Emotional Speech Analysis in Mediation and Court Environments

Ciro Gracia[*], Xavier Binefa[*], Marta Poblet[°]
[*]*Universitat Pompeu Fabra, Barcelona*
[°] *ICREA Researcher at the UAB Institute of Law and Technology, Universitat Autònoma de Barcelona*

Abstract. When people communicate, their states of mind are coupled with the explicit content of the messages being transmitted. The implicit information conveyed by mental states is essential to correctly understand and frame the communication messages. In mediation, professional mediators include empathy as a fundamental skill when dealing with the relational and emotional aspects of a case. In court environments, emotion analysis intends to point out stress or fear as indicators of the truthfulness of certain asserts. In commercial environments, such as call-centers, automatic emotional analysis through speech is focused to detect deception or frustration. Computational analysis of emotions focuses on gathering information from speech, facial expressions, body poses and movements to predict emotional states. Specifically, speech analysis has been reported as a valuable procedure for emotional state recognition. While some studies focus on the analysis of speech features to classify emotional states, others concentrate on determining the optimal classification performance. In this paper we analyze current approaches to computational analysis of emotions through speech and consider the replication of their techniques and findings in the domains of mediation and legal multimedia.

Keywords: Emotions, Speech, Legal Multimedia, Legal Discourse, Signal Processing, Pattern Recognition.

1. Introduction

The study of cognitive and social aspects of human emotions has proved most successful during the last decades. In fact, the ability to infer others' emotional states has been pointed out as a fundamental skill in many commercial and social scenarios, where human relationship aspects have a critical weight. In the dispute resolution domain and, more specifically, in the mediation field, professional mediators use their ability to recognize emotions to improve or promote dialog between parties, thus fostering potential agreements. A good mediator is someone expected to equally work with both the emotional and relational aspects of a case, beyond the commonly known issues of the matter.

In law enforcement scenarios, as in legal multimedia, there is interest for detecting evidences of stress in voice. It is known that persons under stress suffer physical symptoms, as muscular micro-temblors in the speech production apparatus, which are reflected in the speech signal. Measuring this stress evidences provides information about possible untruthfulness or degree of potential threat in a flexible and non intrusive way. These applications are especially interesting because some emotional states fall into the sympathetic nervous system, producing an almost universal mechanical and predictable response.

Emotional speech recognition can be specifically related to the field of automatic emotion analysis, within the broad area of human-computer interaction (HCI). Automatic emotion analysis merges computer science techniques coming from the pattern recognition field, with other different areas of knowledge, such as psychology, neurology, or linguistic analysis, thus becoming a promising and interdisciplinary research field.

From a pattern recognition point of view, the analysis of human emotions is a relatively new field. Piccard published in 1997 an essay on affective computing which introduced the world of emotions and its roles inside human-computer communication (Picard, 1997). The idea of being capable of recognizing the emotional state of humans by analyzing its non verbal communication attracted the interest of pattern recognition researchers all over the world. Strategies using different sources of information, as speech analysis, facial expression, body movements and pose, were therefore introduced and experimentally studied.

Amongst those different approaches, we will focus our attention on the analysis of human speech. Human communication is heavily based on speech which, in many cases, also contains a fundamental part of human non-verbal communication, coded, for example, in elements such as intonation patterns, loudness, and a wide collection of other behavioral cues. Additionally, with the coming of digital multimedia, speech became one of the most cheap and flexible data sources for analysis. It also provided an easy, non-intrusive way to obtain emotional samples and data to be analyzed, even beating in this sense video recordings, where the presence of a camera pointing at a subject affects things as its reactions or its predisposition for showing its emotional states.

Emotion analysis from speech uses different methods for analyzing the speech signal, generally focusing on the non-verbal aspects of the speech. The analysis is complex insofar speech mixes verbal and non-verbal communication and, although both are—to some extent—independent from each other, there is still discussion about the lines separating them.

