 males), 224 recordings.

German: Speech database by Dellaert et al. (1996).

> Emotions: *Anger, fear, happiness, sadness, neutral*
> Elicitation: Recordings of speakers reading a variety of sentences in different emotional states
> Size: five actors, 50 sentences, 1,000 utterances.

German: Speech database by Klasmeyer (1996).

> Emotions: *Anger, boredom, fear, happiness, sadness, neutral*
> Elicitation: Recordings of actors uttering short sentences with non-emotional content
> Size: 10 sentences per emotion, three actors.

German: Emotional speech material used by Alter et al. (1999) and Alter et al. (2000).

> Emotions: *Anger, happiness, neutral*
> Elicitation: Recordings of sentences with emotional content spoken by a trained female speaker in a sound proof room at the University of Leipzig, ratings by 30 subjects on a 5-point scale indicating the three emotions
> Size: 148 sentences, one speaker.

German: Database used and produced by Batliner et al. (2000).

> Emotions: *Anger, neutral*
> Elicitation: Recordings of an acting person produced within the VERBMOBIL scenario, recordings of naive subjects reading emotional sentences, recordings of angry and neutral persons in a WOZ scenario
> Size: Acted data: 1,240 neutral turns and 96 angry turns, Read data: 50 neutral and 50 emotional sentences, WOZ data: 2,395 turns (20 dialogues) planned to be extended.

German: Database of affect bursts (Schröder 2000).

Emotions: *Admiration, anger, boredom, contempt, disgust, elation, relief, startle, threat, worry*

Elicitation: Speakers reading silently a frame story, recordings of an produced affect burst of their choice plus two produced affect bursts from a list

Size: 180 vocalizations (30 vocalization per speaker), six speakers (three females, three males, thereof four amateur actors).

German: Lego corpus (Kehrein 2001).

Emotions: *Wide range of emotions*

Elicitation: Recordings of dialogues between two persons interactively trying to build a Lego kit without seeing each other

Size: 180 min of speech thereof 372 emotional turns.

German: SmartKom database (Schiel et al., 2002; Wahlster, 2006).

Emotions: *Anger, gratification, helplessness, irritation, joy, pondering, reflecting, surprise, neutral, unidentifiable episodes*

Elicitation: Audio–visual recordings of human–computer information system dialogues in a WOZ scenario

Size: 224 speakers, 448 recordings, 4-5 min sessions.

German: Speech database used by Tato et al. (2002).

Emotions: *Angry, bored, happy, sad, neutral*

Elicitation: Recordings of speakers put in an emotional state and reading commands to the Sony entertainment robot AIBO

Size: 2,800 utterances, 40 commands, 14 speakers (seven females, seven males).

German: SYMPAFLY database (Batliner et al. 2004b).

Emotions: *Angry, compassionate, emphatic, helpless, ironic, joyful, panic, surprised, touchy, neutral*

Elicitation: Naive users book flights using a machine dialogue system

Size: 270 dialogues and 29,200 words in total, three parts with increasing system performance and 62–110 speakers per part.

German: Berlin Database of Emotional Speech (Burkhardt et al. 2005).

Emotions: *Anger, boredom, disgust, fear, joy, sadness, neutral*

Elicitation: Recordings of non-professional actors uttering sentences with non-emotional content in each emotion

Size: More than 800 utterances, 10 speakers (five females, five males).

German: EMO-SI database (Schuller et al. 2005).

Emotions: *Anger, disgust, fear, joy, sadness, surprise, neutrality*

Elicitation: Spontaneous and acted emotions in short phrases of car interaction dialogues

Size: 39 speakers (three females, 36 males), 2,730 samples (70 samples per speaker).

German: Emotional database by Kim and André (2006).

Emotions: *High arousal (negative valence, positive valence), low arousal (negative valence, positive valence)*

Elicitation: Recordings of users playing a quiz (including biosensor data)

Size: 45 min per speaker, three speakers.

Greek: Greek Emotional Database (Fakotakis 2004).

Emotions: *Anger, fear, joy, sadness, neutral*

Elicitation: Recordings of a professional actress reading semantically neutral words and sentences in different emotions

Size: 10 single words, 20 short sentences, 25 long sentences, 12 passages of fluent speech, one female speaker.

Hebrew: Emotional speech corpus by Amir et al. (2000).

Emotions: *Anger, disgust, fear, joy, sadness, neutral*

Elicitation: Recordings of students telling of personal experiences evoking certain emotions (including biosensor data)

Size: 31 speakers (15 females, 16 males).

Japanese: Nicholson et al. (1999).

Emotions: *Anger, disgust, fear, joy, sadness, surprise, teasing, neutral*

Elicitation: Speakers were asked to read a word list in eight emotions trying to imitate emotional recordings produced by radio actors

Size: 50 females and 50 males native Japanese speakers uttering a list of 100 context-free Japanese words eight times (once per emotion), each of the Japanese phonemes equally represented within the list.

Japanese: Speech database by Oudeyer (2003).

Emotions: *Anger, joy/pleasure, sorrow/sadness/grief, normal/neutral*

Elicitation: Recordings of professional speakers uttering short sentences or phrases and imagining themselves uttering these sentences to a pet robot

Size: 4,800 utterances (200 per speaker and emotion), six speakers (female and male).

Japanese: Prosodic corpus by Hirose et al. (2004).

Emotions: *Anger, calm, joy, sadness*

Elicitation: Recordings of a female narrator reading sentences with emotional content

Size: Approx. 1,600 utterances (around 400 sentences per emotion), one female speaker.

Japanese: Emotional speech database by Iwai et al. (2004).

Emotions: *Anger, joy, sadness, neutral*

Elicitation: Recordings of students uttering the word "okaasan" (Japanese: "mother") in four emotions

Size: 766 utterances, three male speakers.

Japanese: Emotional speech database by Takahash et al. (2005).

Emotions: *Angry, bright, excited, raging, neutral*

Elicitation: Recordings of expressive speech sounds narrated by professional actors

Size: 1,500 expressive speech sounds, eight speakers.

Japanese: Emotional speech database by Nisimura et al. (2006).

Emotions: *Anger, contempt, contentment, depression, excitement, fear, joy, mirth, pleasure, pressure, sadness, surprise, tension, tiredness, displeasure, neutral*

Elicitation: Recordings of children's utterances extracted from a public spoken

dialogue system

Size: 2699 utterances.

Korean: Emotional speech database by Kim et al. (2004b).

Emotions: *Anger, joy, sadness, neutral*

Elicitation: Recordings of speakers uttering short sentences in different emotional states

Size: 400 utterances, five different sentences (less than 1.5 sec. duration), four male speakers.

Korean: Database produced by Media and Communication Signal Processing Laboratory, Prof. C.Y. Lee of Yonsei University (Kim et al. 2005).

Emotions: *Angry, joyful, sad, neutral*

Elicitation: Ten speakers uttering dialogic sentences expressing natural emotions with easy pronunciation; afterwards subjective emotion recognition by human listeners for verification

Size: 5,400 sentences: 45 dialogic sentences, three repetitions, four emotions, 10 speakers (five females, five males).

Russian: RUSLANA database (Makarova et al. 2002).

Emotions: *Anger, fear, happiness, sadness, surprise, neutral*

Elicitation: Recordings of actors expressing emotional sentences

Size: 61 actors (49 females, 12 males), 610 utterances.

Spanish: SES Spanish Emotional Speech database (Montero et al. 1999).

Emotions: *Anger, happiness, sadness, surprise, neutral*

Elicitation: Recordings of an actor reading neutral texts in different emotional states

Size: three passages and 15 sentences acted by one speaker in four emotions plus neutral style.

Spanish: Emotional speech database by Iriondo et al. (2000).

Emotions: *Desire, disgust, fear, fury, joy, sadness, surprise*

Elicitation: Recordings of actors reading texts in different emotional states and intensities

Size: eight actors (four females, four males), three intensities, 336 utterances.

Spanish: Emotional speech database by Álvarez Martínez and Barrientos Cruz (2005).

Emotions: *Anger, fear, happiness, sadness, neutral*

Elicitation: Recordings of actors and actresses uttering sentences in different emotional states plus extracted utterances from DVD movies

Size: 380 utterances (300 utterances with four different sentences as synthetic data set (actors), 80 utterances as real data set (DVD movies)), 15 non-professional speakers (female and male) in the synthetic data set.

Swedish: Emotional speech database by Abelin and Allwood (2000).

Emotions: *Anger, disgust, dominance, fear, joy, sadness, shyness, surprise*

Elicitation: Recordings of a speaker uttering a non-emotional phrase in different emotional states

Size: 1 male speaker.

Swedish: VeriVox database by Karlsson (1999).

Emotions: *Stress ranked from zero to nine identifying 5 as normal stress level*

Elicitation: Recordings of male speakers reading texts in different tasks and stress levels

Size: 50 males speakers, 30 min per speaker.

Swedish: The Voice Provider Material (VP, Neiberg et al. 2006).

Emotions: *Emphatic, negative, neutral*

Elicitation: Recordings of voice-controlled telephone services (traffic information, postal assistance, etc.)

Size: 7619 utterances, thereof 160 emphatic, 335 negative and 7,124 neutral.

No specific language: Corpus of infants' cries (Matsunaga et al. 2006).

Emotions: *Anger, hunger, pampered, sadness, sleepiness*

Elicitation: Infants' cries recorded by their mothers at home using a digital recorder, emotional judgment by the mothers taking into consideration facial expressions, behavior, etc., emotional intensity ranked from zero (emotion not contained at all) to four (emotion fully contained)

Size: 402 cries, 23 infants (12 females, 11 males, age: 8–13 months).