Advances in emotion recognition require databases for training, testing and validation. Several databases have been constructed, within different scenarios and with a range of different languages, most of them released from year 2000 onwards (Ververidis and Kotropulos, 2006). In section 2 we include a brief description of the factors describing emotional databases and the complexity of emotional database construction. In section 3 we describe some major approaches to signal processing techniques in order to obtain useful features, which are a discrete set of descriptors of the audio stream, for the task of emotion recognition. Section 4 contains a description of the recent approaches to feature selection and a classification used for speech emotion recognition. Finally, Section 5 shows experimental results of automatic techniques for emotion recognition. Based on these results, we discuss the replication of this techniques and findings in the domains of mediation and legal multimedia.

2. Emotional Databases

Emotional speech databases must take into account a set of factors in order not to bias or introduce distortion in data. Elements defining the scope of the database, the variety of contexts of the data, requirements of applicability, naturalness of emotions, copyrights and manual data labelling are not easy to conjugate. In fact, the existing surveys can be divided in two classes: those summarising and comparing the newly appearing databases, and those who shade light into the problems and decisions for constructing adequate emotion speech databases (Ververidis and Kotropulos, 2003, 2006; Douglas-Cowie *et al.*, 2003).

Among the factors affecting the database construction decisions are the number of speakers, the language and gender of the speakers, their dialects, and most important, the emotions to be included and how to obtain them truthfully. The emotional categories present in real life constitute a very complex taxonomy that can grow up to 64 classes. This variety notwithstanding, the research on emotional speech recognition has to be limited to certain emotions to be operational. Thus, most of the emotional speech data collections usually encompass no more than five or six emotions. Furthermore, as it is commonly agreed on, there are some primitive emotions that are more universal than others.

In practice, most of the available databases rely on models that describe the basic axis of an emotion. Most of them include emotions such as fear, sadness, joy, anger or neutrality. Figure 1 below shows a model for emotion categorisation frequently used in the computer science world (Osgood *et al*, 1957). According to this model, the computation of emotions is conceptualized as involving three major dimensions of connotative meaning: arousal, valence, and power.

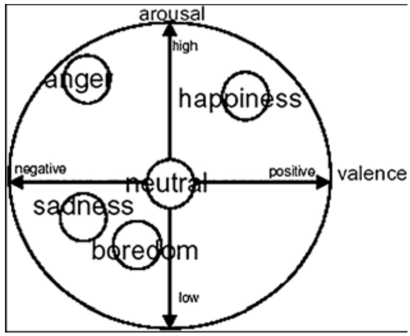

Figure 1: Basic emotional states projected on the arousal and valence axis

The easiest way to collect emotional speech is by having actors simulating such desired emotions. The difficulty regarding this approach is that, strikingly enough, so little is known about the relationship between faked and spontaneous emotional speech. There are many reasons to suspect that there are systematic differences between acted and natural emotional speech, such as the differences already reported in the natural and posed facial expressions from visual cues (Valstar *et al*., 2007). Acted speech is often "previously thought", not improvised—as most of the human speech is—leading to unnatural smoothness in the word flow: a "read" speech effect, which is known to have distinctive characteristics respect to spontaneous speech. It also typically takes place in a non-interactive monologue, so that interpersonal effects are not present. Some studies show that well acted emotional speech can be classified by humans, showing high accuracy for some emotions like anger, boredom or interest. Nevertheless, the differences from posed and non-posed speech and the ability of humans to distinguish them are not yet referred. In any case, for such a study, a database containing spontaneous speech examples would be required for an appropriate comparison.

Collecting spontaneous emotional speech is therefore not an easy task, since makes it difficult to fulfil some further requirements: namely, phonetic balance, required for some speech synthesis techniques, the control of the words and expressions used, or having similar recording sets and acquisition environments. Lost of spontaneity is the price of consistent analysis. Other problems arise from ethics and copyright issues, particularly with natural data. Natural emotional data are often very personal, and subjects may object to allow its diffusion. Radio and television provide rich sources, as chat shows and documentaries do, but their use also raises serious copyright problems. A long-term solution to these problems may be 'bootstrapping', or using truly natural material to guide the production of material that is acted, but genuinely close to nature.