(see also "The HUMAINE Portal" website at `http://emotion-research.net/wiki/Databases`)

Appendix B
Used Abbreviations

ANN	Artificial Neural Network
ANS	Autonomic Nervous System
ASR	Automatic Speech Recognition
ATIS	Air Travel Information System
BDI	Beliefs-Desires-Intentions
BEEP	British English Example Pronunciations dictionary
BEEV	British English Emotional Vocabulary
CSLU	Center for Spoken Language Understanding (Oregon Health & Science University)
DBN	Dynamic Belief Network
DETT	Disposition–Emotion–Trigger–Tendency
DVD	Digital Versatile Disk
EBNF	Extended Backus-Naur Form
ELRA	European Language Resources Association
ETN	Emotion Transition Network
FIA	Form Interpretation Algorithm
GMM	Gaussian Mixture Model
GSM	Global System for Mobile communications
HMIHY	How May I Help You?
HMM	Hidden Markov Model
HTK	Hidden Markov Model Toolkit
IPA	International Phonetic Alphabet
ITU-T	Telecommunication Standardization Sector of the International Telecommunication Union
IVR	Interactive Voice Response
JPEG	Joint Photographic Experts Group
JSGF	Java Speech Grammar Format
KNN	k-Nearest Neighbor
LDC	Linguistic Data Consortium (University of Pennsylvania)
LP	Linear Prediction
MDP	Markov Decision Process
MFCC	Mel-Frequency Cepstral Coefficient
MIMO	Multiple Input Multiple Output

MP3	MPEG-1 Audio Layer 3
MPEG	Moving Picture Experts Group
OCC	Ortony, Clore, Collins
PARADISE	PARAdigm for DIalogue System Evaluation
PDA	Personal Digital Assistant
PLP	Perceptual Linear Prediction
POMDP	Partially Observable Markov Decision Process
PROMISE	PROcedure for Multimodal Interactive System Evaluation
RIFF	Resource Interchange File Format
ROVER	Recognizer Output Voting Error Reduction
SAMPA	Speech Assessment Methods Phonetic Alphabet
SLDS	Spoken Language Dialogue System
SRGS	Speech Recognition Grammar Specification
SVM	Support Vector Machine
TRINDI	Task oRiented INstructional DIalogue
W3C	World Wide Web Consortium
WETN	Word-Emotion Transition Network
WOZ	Wizard-of-OZ
WTN	Word Transition Network
XML	eXtensible Markup Language

References

Abdennaher S, Aly M, Bühler D, Minker W, Pittermann J (2007) BECAM tool – A semi-automatic tool for bootstrapping emotion corpus annotation and management. In: European conference on speech and language processing (EUROSPEECH), Antwerp, Belgium

Abdi H, Valentin D, Edelman B (1999) Neural networks. Sage, Newbury Park, USA

Abelin Å, Allwood J (2000) Cross linguistic interpretation of emotional prosody. In Proceedings of ISCA workshop on speech and emotion, Belfast, United Kingdom

Abrilian S, Devillers L, Buisine S, Martin J-C (2005) EmoTV1: Annotation of real-life emotions for the specification of multimodal affective interfaces. In: 11th international conference on human–computer interaction (HCII'2005), Las Vegas, USA

Ai H, Litman DJ, Forbes-Riley K, Rotaru M, Tetreault J, Purandare A (2006) Using system and user performance features to improve emotion detection in spoken tutoring dialogues. In: International conference on speech and language processing (ICSLP), Pittsburgh, USA, pp 797–800

Alexandris C, Fotinea S-E (2004) Reusing language resources for speech applications involving emotion In: International conference on language resources and evaluation (LREC), Lisbon, Portugal, pp 1383–1386

Allen J (1995) Natural Language Understanding. Benjamin Cummings, Menlo Park, USA

Alm CO, Sproat R (2005) Perceptions of emotions in expressive storytelling. In: International conference on speech and language processing (ICSLP), Lisbon, Portugal, pp 533–536

Alter K, Rank E, Kotz SA, Pfeifer E, Besson M, Friederici AD, Matiasek J (1999) On the relations of semantic and acoustic properties of emotions. In: International congress of phonetic sciences (ICPhS), San Francisco, USA

Alter K, Rank E, Kotz SA, Toepel U, Besson M, Schirmer A, Friederici AD (2000) Accentuation and emotions – Two different systems? In: Proceedings of ISCA workshop on speech and emotion, Belfast, United Kingdom

Álvarez Martínez C, Barrientos Cruz A (2005) Emotion recognition in non-structured utterances for human-robot interaction. In: IEEE International Workshop on Robot and Human Interactive Communication, Nashville, USA, pp 19–23

Amir N, Ron S, Laor N (2000) Analysis of an emotional speech corpus in Hebrew based on objective criteria. In: Proceedings of ISCA workshop on speech and emotion, Belfast, United Kingdom

Andersson G, Paiva A, Arafa Y, Piedade M, Botelho L, Ramos P, Mourão D, Sengers P (2002) Shell for affective user modelling. SAFIRA Deliverable D3.2

André E (2000) The generation of multimedia presentations. In: Dale R, Moisl H, Somers H (eds) Handbook of natural language processing. Marcel Dekker, New York, USA, pp 305–327

André E, Rehm M, Minker W, Bühler D (2004) Endowing spoken language dialogue systems with emotional intelligence. In: Tutorial and Research Workshop Affective Dialogue Systems, Irsee, Germany, pp 178–187

Androutsopoulos I, Aretoulaki M (2003) Natural language interaction. In: Mitkov R (ed) Handbook of computational linguistics, chapter 35. Oxford University Press, New York, USA, pp 629–649

Ang J, Dhillon R, Krupski A, Shriberg E, Stolcke A (2002) Prosody-based automatic detection of annoyance and frustration in human–computer dialog. In: International conference on speech and language processing (ICSLP), Denver, USA, pp 2037–2040

Anscombe E, Geach P (1970) Descartes philosophical writings. The Open University, Nelson

Arnold M (1960) Emotion and personality. Columbia University Press, New York, USA

Aubergé V, Audibert N, Rilliard A (2003) Why and how to control the authentic emotional speech corpora. In: European conference on speech and language processing (EUROSPEECH), Geneva, Switzerland, pp 185–188

Averill JR (1980) A constructivist view of emotion. In: Plutchik R, Kellerman H (eds) Emotion: Theory, research and experience, vol 1. Academic, New York, USA, pp 305–339

Azzini I, Falavignat D, Gretter R, Lanzola G, Orlandi M (2001) First Steps Toward an Adaptive Spoken Dialogue System in Medical Domain. In: European conference on speech and language processing (EUROSPEECH), Aalborg, Denmark, pp 1327–1330

Baken RJ, Orlikoff RF (2000) Clinical measurements of speech and voice. Singular Thomson learning, San Diego, USA

Bakis R (1976) Continuous speech word recognition via centisecond acoustic states. In: Proceedings of the 91st annual meeting of the acoustical society America, Washington DC, USA

Banse R, Scherer KR (1996) Acoustic profiles in vocal emotion expression. J Person Soc Psychol 70(3):614–636

Bard P (1934) On emotional expression after decortication with some remarks on certain theoretical views. Part II. Psychol Rev 41:424–449

Bartels A, Sendlmeier WF, Rolfes M, Burkhardt F (2006) Emo-DB. http://pascal.kgw.tu-berlin.de/emodb/

Batliner A, Fischer K, Huber R, Spilker J, Nöth E (2000) Desperately seeking emotions or: Actors, wizards, and human beings. In: Proceedings of ISCA workshop on speech and emotion, Belfast, United Kingdom

Batliner A, Hacker C, Steidl S, Nöth E, D'Arcy S, Russell M, Wong LP (2004a) "You stupid tin box" – Children interacting with the AIBO robot: A cross-linguistic emotional speech corpus. In: International conference on language resources and evaluation (LREC), Lisbon, Portugal, pp 171–174

Batliner A, Hacker C, Steidl S, Nöth E, Haas J (2004b) User states, user strategies, and system performance: How to match the one with the other. In: International conference on language resources and evaluation (LREC), Lisbon, Portugal, pp 171–174

Batliner A, Steidl S, Hacker C, Nöth E, Niemann H (2005) Private emotions vs. social interaction – towards new dimensions in research on emotion. In: Carberry S, de Rosis F (eds) Proceedings of the workshop on adapting the interaction style to affective factors at UM'05, Edinburgh, United Kingdom

Baum LE, Eagon JA (1967) An inequality with applications to statistical estimation for probalistic functions of markov processes and to a model for ecology. Am Math Soc Bull 73:360–363

Baum LE, Petrie T, Soules G, Weiss N (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Ann Math Stat 41(1):164–171

Baum LE, Sell GR (1968) Growth functions for transformations on manifolds. Pac J Math 27:211–227

Beller G, Marty A (2006) Talkapillar: Outil d'analyse de corpus oraux. In: Rencontres Jeunes Rechercheurs RJC-ED268, Paris, France

Benyon D, Turner P, Turner S (2005) Designing interactive systems: People, activities, contexts, technologies. Addison-Wesley, Harlow, United Kingdom

Beringer N, Kartal U, Louka K, Schiel F, Türk U (2002) PROMISE – A procedure for multimodal interactive system evaluation. In: Proceedings of the LREC workshop on multimodal resources and multimodal systems evaluation, Las Palmas, Spain, pp 77–80

Bernsen NO, Dybkjær H, Dybkjær L (1994) Wizard of Oz prototyping: How and when? In: CCI working papers in cognitive science and HCI, WPCS-94-1. Centre for Cognitive Science, Roskilde University, Roskilde, Denmark

Boersma P (2001) Praat, a system for doing phonetics by computer. Glot Int 5(9/10):341–345

Boersma P (2002) Praat Website. http://www.praat.org/

Bogert BP, Healy MJR, Tukey JW (1963) The quefrency alanysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum, and Saphe cracking. In: Rosenblatt M (ed) Proceedings of the symposium on time series analysis, Providence, USA, pp 209–243

Bolinger D (1989) Intonation and its uses. Stanford University Press, Stanford, USA

Bosma W, André E (2004) Exploiting emotions to disambiguate dialogue acts. In: International conference on intelligent user interfaces (IUI), ACM Press, Funchal, Portugal, pp 85–92

Bourgeois J (2005) An LMS viewpoint on the local stability of second order blind source separation. In: IEEE workshop on statistical signal processing, Bordeaux, France, pp 1096–1101

Bourgeois J, Freudenberger J, Lathoud G (2005) Implicit control of noise canceller for speech enhancement. In: European conference on speech and language processing (EUROSPEECH), Lisbon, Portugal, pp 2065–2068

Braun A, Katerbow M (2005) Emotions in dubbed speech: An intercultural approach with respect to F0. In: International conference on speech and language processing (ICSLP), Lisbon, Portugal, pp 521–524

Brennan SE, Hulteen EA (1995) Interaction and feedback in a spoken language system: A theoretical framework. Knowl Syst 9:143–151

Brown P, Levinson SC (1987) Politeness – Some universals in language Use. Cambridge University Press, Cambridge, United Kingdom

Bühler D, Hamerich SW (2005) Towards VoiceXML compilation for portable embedded applications in ubiquitous environments. In: European conference on speech and language processing (EUROSPEECH), Lisbon, Portugal, pp 3397–4000

Bühler D, Riegler M (2005) Integrating dialogue management and domain reasoning. In: Proceedings of SPECOM, Patras, Greece, pp 409–412

Bui TH, Zwiers J, Poel M, Nijholt A (2006) Toward affective dialogue modeling using partially observable Markov decision processes. In: Proceedings of workshop emotion and computing, 29th annual German conference on artificial intelligence, Bremen, Germany

Burkhardt F, Ajmera J, Englert R, Stegmann J, Burleson W (2006a) Detecting anger in automated voice portal dialogs. In: International conference on speech and language processing (ICSLP), Pittsburgh, USA, pp 1053–1056

Burkhardt F, Audibert N, Malatesta L, Türk O, Arslan L, Aubergé V (2006b) Emotional prosody – Does culture make a difference? In: 3rd international conference on speech prosody, Dresden, Germany, pp 245–248

Burkhardt F, Paeschke A, Rolfes M, Sendlmeier WF, Weiss B (2005) A database of german emotional speech. In: European conference on speech and language processing (EUROSPEECH), Lisbon, Portugal, pp 1517–1520

Burton RR (1976) Semantic grammar: An engineering technique for constructing natural language understanding systems. Technical report 3353, Bolt Beranek & Newman, Inc, Cambridge, USA

Burzio L (1993) English stress, vowel length, and modularity. J Linguist 29(2):359–418

Busso C, Narayanan S (2006) Interplay between linguistic and affective goals in facial expression during emotional utterances. In: Proceedings of 7th international seminar on speech production (ISSP), Ubatuba, Brazil, pp 549–556