Labelling and describing emotional databases also remain difficult tasks. Sometimes there is a varying emotional intensity, a coincidence of different emotions, different interpretations of the expression shown (depending on the person listening to the audio), etc. Unfaithfulness of the labelling puts statistic tractability of data within a compromise. A common reference method for labelling is to let different people classify emotional speech and only keep samples showing at least a certain degree of consensus.

Other contextual variables referring to the emotional database construction problem are expressed next. First, the modality of data: facial expressions shown in visual information carry fundamental information for emotion recognition. Then, including or not visual information will be an important variable. Secondly, there is the language used in the database: since different languages produce different phonetic and acoustic features, the studies based on the acoustic features will produce different models and different results. Furthermore, many signs of emotion are often defined by syntactic patterns and emotionally marked words, which depend on the language. For example, a study of English words with 'emotional connotation' can be found in the Semantic Atlas of Emotional Concepts (Averill, 1975). Thirdly, the recording setup: different microphones or the level of noise present during the recording affects the performance and generality of the different methods.

Finally, the accessibility of the constructed database is a key factor in order to promote research, algorithm comparison over same data, and to drive progress towards a better understanding and automatic handling of emotional speech.

3. Audio Signal Features Describing Emotional Speech

The first step towards automatic speech emotion classification is to extract relevant and measurable characteristics from the raw audio stream: the spectrogram. The selection of such features is crucial since their objective is to greatly reduce the amount of data to be analysed but still retain all the information relevant for emotion detection.

In fact, a wide variety of characteristics describing speech signal can be computed. Some are focused in short time statistics of spectrogram, while others reflect vocal tract description. Most of them are based on short term analysis of speech signal, which is very common in signal processing and automatic speech recognition. Also, most of the features are related to intonation characteristics, its temporal evolution or its short term basic statistics including mean, variance, maximum and minimum of several variables. Pitch (also known as F0) is a very important element, since both its tracking along the different time samples and its mean value strongly determine some of the stress emotions such as anger or fear. Other features related to speech recognition have been explored and try to estimate vocal tract characteristics that reflect phonetic content. Mel Filtered Cepstral Coefficients have been one of the most successful features in speech recognition due to its capability of representing human auditory response. Obtaining a clear vocal tract response improves the quality of the information about articulation effort, a characteristic of stressed voice. Different studies analyze and compare a wide range of features and its capability to classify basic emotions given a fixed classifier (Ververidis, Kotropoulos and Pitas, 2004). Yet, there is still much discussion on which features can be valuable to efficiently recognise basic emotional states. This motivates an exhaustive analysis, using feature selection algorithms to determine optimal feature sets for the multiclass emotion recognition task.

3.1 VOICE STRESS CHARACTERISTICS

Voice Stress Analysis (VSA) technology has already been introduced to the deceit detection field. It has been originated from the concept of micro muscle tremors (MMT) which was considered to be a source of stress detection. In moments of stress, especially if a person is exposed to jeopardy, the body prepares itself for fight or flight by increasing the readiness of its muscles to spring into action. This in turns causes the muscle vibrations to increase.

The micro tremors occur in the muscles that construct the vocal tract, being transmitted also through the speech. Voice stress studies cover a wide range of natural stressed speech origins, from dangerous situations, as aircraft/helicopter cockpit recordings, astronauts communications, law enforcement, emergencies call centres, etc. In general, stressed voice reflects some of the Lombard effect characteristics which include the involuntary tendency of speakers to increase the intensity of their voice to enhance its audibility, increasing phonetic fundamental frequencies and shifting formant centre frequencies. Pitch (F0) has been one of the most successfully studied acoustic descriptors of stressed speech. Short term perturbations, long term variability and higher mean value of F0 value reflects a significant increase with the stress level (F0 range can increase by even 3 octaves), but include large individual differences.

3.2 OTHER EMOTION CUES

Contour trends (Ververidis and Kotropoulos, 2006) are one of the fundamental tools to study the temporal evolution of the signal characteristics. Contour trends reflect some physical semantics on certain emotions. For example, pitch and intensity trends can be extracted in several ways by segmenting the signal in turns levels or syllabic units.

Inside each segment, statistics about its mean value, its rising or falling slope and plateaux are extracted and analysed. The speech rate is defined as the inverse duration of the voiced part of speech. The number of syllabic units in time is also used for the detection of certain emotion cues.

For example, anger usually produces higher energy and pitch levels. Differences between genders are remarkable in terms of the intensity of the energy and speech levels and in the speech rate: males express anger with slow speech rate while females employ fast speech rate under similar circumstances. In contrast, disgust and sadness usually show a low pitch level and a flat-like pitch contour.

The emotional state of fear is correlated with a high pitch level and a raised intensity level. Figure 2 summarize some basic cues for basic emotions covered in the emotional database of the University of Maribor

Table I. Cues analyzed by using 26 different features (DSPLAB of the University of Maribor)

	Values over average	Value near average	Value near average
Anger	Mean F0 Maximal F0 Standard deviation of DF0 Duration of syllables	Duration of fricatives	Duration of vocals Duration of plosives Duration of consonants Minimal DF0
Disgust	Duration of fricatives	Duration of vocals Duration of consonants Duration of plosives	Mean F0 Maximal F0 Duration of syllables Maximal DF0 DF0 range
Fear	Maximal DF0 Duration of plosives	DF0 range Duration of fricatives	Mean F0 Maximal F0 Minimal DF0 Standard deviation of DF0 Duration of syllables Duration of vocals Duration of consonants
Neutral (slow/soft)	Maximal F0 Standard deviation of DF0 Maximal DF0 DF0 range	Mean F0 Duration of vocals	Duration of syllables Duration of plosives Duration of fricatives Duration of consonants
Joy	Mean F0 Maximal F0 Standard deviation of DF0 Duration of syllables	Minimal DF0	Duration of vocals Duration of fricatives Duration of consonants Duration of plosives
Neutral (fast/loud)	Maximal DF0 DF0 range Duration of plosives	Duration of syllables Standard deviation of DF0 Duration of vocals	Mean F0 Duration of syllables
Surprise	Mean F0 Maximal F0 Minimal DF0 Standard deviation of DF0 Mean F03 Duration of syllables		Duration of vocals Duration of fricatives Duration of consonants Duration of plosives
Sadness	Duration of fricatives Duration of consonants Duration of plosives		Mean F0 Maximal F0 Minimal DF0 Standard deviation of DF0 Maximal DF0 DF0 range Duration of syllables Duration of vocals

4. Approaches to Emotional Speech Classification

Different approaches have been taken in order to classify emotional speech. Depending on the nature of the problem, it is possible to distinguish between the following situations: (i) speaker dependent recognition, where data from different emotional states on different phonetic contexts are uttered by the same speaker during the same or different sessions, and (ii) speaker independent recognition, where different speakers, which may have different contexts like gender or age, produce the data.

Speaker independent emotion recognition has been reported to be a difficult task, while speaker dependent recognition gets better results, which can have an 80-90% of success. In turns, speaker independent systems show a recognition rate between 50-55%. This is a naturally complex facet of speech emotion recognition since even humans are not able to obtain a higher recognition performance in a speaker-independent modality. Schuller *et al.* showed that the recognition performance of 12 human subjects dropped from 87.3% for the task of determining the expressed emotions of known persons to 64.7% for those expressed by an unknown subject (Schuller *et al.*, 2005)