Campbell JP (1997) Speaker recognition: A tutorial. Proc IEEE 85(9):1437–1462

Campbell N (2000) Databases of emotional speech. In: Proceedings of ISCA workshop on speech and emotion, Belfast, United Kingdom

Campbell N (2002) The Recording of emotional speech – JST/CREST database research. In: International conference on language resources and evaluation (LREC), vol 6. Las Palmas, Spain, pp 2029–2032

Campbell N, Devillers L, Douglas-Cowie E, Aubergé V, Batliner A, Tao J (2006) Resources for the processing of affect in interactions. In: ELRA (ed) International conference on language resources and evaluation (LREC) Genova, Italy, pp 25–28

Cannon WB (1927) The James-Lange theory of emotion: A critical examination and an alternative theory. Am J Psychol 39:106–124

Carletta JC (1996) Assessing the reliability of subjective codings. Computat Ling 22(2):249–254

Carlson R, Granström B (1977) Speech synthesis. In: Hardcastle WJ, Laver J (eds) The handbook of phonetic sciences. Blackwell, Oxford, United Kingdom, pp 768–788

Carofiglio V, de Rosis F (2005) In favour of cognitive models of emotions. In: Workshop on mind-minding agents at AISB, Hatfield, United Kingdom

Carofiglio V, de Rosis F, Grassano R (2002) Mixed emotion modeling. In: Proceedings of the AISB02 symposium on animating expressive characters for social interactions, London, United Kingdom, pp 5–10. The Society for the Study of Artificial Intelligence and Simulation of Behaviour

Carofiglio V, de Rosis F, Grassano R (2003) Dynamic models of mixed emotion activation. In: Cañamero L, Aylett R (eds) Animating expressive characters for social interactions. John Benjamins, Amsterdam, The Netherlands

Cauldwell RT (2000) Where did the anger go? The role of context in interpreting emotion in speech. In: Proceedings of ISCA workshop on speech and emotion, Belfast, United Kingdom

Chateau N, Maffiolo V, Blouin C (2004) Analysis of emotional speech in voice mail messages: The influence of speakers' gender. In: International conference on speech and language processing (ICSLP), Jeju, Korea

Chu S-W, O'Neill I, Hanna P, McTear M (2005) An approach to multi-strategy dialogue management. In: International conference on speech and language processing (ICSLP), Lisbon, Portugal, pp 865–868

Chuang Z-J, Wu C-H (2004) Emotion recognition using acoustic features and textual content. In: Proceedings of the IEEE international conference on multimedia and expo (ICME), vol 1. Taipei, Taiwan, pp 53–56

Chung S (2000) L'expression et la perception de l'émotion extraite de la parole spontanée: Évidences du coréen et de l'anglais. PhD thesis, Université, de la Sorbonne Nouvelle, Paris III

Clavel C, Vasilescu I, Devillers L, Ehrette T (2004) Fiction Database for Emotion Detection in Abnormal Situations. In: International conference on speech and language processing (ICSLP), Jeju, Korea

Clavel C, Vasilescu I, Devillers L, Ehrette T, Richard G (2006) Fear-type emotions of the SAFE corpus: Annotation issues. In: International conference on language resources and evaluation (LREC), Genova, Italy, pp 1099–1104

Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Measure 20:37–40

Cohen MH, Giangola JP, Balogh J (2004) Voice user interface design. Addison-Wesley, Harlow, United Kingdom

Cohen PR (1995) Empirical methods for artificial intelligence. MIT Press, Boston, USA

Conati C, Gertner A, VanLehn K (2002) Using Bayesian networks to manage uncertainty in student modeling. J User Model User-Adapted Interact 12(4):371–417

Cornelius RR (2000) Theoretical approaches to emotion. In: Proceedings of ISCA workshop on speech and emotion, Belfast, United Kingdom

Cowie R (2000) Describing the emotional states expressed in speech. In: Proceedings of ISCA workshop on speech and emotion, Belfast, United Kingdom

Cowie R, Douglas-Cowie E (1996) Automatic statistical analysis of the signal and prosodic signs of emotion in speech. In: International conference on speech and language processing (ICSLP), vol 3. Philadelphia, USA, pp 1989–1992

Cowie R, Douglas-Cowie E, Apolloni B, Taylor J, Romano A, Fellenz W (1999a) What a neural net needs to know about emotion words. In: Proceedings of the 3rd IMACS international multiconference on circuits, systems, communications and computers (CSCC '99), Athens, Greece, pp 5311–5316

Cowie R, Douglas-Cowie E, Cox C, Cemgil AT (2004) D09: Final version of non-verbal speech parameter extraction module. ERMIS project IST-2000-29319

Cowie R, Douglas-Cowie E, Romano A (1999b) Changing emotional tone in dialogue and its prosodic correlates. In: ESCA tutorial and research workshop on dialogue and prosody, Veldoven, the Netherlands, pp 41–46

Cowie R, Douglas-Cowie E, Savvidou S, McMahon E, Sawey M, Schröder M (2000) "Feeltrace": An instrument for recording perceived emotion in real time. In: Proceedings of ISCA workshop on speech and emotion, Belfast, United Kingdom

Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias SD, Fellenz WA, Taylor JG (2001) Emotion recognition in human-computer interaction. IEEE Signal Proc Mag 12:32–80

Cowie R, Schröder M (2004) Piecing together the emotion jigsaw. In: Workshop on multimodal interaction and related machine learning algorithms (MLMI04), Martigny, Switzerland, pp 305–317

Craggs R (2004) Annotating emotion in dialogue – Issues and approaches. In: 7th annual CLUK research colloquium, Birmingham, United Kingdom

Craggs R, McGee Wood M (2003) Annotating emotion in dialogue. In: Proceedings of the 4th SIGdial workshop on discourse and dialogue, Sapporo, Japan

Cristiani N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge, United Kingdom

Crystal D (1969) Prosodic systems and intonation in English. Cambridge University Press, Cambridge, United Kingdom

d'Alessandro C, Mertens P (1995) Automatic pitch contour stylization using a model of tonal perception. Comput Speech Lang 9(3):257–288

Dan-Ning Jiang L-HC (2004) Classifying emotion in Chinese speech by decomposing prosodic features. In: International conference on speech and language processin (ICSLP), Jeju, Korea

Darwin C (1872) The expression of the emotions in man and animals. John Murray, London, United Kingdom

Dasarathy BV (1991) Nearest neighbor (NN) norms: NN pattern classification techniques. IEEE Computer Society Press, Los Alamitos, USA

De Mori R (1998) Spoken dialogue with computers. Academic Press, Orlando, USA

Dellaert F, Polzin T, Waibel A (1996) Recognizing emotion in speech. In: International conference on speech and language processing (ICSLP), Philadelphia, USA, pp 1970–1973

Deller Jr. JR, Proakis JG, Hansen JH (1993) Discrete-time processing of speech signals. Prentice Hall PTR, Upper Saddle River, USA

Delorme F, Lehuen J (2003) Dialog planning and domain knowledge modeled in terms of tasks and methods – A flexible framework for dialog managing. In: Proceedings of the international symposium on methodologies for intelligent systems, Maebashi City, Japan, pp 689–693

Descartes (1649) Les Passions de l'âme. Paris, France

Devillers L, Cowie R, Martin J-C, Douglas-Cowie E, Abrilian S, McRorie M (2006) Real life emotions in French and English TV video clips: An integrated annotation protocol combining continous and discrete approaches. In: International conference on language resources and evaluation (LREC), Genova, Italy, pp 1105–1110

Devillers L, Lamel L, Vasilescu I (2003) Emotion detection in task-oriented spoken dialogues. In: proceedings of the IEEE international conference on multimedia and expo (ICME), vol 3. Baltimore, USA, pp 549–552

Devillers L, Vasilescu I, Lamel L (2002) Annotation and detection of emotion in a task-oriented human-human dialog corpus. In: ISLE Workshop on dialogue tagging, Edinburgh, United Kingdom

Douglas-Cowie E, Cowie R, Schröder M (2000) A new emotion database: Considerations, sources and scope. In: proceedings of ISCA workshop on speech and emotion, Belfast, United Kingdom

Duff D, Gates B, LuperFoy S (1996) An architecture for spoken dialogue management. In: International conference on speech and language processing (ICSLP), vol 2. Philadelphia, USA, pp 1025–1028

Durbin R, Eddy SR, Krogh A, Mitchison G (1999) Biological sequence analysis: Probabilistic models of proteins and nucleic acids. Cambridge University Press, Cambridge, United Kingdom

Dybkjær L, Bernsen NO (2001) Usability evaluation in spoken language dialogue systems. In: Proceedings of the ACL workshop on evaluation methodologies for language and dialogue systems, Toulouse, France, pp 9–18

Ecma International (1999) ECMAScript language specification. ECMA Standard 262, 3rd Edition

Ecma International (2005) ECMAScript for XML (E4X) specification. ECMA Standard 357, 2nd Edition

Ekman P (1977) Biological and cultural contributions to body and facial movement. In: Blacking J (ed) The anthropology of the body. Academic, London, United Kingdom, pp 34–84

Ekman P (1992) An argument for basic emotions. Cogn Emot 6(3–4):169–200

Ekman P (1999) Basic emotions. In: Dalgleish T, Power M (eds) Handbook of cognition and emotion. John Wiley & Sons, Ltd., New York, USA, pp 301–320

Ekman P, Friesen WV, Ellsworth P (1982) What emotion categories or dimensions can observers judge from facial behaviour? In: Ekman P (ed) Emotion in the human face. Cambridge University Press, New York, USA, pp 39–55

Engberg IS, Hansen AV, Andersen O, Dalsgaard P (1997) Design, recoding and verification of a danish emotional speech database. In: European conference on speech and language processing (EUROSPEECH), Rhodes, Greece, pp 1695–1698

Fakotakis N (2004) Corpus design, recording and phonetic analysis of Greek emotional database. In: International conference on language resources and evaluation (LREC), Lisbon, Portugal, pp 1391–1394

Fallside F, Woods WA (1985) Computer speech processing. Prentice-Hall, Englewood Cliffs, USA

Fék M, Németh G, Olaszy G, Gordos G (2004) Design of a hungarian emotional database for speech analysis and synthesis. In: Proceedings of affective dialogue systems, tutorial and research workshop, ADS 2004, Kloster Irsee, Germany, pp 113–116

Feldman JA, Ballard DH (1982) Connectionist models and their properties. Cogn Sci 6(3):205–254

Fernandez R, Picard RW (2003) Modeling drivers' speech under stress. Speech Commun 40: 145–149

Fillmore CJ (1968) The case for case. In: Bach E, Harms R (eds) Universals in linguistic theory. Holt, Rhinehart and Winston, New York, USA, pp 1–88

Fiscus JG (1997) A post-processing system to yield reduced word error rates: Recogniser output voting error reduction (ROVER). In: Proceedings of the IEEE workshop on automatic speech recognition and understanding, Santa Barbara, USA, pp 347–352

Fleiss JL (1971) Measuring nominal scale agreement among many raters. Psychol Bull 76(5): 378–382

Forbes-Riley KM, Litman DJ (2004) Predicting emotion in spoken dialogue from multiple knowledge sources. In: Proceedings of the human language technology conference of the north american chapter of the association for computational linguistics (NAACL-HLT), pp 201–208

Fotinea S-E, Bakamidis S, Athanaselis T, Dologlou I, Carayannis G, Cowie R, Douglas-Cowie E, Fragopanagos NF, Taylor JG (2003) Emotion in speech: Towards an integration of linguistic, paralinguistic, psychological analysis. In: Joint international conference ICANN/ICONIP, Istanbul, Turkey, pp 1125–1132

Fox NA (1992) If it's not left it's right. Am Psychol 46:863–872

France DJ, Shiavi RG, Silverman S, Silverman M, Wilkes DM (2000) Acoustical properties of speech as indicators of depression and suicidal risk. IEEE Trans Biomed Eng 47(7):829–837

Frederking RE (1996) Grice's maxims: Do the right thing. In: Working notes of the AAAI'96 spring symposium on computational implicature: Computational approaches to interpreting and generating conversational implicature, Palo Alto, USA, pp 21–26

Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 55(1):119–139

Frijda NH (1970) Emotion and recognition of emotion. In: Arnold M (ed) Feelings and emotions. Academic, New York, USA

Frijda NH (1986) The emotions. Cambridge University Press, Cambridge, United Kingdom

Fujisawa T, Cook ND (2004) Identifying emotion in speech prosody using acoustical cues of harmony. In: International conference on speech and language processing (ICSLP), Jeju, Korea

Gharavian D, Ahadi SM (2005) The effect of emotion on Farsi speech parameters: A statistical evaluation. In: Proceedings of SPECOM, Patras, Greece, pp 463–466

Gibbon D, Lüngen H (1999) Consistent vocabularies for spoken language machine translation systems. In: Multilinguale Corpora. Codierung, Strukturierung, Analyse, Frankfurt, Germany, pp 169–178

Glass J, Goddeau D, Hetherington L, McCandless M, Pao C, Phillips M, Polifroni J, Seneff S, Zue V (1995) The MIT ATIS system: December 1994 progress report. In: Proceedings of the ARPA spoken language systems technology workshop, Austin, USA

Glass J, Seneff S (2003) Flexible and personalizable mixed-initiative dialogue systems. In: Proceedings of the HLT-NAACL 2003 workshop on research directions in dialogue processing, Edmonton, Canada

Gmytrasiewicz PJ, Lisetti CL (2000) Using decision theory to formalize emotions for multi-agent system applications: Preliminary report. In: 4th international conference on multi-agent systems, Boston, USA, pp 391–392

González-Bernal JA, López-López A, Muñoz-Arteaga J, Montes y Gómez M, Reyes-García CA, Villaseñor-Pineda L (2002) Natural language dialogue system for information retrieval. In: Proceedings of the international workshop on research and development of human communication technologies for conversational interaction and learning, Puebla, Mexico

Gorin AL, Riccardi G, Wright JH (1997) How may I help you? Speech Commun 23(1/2):113–127

Gratch J, Marcella S (2004) A domain-independent framework for modeling emotion. J Cogn Syst Res 5(4):269–306

Gray JA (1982) The neuropsychology of anxiety. Oxford University Press, Oxford, United Kingdom

Greasley P, Sherrard C, Waterman M (2000) Emotion in language and speech: Methodological issues in naturalistic approaches. Lang Speech 43:355–375

Grice HP (1975) Logic and conversation. In: Cole P, Morgan JL (eds) Syntax and semantics, Vol. 3: Speech Acts. Academic, New York, USA, pp 41–58

Gussenhoven C (2004) The phonology of tone and intonation. Cambridge University Press, Cambridge, United Kingdom

Gutiérrez-Arriola JM, Montero JM, Vallejo JA, Córdoba R, San-Segundo R, Pardo JM (2001) A new multi-speaker formant synthesizer that applies voice conversion techniques. In: European conference on speech and language processing (EUROSPEECH), Aalborg, Denmark, pp 357–360

Hagen E, Popowich F (2000) Flexible speech act based dialogue management. In: Dybkjær L, Hasida K Traum D (eds) Proceedings of the first SIGdial workshop on discourse and dialogue. Association for computational linguistics, Somerset, NJ, pp 131–140

Hajdinjak M, Mihelič F (2006) The Paradise evaluation framework: Issues and findings. Computat Linguist 32(2):263–272

Hansen JH, Bou-Ghazale SE, Sarikaya R, Pellom B (1998) Getting started with the SUSAS: Speech under simulated and actual stress database. Technical report RSPL-98-10, Robust Speech Processing Laboratory, Duke University, Durham, USA

Hassel L, Hagen E (2005) Adaptation of an automotive dialogue system to users' expertise. In: 6th SIGdial workshop on discourse and dialogue, Lisbon, Portugal, pp 222–226

Henton C (2005) Bitter pills to swallow. ASR and TTS have drug problems. Int J Speech Technol 8(3):247–257

Hermansky H (1990) Perceptual linear predictive (PLP) analysis of speech. J Acoust Soc Am 87(4):1738–1752

Hermansky H, Morgan N (1994) RASTA processing of speech. IEEE Trans Speech Audio Process 2(4):578–589

Hillard D, Hoffmeister B, Ostendorf M, Schlüter R, Ney H (2007) iROVER: Improving system combination with classification. In: Proceedings of the human language technology conference of the north American chapter of the association for computational linguistics (NAACL-HLT), Rochester, USA, pp 44–51

Hirose K, Sato K, Minematsu N (2004) Improvement in corpus-based generation of F_0 contours using generation process model for emotional speech synthesis. In: International conference on speech and language processing (ICSLP), Jeju, Korea

Holzapfel H, Fügen C, Denecke M, Waibel A (2002) Integrating emotional cues into a frame-
  work for dialogue management. In: proceedings of the international conference on multimodal
  interfaces, Pittsburgh, USA, pp 141–148

Howard RA (1971) Dynamic probabilistic systems, Vol 1: Markov chains. John Wiley & Sons,
  Ltd, New York, USA

Hunt A, McGlashan S (2004) Speech recognition grammar specification, Version 1.0. `http://
  www.w3.org/TR/speech-grammar/`

Hurtado LF, Griol D, Segarra E, Sanchis E (2006) A stochastic approach for dialog management
  based on neural networks. In: International conference on speech and language processing
  (ICSLP), Pittsburgh, USA, pp 49–52

Iida A, Campbell N, Higuchi F, Yasumura M (2003) A corpus-based speech synthesis system with
  emotion. Speech Commun 40:161–187

Ikehara S, Miyazaki M, Shirai S, Yokoo A, Nakaiwa H, Ogura K, et al (1999) A japanese lexicon.
  Iwanami Shoten, Tokyo, Japan

Ioannou SV, Raouzaiou AT, Tzouvaras VA, Mailis TP, Karpouzis KC, Kollias SD (2005) Emo-
  tion recognition through facial expression analysis based on a neurofuzzy network. Neural
  Networks 18:423–435

Iriondo I, Guaus R, Rodríguez A, Lázaro P, Montoya N, Blanco JM, Bernandas D, Oliver JM, Tena
  D, Longhi L (2000) Validation of an acoustical modelling of emotional expression in spanish
  using speech synthesis techniques. In: Proceedings of ISCA workshop on speech and emotion,
  Belfast, United Kingdom

Isard SD, Miller DA (1986) Diphone synthesis techniques. In: Proceedings of IEE international
  conference on speech input/output, London, United Kingdom, pp 77–82

Ito A, Shimada K, Suzuki M, Makino S (2006) A user simulator based on VoiceXML for evaluation
  of spoken dialog systems. In: International conference on speech and language processing
  (ICSLP), Pittsburgh, USA, pp 1045–1048

ITU-T P.supp24 (2005) P-series recommendation – Supplement 24. Parameters describing the
  interaction with spoken dialogue systems

Iwai A, Yano Y, Okuma S (2004) Complex emotion recognition system for a specific user using
  SOM based on prosodic features. In: International conference on speech and language process-
  ing (ICSLP), Jeju, Korea

Izard CE (1971) The face of emotion. Appleton century crofts, New York, USA

James W (1884) What is an emotion? Mind, 9:188–205

Jelinek F, Lafferty J, Magerman DM, Ratnaparkhi A, Roukos S. (1994) Decision tree parsing
  using a hidden derivation model. In: Proceedings of the ARPA workshop on human language
  technology, pp 260–265

Jelinek F, Lafferty J, Mercer R (1992) Basic methods of probabilistic context free grammars.
  Speech recognition and understanding. Recent advances, 75:345–360

Jiang X, Tian L, Han M (2005) Separability and recognition of emotion states in multilingual
  speech. In: Proceedings of the international conference on communications, circuits and sys-
  tems, vol 2. Hong Kong, China, pp 861–864

Johannesson R (1988) Informationsteori – Grundvalen för telekommunikation. Studenlitteratur
  AB, Lund, Sweden

Johnston M, Bangalole S, Vasireddy G, Stent A, Eblen P, Walker MA, Whittaker S, Maloor P
  (2002) MATCH: An architecture for multimodal dialogue systems. In: Proceedings of the 40th
  annual meeting of the association for computational linguistics (ACL2002), Philadelphia, USA,
  pp 376–383

Johnstone T, Scherer KR (1999) The effects of emotions on voice quality. In: International congress
  of phonetic sciences (ICPhS), San Francisco, USA, pp 2029–2032

Jokinen K, Kerminen A, Kaipainen M, Jauhiainen T, Wilcock G, Turunen M, Hakulinen J, Kuusisto
  J, Lagus K (2002) Adaptive dialogue systems – interaction with interact. In: Proceedings of
  the 3rd SIGdial workshop on discourse and dialogue, Philadelphia, USA, pp 64–73

Jönsson A, Dahlbäck N (2000) Distilling dialogues – A method using natural dialogue corpora for dialogue systems development. In: 6th applied natural language processing conference (ANLP), Seattle, USA, pp 44–51

Juang B-H, Rabiner LR (1991) Hidden markov models for speech recognition. Technometrics 33(3):251–272

Jun S-A (2005) Prosodic typology. The phonology of intonation and phrasing. Oxford University Press, New York, USA

Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans ASME J Basic Eng 82:35–45

Kamm C, Narayanan S, Dutton D, Ritenour R (1997) Evaluating spoken dialog systems for telecommunication systems. In: European conference on speech and language processing (EUROSPEECH), Rhodes, Greece, pp 22–25

Karlsson I (1999) Within-speaker variability in the VeriVox database. In: Proceedings of the twelfth swedish phonetics conference (Fonetik 99), number 81 in Gothenburg Papers in Theoretical Linguistics, Gothenburg, Sweden, pp 93–96

Kehrein R (2001) Linguistische und psychologische Aspekte der Erforschung des prosodischen Emotionsausdrucks. In: Germanische Linguistik (GL), 157–158:91–123

Kim EH, Hyun KH, Kwak YK (2005) Robust emotion recognition feature, frequency range of meaningful signal. In: IEEE international workshop on robots and human interactive communication (RO-MAN), Nashville, USA, pp 667–671

Kim I-S (2006) Automatic speech recognition: Reliability and pedagogical implications for teaching pronunciation. Educ Technol Soc 9(1):322–334

Kim J, André E (2006) Emotion recognition using physiological and speech signal in short-term observation. In: Tutorial and research workshop on perception and interactive technologies (PIT06), Irsee, Germany, pp 53–64