Several approaches to emotion recognition on speech have been directed to explore and select the most adequate pattern recognition techniques to address the problems of feature selection and classification. Some studies have been centred in exploring massive sets of features in order to obtain a subset that optimally classifies different emotions. It is a non-trivial task, since using the whole set of features is not a good option. A smaller feature vector provides a better generalization performance [14] and reduces the well-known problem of the curse of dimensionality. Several different methods has been proposed for tackling the problem of feature selection, as: In [15] Schuller designed a genetic algorithm that generated features by combining low level descriptors with systematic derivation. Descriptive statistical analysis is used to find an optimal representation within the feature space. The feature set depends on the used classifier. In [5] correlation-based exclusion was used in order to exclude highly correlated features. Discriminant analysis has also been applied in its heteroscedastic form in order to maximize the separation between emotion classes while reducing the dimensionality of the feature space in [4].

Different tactics has been followed in order to choose the temporal scope of the features. Vlasenko *et al.* have combined short-time frame-level analysis with turns or chunk level analysis (Vlasenko *et al.*, 2007). Features modelling phoneme, word or sentence scopes are applied to speech emotion classification by Schuller *et al.* (Schuller *et al*, 2008). They use a very large feature space combined with the use of feature selection algorithms. Results show that features modelling larger time units are beneficial for emotion recognition, especially after assigning the optimal scope level to each specific emotion. Voiced versus non voiced sounds for emotional speech classification have also recently being explored showing that speaker independent features can be obtained by performing a hierarchical classification of emotions (Kim *et al*., 2008, 2009). This procedure reportedly produces lower confusions between anger/joy and neutral/sadness.

Classification schemes for emotion recognition have acquired wide variability. Several classification schemes have been proposed in to model emotional speech, from both the generative and the discriminative point of view. Furthermore, some approaches combine generative and discriminative techniques in order to model temporal sequences of features or to obtain combined properties from both methods.

Gaussian Mixture Models (GMM) have been one of the most used techniques for classifying acoustic vectors in speaker/gender recognition, but its inability to model temporal sequences produced their combination with other models. For example, GMM and Support Vector Machines (SVM) have been applied jointly in several forms. Thus, Hu *et. al* have built a universal background model of speech using a GMM (Hu *et. al,* 2007). To classify emotional speech, the universal model is adapted to the new coming utterance, and the resulting parameters are a high dimensional vector, further classified by a SVM adapted to multiclass classification. Chandrakala and Sekhar present another example of combination of generative models and SVM for speech emotion recognition where different generative models as GMM and Hidden Markov Models (HMM) are modelled for each of M training emotional utterances. Then each training utterance is scored by the M models, building a vector of a fixed length. Finally, a SVM classifier uses these vectors to model boundaries between classes (Chandrakala and Sekhar, 2009).

Other recent techniques include multisurface proximal SVM (Yang and Pu, 2008) Boosted GMM (Tang *et al*., 2009) or hierarchical classification using GMM (Kim *et al*., 2008) and they all reported good recognition results for speaker independent task.

5. Discussion

We have presented the current problems and some of the facing approaches for the task of emotion recognition through speech. Although emotions are not still well understood, the recent advances in emotional speech databases and machine learning approaches have provided a feasible way to perform initial applications.

Emotions also play a role in Internet and digital multimedia. Internet email has fostered a revolution in textual communication by providing a costless and fast way for international messaging. Similarly, videoconferencing and voice IP have a promising future empowering interpersonal information exchange.