Kim KH, Bang SH, Kim SR (2004a) Emotion recognition system using short-term monitoring of physiological signals. Med Biol Eng Comput 42:419–427

Kim S-J, Kim K-K, Hahn M (2004b) Study on emotional speech features in Korean with its application to voice color conversion. In: International conference on speech and language processing (ICSLP), Jeju, Korea

Kipp M (2001) ANVIL – A generic annotation tool for multimodal dialogue. In: European conference on speech and language processing (EUROSPEECH), Aalborg, Denmark, pp 1367–1370

Klasmeyer G (1996) Perceptual cues for emotional speech. In: Workshop on the auditory basis of speech perception, Keele, United Kingdom, pp 154–157

Klasmeyer G, Johnstone T, Bänzinger T, Sappok C, Scherer KR (2000) Emotional voice variability in speaker verification. In: Proceedings of ISCA workshop on speech and emotion, Belfast, United Kingdom

Kleinginna PR, Kleinginna AM (1981) A categorized list of emotion definitions, with suggestions for a consensual definition. Motivat Emot 5(4):345–359

Kochanski G, Grabe E, Coleman J, Rosner B (2005) Loudness predicts prominence: Fundamental frequency lends little. J Acoust Soc Am 118(2):1038–1054

Koike K, Suzuki H, Saito H (1998) Prosodic parameters in emotional speech. In: International conference on speech and language processing (ICSLP), Sydney, Australia, pp 679–682

Komatani K, Kawahara T (2000) Flexible mixed-initiative dialogue management using concept-level confidence measures of speech recognizer output. In: Proceedings of the 18th conference on computational linguistics, Saarbrücken, Germany, pp 467–473

Krahmer E, Landsbergen J, Pouteau X (1997) How to obey the 7 commandments for spoken dialogue systems. In: Proceedings of the (E)ACL workshop on interactive spoken dialog systems, Madrid, Spain, pp 82–89

Kshirsagar S, Magnenat-Thalmann N (2002) A Multilayer Personality Model. In: Proceedings of the 2nd international symposium on smart graphics, Hawthorne, USA, pp 107–115

Kumar R, Rosé CP, Litman DJ (2006) Identification of confusion and surprise in spoken dialog using prosodic features. In: International conference on speech and language processing (ICSLP), Pittsburgh, USA, pp 1842–1845

Kwon O-W, Chan K, Hao J, Lee T-W (2003) Emotion recognition by speech signals. In: European conference on speech and language processing (EUROSPEECH), Geneva, Switzerland, pp 125–128

Ladd DR (1996) Intonational phonology. Cambridge University Press, Cambridge, United Kingdom

Lamere P, Kwok P, Walker W, Gouvea E, Singh R, Raj B, Wolf P (2003) Design of the CMU Sphinx-4 decoder. In: European conference on speech and language processing (EURO-SPEECH), Geneva, Switzerland, pp 1191–1184

Lange CG (1885) Om Sindsbevægelser. Et psyko-fysiologisk Studie. Rasmussen, Copenhagen, Denmark

Larson JA (2001) VoiceXML 2.0 and the W3C Speech Interface Framework. In: IEEE Workshop on automatic speech recognition and understanding (ASRU), Madonna di Campiglio, Italy, pp 5–8

Larson JA (2002) VoiceXML: Introduction to developing speech applications. Prentice Hall, Upper Saddle River, USA

Larsson S (2000) Godis Demo. http://www.ling.gu.se/leifg/WebDemos/godis2.0a/

Larsson S, Berman A, Grönqvist L, Kronlid F (2002) TRINDIKIT 3.0 Manual, D6.4, Siridus Project

Larsson S, Traum D (2000) Information state and dialogue management in the TRINDI dialogue move engine toolkit. Nat Lang Eng 6:323–340

Laskowski K, Burger S (2006) Annotation and analysis of emotionally relevant behavior in the ISL meeting corpus. In: International conference on language resources and evaluation (LREC), Genova, Italy, pp 1111–1116

Laukka P (2004) Vocal expression of emotion. PhD thesis, Uppsala University

Lazarus RS (1982) Thoughts on the relations between emotion and cognition. Am Psychol 37:1019–1024

Lazarus RS (1991) Emotion and adaptation. Oxford University Press, New York, USA

Lazarus RS, Smith CA (1988) Knowledge and appraisal in the cognition-emotion relationship. Cogn Emot 2:281–300

Lecœuche R (2001) Learning optimal dialogue management rules by using reinforcement learning and inductive logic programming. In: Proceedings of the 2nd meeting of the NAACL, Pittsburgh, USA

Lee CM, Narayanan S (2003) Emotion recognition using a data-driven fuzzy interference system. In: European conference on speech and language processing (EUROSPEECH), Geneva, Switzerland, pp 157–160

Lee CM, Yildirim S, Bulut M, Kazemzadeh A, Busso C, Deng Z, Lee S, Narayanan S (2004) Emotion recognition based on phoneme classes. In: International conference on speech and language processing (ICSLP), Jeju, Korea

Lee S, Bresch E, Adams J, Kazemzadeh A, Narayanan S (2006) A study of emotional speech articulation using a fast magnetic resonance imaging technique. In: International conference on speech and language processing (ICSLP), Pittsburgh, USA, pp 2234–2237

Lee S, Yildirim S, Kazemzadeh A, Narayanan S (2005) An articulatory study of emotional speech production. In: International conference on speech and language processing (ICSLP), Lisbon, Portugal, pp 497–500

Lefèvre F Gauvain J-L Lamel L (2001) Improving genericity for task-independent speech recognition. In: European conference on speech and language processing (EUROSPEECH), Aalborg, Denmark, pp 1241–1244

Lesher GW, Moulton BJ, Higginbotham DJ (1999) Effects of ngram order and training text size on word prediction. In: Proceedings of the RESNA 1999 annual conference, Long Beach, USA, pp 52–54

Levenshtein VI (1966) Binary codes capable of correcting deletions, insertions, reversals. Sov Phys Doklady 10(8):707–710

Levin E, Narayanan S, Pieraccini R, Biatov K, Bocchieri E, Di Fabbrizio G, Eckert W, Lee S, Pokrovsky A, Rahim M, Ruscitti P, Walker M (2000a) The AT&T DARPA communicator

mixed-initiative spoken dialog system. In: International conference on speech and language processing (ICSLP), Beijing, China, pp 122–125

Levin E, Pieraccini R (1995) "CHRONUS", the next generation. In: Proceedings of the DARPA Speech and Natural Language Workshop, Austin, USA, pp 269–271

Levin E, Pieraccini R, Eckert W (2000b) A Stochastic Model of Human Machine Interaction for Learning Dialog Strategies. IEEE Trans Speech Audio Process 8(1):11–23

Li J, Najmi A, Gray RM (2000) Image classification by a two-dimensional hidden Markov model. IEEE Trans Signal Process 48(2):517–533

Li Y, Zhao Y (1998) Recognizing Emotions in Speech Using Short-term and Long-term Features. In: International conference on speech and language processing (ICSLP), Sydney, Australia, pp 2255–2258

Linhard K, Haulick T (1999) Noise subtraction with parametric recursive gain curves. In: European conference on speech and language processing (EUROSPEECH), Budapest, Hungary, pp 2611–2614

Lippmann RP (1997) Speech recognition by machines and humans. Speech Commun 22(1):1–15

Liscombe J, Riccardi G, Hakkani-Tür D (2005) Using context to improve emotion detection in spoken dialog systems. In: International conference on speech and language processing (ICSLP), Lisbon, Portugal, pp 1845–1848

Liscombe J, Venditti J, Hirschberg J (2003) Classifying subject ratings of emotional speech using acoustic features. In: Proceedings of the 8th European conference on speech communication and technology, Geneva, Switzerland, pp 725–728

Litman DJ, Kearns MS, Singh SB, Walker MA (2000) Automatic optimization of dialogue management. In: Proceedings of the 17th conference on computational linguistics, Saarbrücken, Germany, pp 502–508

Litman DJ, Pan S (2002) Designing and evaluating an adaptive spoken dialogue system. User Model User-Adapted Interact 12:111–137

Lopes Rodrigues LM, Carvalho M (2004) Emotional and motivational ITS architecture. In: Proceedings of the IEEE international conference on advanced learning techniques (ICALT), Joensuu, Finland, pp 276–280

López-Jaquero V, Montero F, Fernández-Caballero A, Lozano MD (2005) Towards adaptive user interfaces generation. In: Camp O, Filipe JBL, Hammoudi S, Piattini M, (eds) Enterprise systems V. Springer, the Netherlands, pp 226–232

Luengo I, Navas E, Hernáez I, Sánchez J (2005) Automatic emotion recognition using prosodic parameters. In: International conference on speech and language processing (ICSLP), Lisbon, Portugal, pp 493–496

LuperFoy S, Loehr D, Duff D, Miller K, Reeder F, Harper L (1998) An architecture for dialogue management, context tracking, pragmatic adaptation in spoken dialogue systems. In: Proceedings of the 17th international conference on computational linguistics, Montreal, Canada, pp 794–801. Association for Computational Linguistics

Ma X, Lee H, Bird S, Maeda K (2002) Models and tools for collaborative annotation. In: International conference on language resources and evaluation (LREC), Las Palmas, Spain, pp 2066–2073

Macherey K, Ney H (2003) Features for tree based dialogue management. In: European conference on speech and language processing (EUROSPEECH), Geneva, Switzerland, pp 601–604

Makarova V, Petrushin VA, Ruslana VA (2002) A database of russian emotional utterances. In: International conference on speech and language processing (ICSLP), Denver, USA, pp 2041–2044

Makhoul J (1975) Linear prediction: A tutorial review. Proc IEEE 63(4):561–580

Marin J-M, Mengersen K, Robert CP (2005) Bayesian modelling and inference on mixtures of distributions. In: Dey D, Dao CR (eds) Handbook of statistics, vol 25. Elsevier-Sciences, Amsterdam, the Netherlands

Markov AA (1907) Extension of the limit theorems of probability theory to a sum of variables connected in a chain (in Russian). Bulletin of the imperial academy of sciences, XXII. reprinted and translated to English in Appendix B of Howard (1971)

Martin J-C, Caridakis G, Devillers L, Karpouzis K, Abrilian S (2006) Manual annotation and automatic image processing of multimodal emotional behaviours: Validating the annotation of TV interviews. In: International conference on language resources and evaluation (LREC), Genova, Italy, pp 1127–1132

Martinovski B, Traum D (2003) Breakdown in human-machine interaction: The error is the clue. In: Proceedings of the ISCA tutorial and research workshop on Error handling in dialogue systems, pp 11–16

Masters T (1994) Signal and image processing with neural networks. John Wiley & Sons, Ltd, New York, USA

Matsumoto K, Minato J, Ren F, Kuroiwa S (2005) Estimating human emotions using wording and sentence patterns. In: Proceedings of the IEEE international conference on information acquisition, Hong Kong and Macau, China, pp 421–426

Matsunaga S, Sakaguchi S, Yamashita M, Miyahara S, Nishitani S, Shinohara K (2006) Emotion detection in infants' cries based on a maximum likelihood approach. In: International conference on speech and language processing (ICSLP), Pittsburgh, USA, pp 1834–1837