Negotiation, mediation and emotions are not ignorant to each other either. For some years now, the analysis of the role and impact of emotions on negotiation and mediation processes has provided a fertile terrain for research (see, for instance, Jones and Bodtker 2001; Fischer and Shapiro, 2005; Jones, 2005; Kopelmana *et al.* 2006; Li and Roloff, 2006; Sinaceur and Tiedens, 2006; Van Kleef *et al*, 2006; Jameson, 2007; Steinel *et al.* 2007; Van Kleef, 2008; Overbeck et al., 2010). Similarly, the judicial and legal domains at large have benefitted from the study of emotions in different legal settings (i.e. Blumenthal 2005; Daicoff, 2005, Maroney, 2006; Wessell *et al.*, 2006). From this point of view future mediators may find it as a fundamental tool to extend its relationship capacities. In this scenario, tools that analyze voice or video data from the emotional point of view could be a valuable allied for the management of mediators relational aspects. We summarized tools that will track videoconferencing speech, trying to represent it into an emotional space, helping the mediator to judge the emotional flow of the communication. To provide emotional saliency can be a valuable tool for review and navigate mediation multimedia in order to find out especially remarkable parts. In our previous works with legal multimedia (Binefa *et al.*, 2007), we explored the characteristics of legal multimedia and how to extract patterns from it. Emotional speech analysis on legal multimedia can also become a fruitful way for semantic annotation, as emotions carry semantics that are very interesting in order to further analyze behaviours or remarkable events.

Acknowledgements

This work has been funded by the Spanish projects: (i) ONTOMEDIA: Platform of Web Services for Online Mediation, Spanish Ministry of Industry, Tourism and Commerce (Plan AVANZA I+D, TSI-020501-2008, 2008-2010), (ii) ONTOMEDIA. Web Semántica, ontologías y ODR: plataforma de servicios web para la mediación on line. CSO-2008-05536-SOCI, Spanish MInisytry of Science and Innovation.

References

Averill, J. R. (1975) *A Semantic Atlas of Emotional Concepts*, American Psychological Association, *JSAS Catalog of Selected Documents in Psychology: 5*, 330.

Binefa, X., Gracia, C., Monton, M., Carrabina, J., Montero, C., Serrano, J., Blázquez, M., Benjamins, V.R., Teodoro, E., Poblet, M., Casanovas, P. (2007), *Developing Ontologies for Legal Multimedia Applications*, LOAIT 2007, pp. 87-101.

Blumenthal, J.A. (2005), *Law and the emotions: the problems of affective forecasting*, Ind. L.J., Vol. 80, p. 155.

Chandrakala, C. and Chandra Sekhar, C. (2009), *Combination of generative models and SVM based classifier for speech emotion recognition*, Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, June 14-19, 2009, pp. 1374-1379.

Daicoff, S. (2005) *Law as a Healing Profession: The "Comprehensive Law Movement*", Bepress Legal Series, Paper 1331, available at http://law.cwru.edu/lectures/files/2008-2009/20090410_Daicoff_excerpt.pdf (accessed May 17).

Douglas-Cowie, E. Campbell, N. Cowie, R., Roach, P. (2003), *Emotional speech: Towards a new generation of databases*. Speech Communication Journal, Vol. 40, pp. 33-603.

Fisher, R. and Shapiro, D. (2005), *Beyond Reason: Using Emotions As You Negotiate*. New York: Viking Press.

Hu, H. Ming-Xing, X., Wu, W. (2007) *GMM supervector based SVM with spectral features for speech emotion recognition*, *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Vol. 4, pp. 413–416.

Jameson, J.K., Bodtker, A.M., Porch, D.M., Jordan, W.J. (2007), *Exploring the role of emotion in conflict transformation*, Conflict Resolution Quarterly, Vol. 27, No. 2: pp. 167-192, DOI: 10.1002/crq.254.

Jones, T. S. (2005), Emotion in mediation: implications, applications, opportunities, and challenges, in M. Herrman (ed.), *Blackwell Handbook of Mediation:Theory and Practice*. New York: Blackwell.

Jones, T. S. and Bodtker, A. M. (2001). *Mediating with Heart in Mind: Addressing Emotion in Mediation Practice*, Negotiation Journal, Vol. 17, No. 3, pp. 207–244.

Kim, E. H., Hyun, K.H., Kim, H.H., Kwak, Y.K (2008), *Speech Emotion Recognition Separately from Voiced and Unvoiced Sound for Emotional Interaction Robot*, International Conference on Control, Automation and Systems 2008, pp. 2014-2019.