McCulloch W, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 7:115–133

McDougall W (1926) An introduction to social psychology. Luce, Boston, USA

McGilloway S, Cowie R, Douglas-Cowie E (1995) Prosodic signs of emotion in speech: Preliminary results from a new technique for automatic statistical analysis. In: International congress of phonetic sciences (ICPhS), vol 1. Stockholm, Sweden, pp 250–253

McGilloway S, Cowie R, Douglas-Cowie E, Gielen S, Westerdijk M, and Stroeve S (2000) Approaching automatic recognition of emotion from voice: A rough benchmark. In: Proceedings of ISCA workshop on speech and emotion, Belfast, United Kingdom

McMahon E, Cowie R, Kasderidis S, Taylor JG, Kollias S (2003) What chance that a DC could recognise hazardous mental states from sensor outputs? In: DC Tales, Santorini, Greece

McTear MF (2004) Spoken dialogue technology – Toward the conversational user interface. Springer, London, United Kingdom

Meng H, Pittermann J, Pittermann A, Minker W (2007) Combined speech-emotion recognition for spoken human-computer interfaces. In: IEEE international conference on signal processing and communications (ICSPC), Dubai, United Arab Emirates

Miceli M, de Rosis F, Poggi I (2006) Emotional and non-emotional persuasion. Appl Artif Intell Int J 20(10):849–879

Miller S, Bobrow R, Ingria R, Schwartz R (1994) Hidden understanding models of natural language. In: Proceedings of the 32nd annual meeting of the association for computational linguistics (ACL1994), Las Cruces, USA, pp 25–32

Minker W, Albalate A, Bühler D, Pittermann A, Pittermann J, Strauss P-M, Zaykovskiy D (2006a) Recent trends in spoken language dialogue systems. In: ITI 4th international conference on information and communications technology (ICICT 2006), Cairo, Egypt

Minker W, Pittermann J, Pittermann A, Strauss P-M, Bühler D (2006b) Next-generation human-computer interfaces – Towards intelligent, adaptive and proactive spoken language dialogue systems. In:2nd IEE international conference on intelligent environments, Athens, Greece

Minker W, Waibel A, Mariani J (1999) Stochastically-based semantic analysis, vol 514 of The Kluwer international series in engineering and computer science. Kluwer, Boston, USA

Miwa H, Umetsu T, Takanishi A, Takanobu H (2000) Robot personalization based on the mental dynamics. In: IEEE/RSJ conference on intelligent robots and systems, vol 1, pp 8–14

Montero JM, Gutiérrez-Arriola J, Colás J, Macías-Guarasa J, Enríquez E, Pardo JM (1999) Development of an emotional speech synthesiser in spanish. In: European conference on speech and language processing (EUROSPEECH), Budapest, Hungary, pp 2099–2102

Mowrer OH (1960) Learning theory and behaviour. John Wiley & Sons, Ltd, New York, USA

Mozziconacci S. JL, Hermes DJ (1999) Role of intonation patterns in conveying emotion in speech. In: International congress of phonetic sciences (ICPhS), San Francisco, USA, pp 2001–2004

Nasoz F, Alvarez K, Lisetti CL, Finkelstein N (2003) Emotion recognition from physiological signals for presence technologies. Int J Cogn Technol Work 6(1):4–14

Nass C, Lee KM (2001) Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, consistency-attraction. J Exper Psychol Appl 7(3):171–181

Navas E, Castelruiz A, Luengo I, Sánchez J, Hernáez I. (2004a) Designing and recording an audiovisual database of emotional speech in Basque. In: International conference on language resources and evaluation (LREC), Lisbon, Portugal, pp 1387–1390

Navas E, Hernáez I, Castelruiz A, Luengo I (2004b) Obtaining and evaluating an emotional database for prosody modelling in standard Basque. In: Sojka P Kopeček I Pala K (eds) Proceedings of the 7th international conference on text, speech and dialogue (TSD). Springer, Brno, Czech Republic, pp 393–400

Navas E, Hernáez I, Luengo I (2006) An objective and subjective study of the role of semantics and prosodic features in buidling corpora for emotional TTS. IEEE Trans Audio Speech Lang Proces 14(4):1117–1127

Neiberg D, Elenius K, Laskowski K (2006) Emotion recognition in spontaneous speech using GMMs. In: International conference on speech and language processing (ICSLP), Pittsburgh, USA, pp 809–812

Nicholson J, Takahashi K, Nakatsu R (1999) Emotion recognition in speech using neural networks. In: Proceedings of the 6th international conference on neural information processing (ICONIP), vol 2. Perth, Australia, pp 495–501

Nisimura R, Omae S, Kawahara H, Irino T (2006) Analyzing dialogue data for real-world emotional speech classification. In: International conference on speech and language processing (ICSLP), Pittsburgh, USA, pp 1822–1825

Nogueiras A, Moreno A, Bonafonte A, no J. BM (2001) Speech emotion recognition using hidden Markov models. In: European conference on speech and language processing (EUROSPEECH), Aalborg, Denmark, pp 2679–2682

Nuttall AH (1981) Some windows with very good sidelobe behavior. IEEE Trans Acoust Speech Signal Proces ASSP-29(1):84–91

Nwe TL, Wei FS, De Silva LC (2001) Speech based emotion classification. In: Proceedings of the IEEE region 10 international conference on electrical and electronic technology (TENCON), vol 1. Phuket Island, Singapore, pp 297–301

Oatley K, Jenkins JM (1996) Understanding emotions. Blackwell, Oxford, United Kingdom

Oatley K, Johnson-Laird PN (1987) Towards a cognitive theory of emotions. Cogn Emot 1:29–50

Oatley K, Johnson-Laird PN (1995) Communicative theory of emotions: Empirical tests, mental models & implications for social interaction. In: Martin LL, Tesser A (eds) Goals and affect. Erlbaum, Hillsdale, USA

O'Brien D, Monaghan A. IC (2001) Concatenative synthesis based on a harmonic model. IEEE Trans Speech Audio Process, 9(1):11–20

Okada N, Inui K, Tokuhisa M (1999) Towards affective integration of vision, behavior, speech processing. In: Proceedings of integration of speech and image understanding, Corfu, Greece, pp 49–77

O'Malley MH (1990) Text-to-speech conversion technology. Computer 23(8):17–23

O'Neill I, Hanna P, Liu X, McTear M (2003) The queen's communicator: An object-oriented dialogue manager. In: European conference on speech and language processing (EUROSPEECH), Geneva, Switzerland, pp 593–596

Oppenheim AV, Schafer RW (2004) From frequency to quefrency: A history of the cepstrum. IEEE Signal Proces Mag 21(5):95–106

Ortony A, Clore G, Foss M (1987) The referential structure of the affective lexicon. Cogn Sci 11:341–346

Ortony A, Clore GL, Collins A (1988) The cognitive structure of emotions. Cambridge University Press, Cambridge, United Kingdom

Ortony A, Turner TJ (1990) What's basic about basic emotions. Psychol Rev 97(3):315–331

Osgood CE, Suci GJ, Tanenbaum PH (1957) The measurement of meaning. University of Illinois Press, Urbana, USA

Oudeyer P-Y (2003) The production and recognition of emotions in speech: Features and algorithms. Int J Hum-Comput St 59(1-2):157–183

Oviatt SL (1997) Multimodal interactive maps: Designing for human performance. Hum-Comput Int (Special Issue on Multimodal Interfaces) 12:93–129

Oviatt SL (2000) Taming recognition errors with a multimodal interface. Commun ACM 43(9): 45–51

Page L, Brin S, Motwani R, Winograd T (1998) The PageRank citation ranking: Bringing order to the Web. Technical report, Stanford digital library technologies project

Panksepp J (1982) Toward a general psychobiological theory of emotions. Behav Brain Sci 5:407–467

Pao T-L, Chen Y-T, Yeh J-H, Lu J-J (2004) Detecting emotions in mandarin speech. In: Proceedings of the 16th conference on computational linguistics and speech processing ROCLING, Taipei, Taiwan

Park C-H, Si K-B (2003) Emotion recognition and acoustic analysis from speech signal. In: Proceedings of the international joint conference on neural networks (IJCNN), vol 4. Portland, USA, pp 2594–2598

Parunak H. VD, Bisson R, Brueckner S, Matthews R, Sauter J (2006) A model of emotions for situated agents. In: Proceedings of the fifth international joint conference on autonomous agents and multiagent systems, Hakodate, Japan

Pereira C (2000) Dimensions of emotional meaning in speech. In: Proceedings of ISCA workshop on speech and emotion, Belfast, United Kingdom

Peter C, Beale R (eds) (2008) Affect and emotion in human-computer interaction: From theory to applications. (Lecture notes in computer science) Springer, Berlin, Germany

Peter C, Herbon A (2006) Emotion representation and physiology assignments in digital systems. Interact Comput 18(2):139–170

Peter G, Rösner D (1994) User-model-driven generation of instructions. User Model User-Adapted Interact 3(4):289–319

Picard RW (2000a) Affective computing. The MIT Press, Cambridge, USA

Picard RW (2000b) Toward computers that recognize and respond to user emotion. IBM Syst J 39(3,4):705–718

Pierrehumbert J, Hirschberg J (1990) The meaning of intonational contours in the interpretation of discourse. In: Cohen PR, Morgan J, Pollack ME (eds) Intentions in communication. MIT Press, Cambridge, USA, pp 271–311

Pittermann A, Pittermann J (2006a) Getting bored with HTK? Using HMMs for emotion recognition. In: 8th international conference on signal processing (ICSP), vol 1. Guilin, China, pp 704–707

Pittermann J, Minker W, Pittermann A, Bühler D (2007a) ProblEmo – Problem solving and emotion awareness in spoken dialogue systems. In: 3rd IET international conference on intelligent environments, Ulm, Germany

Pittermann J, Pittermann A (2006b) A post-processing approach to improve emotion recognition rates. In: 8th international conference on signal processing (ICSP), vol 1. Guilin, China, pp 708–711

Pittermann J, Pittermann A (2006c) An 'emo-statistical' model for flexible dialogue management. In: 8th international conference on signal processing (ICSP), vol 2. Guilin, China, pp 1599–1602

Pittermann J, Pittermann A (2006d) Integrating emotion recognition into an adaptive spoken language dialogue system. In: 2nd IET international conference on intelligent environments, vol 1. Athens, Greece, pp 197–202

Pittermann J, Pittermann A (2007) A data-oriented approach to integrate emotions in adaptive dialogue management. In: International conference on intelligent user interfaces (IUI), Honolulu, USA, pp 270–273

Pittermann J, Pittermann A, Meng H, Minker W (2007b) Towards an emotion-sensitive spoken dialogue system – Classification and dialogue modeling. In: 3rd IET international conference on intelligent environments, Ulm, Germany

Pittermann J, Pittermann A, Minker W (2007c) Design and implementation of adaptive dialogue strategies for speech-based interfaces. J Ubiquitous Comput Intell 1(2):145–152

Pittermann J, Pittermann A, Schmitt A, Minker W (2008a) Comparing evaluation criteria for (automatic) emotion recognition. In: 4th IET international conference on intelligent environments, Seattle, USA

Pittermann J, Rittinger A, Minker W (2005) Flexible dialogue management in intelligent human-machine interfaces. In: The IEE international workshop on intelligent environments, University of Essex, Colchester, United Kingdom

Pittermann J, Schmitt A, Fawzy El Sayed N (2008b) Integrating linguistic cues into speech-based emotion recognition. In: 4th IET international conference on intelligent environments, Seattle, USA

Plutchik R (1980a) A generalpsychorevolutionary theory of emotion. In: Plutchik R, Kellerman H (eds) Emotion: Theory, research, experience: Vol. 1. Theories of emotion. Academic, New York, USA, pp 3–31

Plutchik R (1980b) Emotion: A psychorevolutionary synthesis. Harper & Row, New York, USA

Plutchik R (1994) The psychology and biology of emotion. Harper Collins College Publishers, New York, USA

Polifroni J, Chung G (2002) Promoting portability in dialogue management. In: International conference on speech and language processing (ICSLP), Denver, USA, pp 2721–2724

Polzin TS, Waibel A (1998) Detecting emotions in speech. In: Proceedings of the CMC, Tilburg, The Netherlands

Polzin TS, Waibel A (2000) Emotion-sensitive human-computer interfaces. In: ITRW on speech and emotion, ISCA, pp 201–206

Potapova R, Potapov V (2005) Identification of prosodic features of emotional state of a speaker. In: Proceedings of SPECOM, Patras, Greece, pp 25–32

Power M, Dalgleish T (1997) Cognition and emotion: From order to disorder. Pschology Press, Hove, United Kingdom

Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 77(2):257–286

Rank E, Pirker H (1998) Generating emotional speech with a concatenative synthesizer. In: International conference on speech and language processing (ICSLP), Sydney, Australia

Rao AS, Georgeff MP (1991) Modeling agents within a BDI architecture. In: Proceedings of the 2nd international conference on principles of knowledge representation and reasoning (KR-91), pp 473–484

Razak AA, Komiya R, Abidin M. IZ (2005) Comparison between fuzzy and NN method for speech emotion recognition. In: Proceedings of the 3rd IEEE international symposium on information technology and applications (ICITA), vol 1. Sydney, Australia, pp 297–302

Razak AA, Yusof M. HM, Komiya R (2003) Towards automatic recognition of emotion in speech. In: Proceedings of the 3rd IEEE international symposium on signal processing and information technology (ISSPIT), Darmstadt, Germany, pp 548–551

Reeves B, Nass C (1996) The media equation: How people treat computers, television, new media like real people and places. Cambridge University Press, Cambridge, United Kingdom

Reidsma D, Heylen D, Ordelman R (2006) Annotating emotions in meetings. In: International conference on language resources and evaluation (LREC), Genova, Italy, pp 1117–1122

Reiter E, Dale R (2000) Building natural language generation systems. Cambridge University Press, Cambridge, United Kingdom

Reza FM (1961) An introduction to information theory. McGraw-Hill, New York, USA

Rigoll G, Müller R, Schuller B (2005) Speech emotion recognition exploiting acoustic and linguistic information sources. In: Proceedings of SPECOM, Patras, Greece, pp 61–67

Ripley BD (1996) Pattern recognition and neural networks. Cambridge University Press, Cambridge, United Kingdom

Roark B, Saraclar M, Collins M (2007) Discriminative n-gram language modeling. Comput Speech Lang 21(2):373–392

Robinson T (1997a) Speech analysis – Lecture notes, lent term. Cambridge University, Cambridge, (Online tutorial)

Robinson T (1997b) The British English example pronunciation dictionary

Robson J, Mackenzie-Beck J (1999) Hearing smiles – Perceptual, acoustic and production aspects of labial spreading. In: International congress of phonetic sciences (ICPhS), San Francisco, USA, pp 219–222

Roseman IJ (1979) Cognitive aspects of emotion and emotional behaviour. In: 87th annual convention of the American Psychological Association, New York, USA

Roseman IJ, Spindel MS, Jose PE (1990) Appraisals of emotion-eliciting events: Testing a theory of discrete emotions. J Personality Soc Psychol 59:899–915

Rossi G, Schwabe D, Guimarães R (2001) Designing personalized Web applications. In: Proceedings of the 10th international World Wide Web conference, Hong Kong, China, pp 275–284

Rotaru M, Litman DJ (2005) Using word-level pitch features to better predict student emotions during spoken tutoring dialogues. In: International conference on speech and language processing (ICSLP), Lisbon, Portugal, pp 1845–1848

Rotaru M, Litman DJ, Forbes-Riley K (2005) Interactions between speech recognition problems and user emotions. In: International conference on speech and language processing (ICSLP), Lisbon, Portugal, pp 2481–2484

Roy N, Pineau J, Thrun S (2000) Spoken dialogue management using probabilistic reasoning. In: Proceedings of the 38th annual meeting of the association for computational linguistics (ACL2000), Hong Kong, China

Russell JA (1980) A circumplex model of affect. J Personality Soc Psychol 39:1161–1178

Russell JA, Mehrabian A (1977) Evidence for a three-factor theory of emotions. J Res Personality 11:273–294

Sagisaka Y (2001) Speech synthesis. J ASJ 57(1):11–20

Sakoe H, Chiba S (1978) Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans Acoust Speech Signal Proces (ASSP) 26(1):43–49

Saratxaga I, Navas E, Hernáez I, Luengo I (2006) Designing and recording an emotional speech database for corpus based synthesis in Basque. In: International conference on language resources and evaluation (LREC), Genova, Italy, pp 2126–2129

Schachter S, Singer JE (1962) Cognitive, social, physiological determinants of emotional state. Psychol Rev 69:379–399

Scherer KR (1988) Criteria for emotion-antecedent appraisal: A review. In: Hamilton V, Frijda NH (eds) Cognitive perspectives on emotion and motivation. Kluwer, Dordrecht, the Netherlands, pp 89–126

Scherer KR (2000) Psychological models of emotion. In: Borod JC (ed) The neuropsychology of emotion. Oxford University Press, New York, USA, pp 137–162

Scherer KR, Bänziger T (2004) Emotional expression in prosody: A review and an agenda for future research. In: Proceedings of speech prosody 2004, Nara, Japan, pp 359–366

Scherer KR, Ceschi G (1997) Lost luggage: A field study of emotion-antecedent appraisal. Motivat Emot 21(3):211–235

Schiel F, Steininger S, Türk U (2002) The SmartKom multimodal corpus at BAS. In: International conference on language resources and evaluation (LREC), Las Palmas, Spain

Schlosberg H (1954) Three dimensions of emotion. Psychol Rev 61(2):81–88

Schmidt G, Haulick T (2006) Signal processing for in-car communication systems. Signal Proces 86(6):1307–1326

Schröder M (2000) Experimental study of affect bursts. In: Proceedings of ISCA workshop on speech and emotion, Belfast, United Kingdom

Schröder M (2004) Speech and emotion research: An overview of research frameworks and a dimensional approach to emotional speech synthesis. PhD thesis, Saarland University

Schröder M, Cowie R (2006) Developing a consistent view on emotion-oriented computing. In: Renals S, Bengio S (eds) MLMI 2005, LNCS 3869. Springer, Heidelberg, Germany, pp 194–205

Schroeder MR, Atal BS, Hall JL (1979) Optimizing digital speech coders by exploiting masking properties of the human ear. J Acoust Soc Am 66(6):1647–1652

Schuller B, Müller R, Lang M, Rigoll G (2005) Speaker independent emotion recognition by early fusion of acoustic and linguistic features within ensembles. In: International conference on speech and language processing (ICSLP), Lisbon, Portugal, pp 805–808

Seneff S, Wang C, Chung G (2004) A flexible mixed-initiative speech interface for restaurant information. In: 5th SIGdial workshop on discourse and dialogue, Cambridge, USA

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423, 623–656

Shigeno S (1998) Cultural similarities and differences in the recognition of audio-visual speech stimuli. In: International conference on speech and language processing (ICSLP), vol 2. Sydney, Australia, pp 281–284

Silverman K, Beckman M, Pitrelli J, Ostendorf M, Wightman C, Price P, Pierrehumbert J, Hirschberg J (1992) ToBI: A standard for labeling English prosody. In: International conference on speech and language processing (ICSLP), Banff, Canada, pp 867–870

Sproat R, Hirschberg J, Yarowsky D (1992) A corpus-based synthesizer. In: International conference on speech and language processing (ICSLP), Banff, Canada, pp 563–566

Stallard D (2002) Flexible dialogue management in the Talk'n'Travel system. In: International conference on speech and language processing (ICSLP), Denver, USA, pp 2693–2696

Stevens SS, Volkmann J, Newman EB (1937) A scale for the measurement of the psychological magnitude pitch. J Acoust Soc Am 8(3):185–190

Stibbard R (2000) Automated extraction of ToBI annotation data from the reading/leeds emotional speech corpus. In: Proceedings of ISCA workshop on speech and emotion, Belfast, United Kingdom

Stibbard RM (2001) Vocal expression of emotions in non-laboratory speech: An investigation of the reading/leeds emotion in speech project annotation data. PhD thesis, University of Reading, United Kingdom

Sugimoto T, Ito N, Fujishiro H, Sugeno M (2002) Dialogue management with the semiotic base: A systemic functional linguistic approach. In: Proceedings of the 1st international conference on soft computing and intelligent systems, Tsukuba, Japan

Sun Y, Willett D, Brueckner R, Gruhn R, Bühler D (2006) Experiments on Chinese speech recognition with tonal models and pitch estimation using the Mandarin Speecon data. In: International conference on speech and language processing (ICSLP), Pittsburgh, USA

Sutton S, Cole R, de Villiers J, Schalkwyk J, Vermeulen P, Macon M, Yan Y, Kaiser E, Rundle B, Shobaki K, Hosom P, Kain A, Wouters J, Massaro D, Cohen M (1998) Universal speech tools: The CSLU toolkit. In: International conference on speech and language processing (ICSLP), Sydney, Australia, pp 3221–3224

Swerts M, Krahmer E (2000) On the use of prosody for on-line evaluation of spoken dialogue systems. In: International conference on language resources and evaluation (LREC), Athens, Greece

't Hart J, Cohen A, Collier R (1990) A perceptual study of intonation: An experimental-phonetic approach to speech melody. Cambridge University Press, Cambridge, United Kingdom

Takahash T, Fujii T, Nishi M, Banno H, Irino T, Kawahara H (2005) Voice and emotional expression transformation based on statistics of vowel parameters in an emotional speech database. In: International conference on speech and language processing (ICSLP), Lisbon, Portugal, pp 1853–1856

Tao J (2004) Context based emotion detection from text input. In: International conference on speech and language processing (ICSLP), Jeju, Korea

Tao J, Kang Y, Li A (2006) Prosody conversion from neutral speech to emotional speech. IEEE Trans Audio Speech Lang Proces 14(4):1145–1154

Tato R, Santos R, Kompe R, Pardo JM (2002) Emotional space improves emotion recognition. In: International conference on speech and language processing (ICSLP), Denver, USA, pp 2029–2032

Teixeira A, Silva L, Martinez R, Vaz F (2002) SAPWindows – Towards a versatile modular articulatory synthesizer. In: Proceedings of the 2002 IEEE workshop on speech synthesis, Santa Monica, pp 31–34

Telatar IE (1999) Capacity of multi-antenna gaussian channels. Eur Trans Telecommun 10(6): 585–595

ten Bosch L (2000) Emotions: What is possible in the ASR Framework. In: Proceedings of ISCA workshop on speech and emotion, Belfast, United Kingdom

Tepperman J, Traum D, Narayanan S (2006) "Yeah Right": Sarcasm recognition for spoken dialogue systems. In: International conference on speech and language processing (ICSLP), Pittsburgh, USA, pp 1838–1841

Tesser F, Cosi P, Drioli C, Tisato G (2005) Emotional festival-MBROLA TTS synthesis. In: International conference on speech and language processing (ICSLP), Lisbon, Portugal, pp 505–508

Tolkmitt FJ, Scherer KR (1986) Effect of experimentally induced stress on vocal parameters. J Exper Psychol Hum Percept Perform 12(3):302–313

Tomkins SS (1984) Affect theory. In: Scherer KR, Ekman P (eds) Approaches to emotion. Erlbaum, Hillsdale, USA, pp 163–195

Torres F, Sanchis E, Segarra E (2003) Development of a stochastic dialog manager driven by semantics. In: European conference on speech and language processing (EUROSPEECH), Geneva, Switzerland, pp 605–608

Tsuzuki R, Zen H, Tokuda K, Kitamura T, Bulut M, Narayanan SS (2004) Constructing emotional speech synthesizers with limited speech database. In: International conference on speech and language processing (ICSLP), Jeju, Korea

Turunen M, Hakulinen J (2001) Agent-based adaptive interaction and dialogue management architecture for speech applications. Lecture Note Comput Sci 2166:357–364

Ukelson J, Rodeh M (1989) A dialogue manager for efficient adaptive man-machine dialogues. In: Proceedings of the 13th annual international computer software and applications conference (COMPSAC), Orlando, USA, pp 588–595

Van Bezooijen R. A. MG (1984) Characteristics and recognizability of vocal expressions of emotion. Foris Publictions, Dordrecht, the Netherlands

Väyrynen E, Seppänen T, Toivanen J (2003) An experiment in emotional content classification of spoken Finnish using prosodic features. In: Finnish signal processing symposium, Tampere, Finland, pp 264–267

Veeravalli AG, Pan WD, Adhami R, Cox PG (2005) A tutorial on using hidden Markov models for phoneme recognition. In: Proceedings of the thirty-seventh southeastern symposium on system theory, Tuskegee, USA, pp 154–157

Veldhuijzen van Zanten G (1998) Adaptive mixed-initiative dialogue management. In: Proceedings of the IEEE 4th workshop on interactive voice technology for telecommunications applications (IVTTA), Torino, Italy, pp 65–70

Veldhuijzen van Zanten G (1999) User modelling in adaptive dialogue management. In: European conference on speech and language processing (EUROSPEECH), Budapest, Hungary, pp 1183–1186

Velten E (1968) A laboratory task for induction of mood states. Behav Res Ther 6:473–482

Viterbi AJ (1967) Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Trans Inform Theory IT-13:260–269

Vogt T, André E (2006) Improving automatic emotion recognition from speech via gender differentiation. In: International conference on language resources and evaluation (LREC), Genova, Italy, pp 1123–1126

von Neumann J, Morgenstern O (1944) Theory of games and economic behavior. Princeton University Press, Princeton, USA

Wagner T, Dieckmann U (1995) Sensor-fusion for robust identification of persons: A field test. In: Proceedings of the international conference on image processing (ICIP), vol 3. Washington, DC, USA, pp 516–519

Wahlster W (ed) (2006) SmartKom: Foundations of multimodal dialogue systems. Springer, Berlin, Germany

Walker MA, Aberdeen J, Boland J, Bratt E, Garafolo J, Hirschman L, Le A, Lee S, Narayanan S, Papineni K, Pellom B, Polifroni J, Potamianos A, Prabhu P, Rudnicky A, Sanders G, Seneff S, Stallard D, Whittaker S (2001) DARPA Communicator dialog travel planning systems: The June 2000 data collection. In: Dalsgaard P, Lindberg B, Benner H, Tan Z (eds) European conference on speech and language processing (EUROSPEECH), Aalborg, Denmark, pp 1371–1374

Walker MA, Cahn JE, Whittaker SJ (1997a) Improvising linguistic style: social and affective bases of agent personality. In: Johnson WL, Hayes-Roth B (eds) Proceedings of the first international conference on autonomous agents (Agents'97), Marina del Rey, USA, ACM Press, pp 96–105

Walker MA, Litman DJ, abd Alicia Abella C. AK (1998) Evaluating spoken dialogue agents with PARADISE: Two case studies. Comput Speech Lang 12(4):317–348

Walker MA, Litman DJ, Kamm CA, Abella A (1997b) PARADISE: A framework for evaluating spoken dialogue agents. In: Proceedings of the 35th annual meeting of the association of computational linguistics ACL/EACL, Madrid, Spain, pp 271–280

Walker W, Lamere P, Kwok P (2002) FreeTTS – A performance case study. Technical report TR-2002-114, Sun Microsystems, Inc., Burlington, USA

Wallace GK (1991) The JPEG still picture compression standard. Commun ACM 34(4):30–44

Wardhaugh R (1992) An introduction to sociolinguistics. Blackwell textbooks in linguistics. Blackwell, Chichester, United Kingdom

Watson JB (1930) Behaviorism. University of Chicago Press, Chicago, USA

Weiner B, Graham S (1984) An attributional approach to emotional development. In: Izard CE, Kagan J, Zajonc RB (eds) Emotions, cognition and behavior. Cambridge University Press, New York, USA, pp 167–191

Weizenbaum J (1966) ELIZA – A computer program for the study of natural language communication between man and machine. Commun ACM 9(1):36–45

Welch LR (2003) Hidden Markov models and the Baum-Welch algorithm. IEEE Inform Theory Soc Newslett 53(4):1, 10–13

Werner S, Hoffmann R (2007) Spontaneous speech synthesis by pronunciation variant selection – A comparison to natural speech. In: European conference on speech and language processing (EUROSPEECH), Antwerp, Belgium, pp 1781–1784

Whissel CM (1989) The dictionary of affect in language. In: Plutchik R, Kellerman H (eds) Emotion: Theory, research and experience, vol 4. Academic, New York, USA, pp 113–131

Whiteside SP (1998) Simulated emotions: An acoustic study of voice and perturbation measures. In: International conference on speech and language processing (ICSLP), Sydney, Australia, pp 699–703

Wichmann A (2000) The attitudinal effects of prosody, how they relate to emotion. In: Proceedings of ISCA workshop on speech and emotion, Belfast, United Kingdom

Williams JD (2006) Partially observable Markov decision processes for spoken dialogue management. PhD thesis, Cambridge University

Williams JD, Poupart P, Young S (2005) Partially observable Markov decision processes with continuous observations for dialogue management. In: Proceedings of the 6th SIGdial workshop on discourse and dialogue, Lisbon, Portugal

Williams JD, Young S (2007) Partially observable markov decision processes for spoken dialog systems. Comput Speech Lang 21(2):393–422

Wilting J, Krahmer E, Swerts M (2006) Real vs. acted emotional speech. In: International conference on speech and language processing (ICSLP), Pittsburgh, USA, pp 805–808

Witten IH, Frank E (2000) Data mining: Practical machine learning tools and techniques with Java implementations. Morgan Kaufmann, San Francisco, USA

Wolz U (1990) An object oriented approach to content planning for text generation. In Proceedings of the fifth international workshop on natural language generation, Dawson, USA, pp 95–104

Woodland PC, Young SJ (1993) The HTK tied-state continuous speech recogniser. In: European conference on speech and language processing (EUROSPEECH), Berlin, Germany, pp 2207–2210

Wu W, Zeng TF, Xu M-X, Bao H-J (2006) Study on speaker verification on emotional speech. In: International conference on speech and language processing (ICSLP), Pittsburgh, USA, pp 2102–2105

Wu X-J, Zheng F, Wu W-H (2002) A hybrid dialogue management approach for a flight spoken dialogue system. In: Proceedings of the first international conference on machine learning and cybernetics, Beijing, China, pp 824–829

Wu X-J, Zheng F, Xu M (2001) Topic forest: A plan-based dialog management structure. In: International conference on acoustics, speech and signal processing (ICASSP), Salt Lake City, USA

Wundt W (1924) An introduction to psychology. Allen & Unwin, London. (Translated by R. Pintner, original work published in 1912)

Yacoub S, Simske S, Lin X, Burns J (2003) Recognition of emotions in interactive voice response systems. In: European conference on speech and language processing (EUROSPEECH), Geneva, Switzerland, pp 729–732

Yang L, Campbell N (2001) Linking form to meaning: The expression and recognition of emotions through prosody. In: Proceedings of the 4th ISCA workshop on speech synthesis, Perthshire, United Kingdom

Yankelovich N (1996) How do users know what to say? ACM Interact 3(6):32–43

Yildirim S, Bulut M, Lee CM, Kazemzadeh A, Busso C, Deng Z (2004) An acoustic study of emotions expressed in speech. In: International conference on speech and language processing (ICSLP), Jeju, Korea

Young S (1994) The HTK hidden Markov model toolkit: Design and philosophy. Cambridge University Engineering Department, UK, Tech. Rep. CUED/F-INFENG/TR152

Young S (2001) Statistical modelling in continuous speech recognition (CSR). In: Proceedings of the international conference on uncertainty in artificial intelligence, Seattle, USA

Young S, Evermann G, Gales M, Hain T, Kershaw D, Liu X, Moore G, Odell J, Ollason D, Povey D, Valtchev V, Woodland P (2006) The HTK book (for HTK Version 3.4). Cambridge University Engineering Department

Yu C, Aoki PM, Woodruff A (2004) Detecting user engagement in everyday conversations. In: International conference on speech and language processing (ICSLP), Jeju, Korea

Yuan J, Shen L, Chen F (2002) The acoustic realization of anger, fear, joy and sadness in Chinese. In: International conference on speech and language processing (ICSLP), Denver, USA, pp 2025–2028

Yule GU (1927) On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. Philos Trans Roy Soc Lond 226:267–298

Zajonc RB (1984) On the primacy of affect. Am Psychol 39:124–129

Zell A, Mache N, Sommer T, Korb T (1991) The SNNS neural network simulator. In: Proceedings of 15. Fachtagung für Künstliche Intelligenz, Bonn, Germany, pp 254–263

Zhang S, Ching PC, Kong F (2006) Automatic recognition of speech signal in Mandarin. In: International conference on speech and language processing (ICSLP), Pittsburgh, USA, pp 1810–1813

Zhang T, Hasegawa-Johnson M, Levinson SE (2004) Children's emotion recognition in an intelligent tutoring scenario. In: International conference on speech and language processing (ICSLP), Jeju, Korea

Zhou L, Shi Y, Feng J, Sears A (2005) Data mining for detecting errors in dication speech recognition. IEEE Trans Speech Audio Process 13(5):681–688

Zinn C (2004) Flexible dialogue management in natural-language enhanced tutoring. In: Konvens 2004 workshop on advanced topics in modeling natural language dialog, Vienna, Austria, pp 28–35

Index