Kim, E. H., Hyun, K.H., Kim, H.H., Kwak, Y.K (2009), *Improved emotion recognition with a novel speaker-independent feature*, IEEE/ASME Transactions on Mechatronics, Vol. 14, No. 3, June 2009, pp. 317-325.

Kopelmana, S, Rosette, A.S., Thompsonc, L. (2006), *The three faces of Eve: Strategic displays of positive, negative, and neutral emotions in negotiations*, Organizational Behavior and Human Decision Processes, Vol. 99: pp. 81–101.

Li, S. and Roloff, M. E. (2006), *Strategic emotion in negotiation: cognition, emotion, and culture*, in G. Riva, M. T. Anguera, B. K. Wiederhold, and F. Mantovani (eds.), *From Communication to Presence: Cognition, Emotions, and Culture Towards the Ultimate Communicative Experience*, Amsterdam: IOS Press, 2006.

Maroney, T.A. (2006), *Law and emotion: a proposed taxonomy of an emerging field*, Law and Human Behavior, Vol. 30, pp.119–142, DOI 10.1007/s10979-006-9029-9.

Moataz, M. H., El Ayadi, M., Kamel, S., Karray, F. (2007), *Speech emotion recognition using Gaussian mixture vector autoregressive models*. IEEE International Conference on Acoustics, Speech and Signal Processing 2007. (ICASSP 2007), pp. 957-960.

Osgood, C. E., Suci, J. G., Tannenbaum, P. H. (1957), *The Measurement of Meaning,* Urbana, IL: Univ. Illinois, 1957, pp. 31–75.

Overbeck, J.R., Neale, M.A., Govan, C. L. (2010), *I feel, therefore you act: Intrapersonal and interpersonal effects of emotion on negotiation as a function of social power*, Organizational Behavior and Human Decision Processes (in press).

Picard, R.W. (1997), *Affective computing*, MIT Press, Cambridge, Mass.

Schuller, B, Reiter, S., Muller, R. Al-Hames, M., Lang, M., Rigoll, G. (2005) *Speaker Independent Speech Emotion Recognition by Ensemble Classification*, ICME, 2005 IEEE International Conference on Multimedia and Expo, pp.864-867.

Schuller, B., Reiter, S., Rigoll, G. (2006) *Evolutionary feature generation in speech emotion recognition*, ICME 2006: pp. 5-8.

Schuller, B., Vlasenko, B., Arsic, D., Rigoll, G., Wendemuth, A. (2008) *Combining speech recognition and acoustic word emotion models for robust text-independent emotion recognition*. ICME 2008: IEEE International Conference on Multimedia & Expo, pp. 1333-1336.

Sinaceur, M. and Tiedens, L. Z. (2006). *Get Mad and Get More Than Even: When and Why Anger Expression Is Effective in Negotiations*, Journal of Experimental Social Psychology, Vol. 42, No. 3, pp. 314–322.

Steinel, W, Van Kleef, G.A., Harinck, F. (2008), *Are you talking to me?! Separating the people from the problem when expressing emotions in negotiation*, Journal of Experimental Social Psychology, Vol. 44: pp. 362–369.

Tang, H., Chu, S.M., Hasegawa-Johnson, M., Huang, T:S: (2009), *Emotion recognition from speech via boosted Gaussian mixture models*, ICME 2009.

Valstar, M.F., Gunes, H., Pantic, M. (2007), *How to distinguish posed from spontaneous smiles using geometric features*, ICMI 2007, pp. 38-45.

Van Kleef, G.A. (2008) *Emotion in Conflict and Negotiation: Introducing the Emotions as Social Information (EASI) Model*, in N. M. Ashkanasy and C. L. Cooper (eds.) *Research Companion to Emotion in Organizations*. Massachusetts: Edward Elgar Publishing: pp. 392-404.

Van Kleef, G.A., Carsten K., De Dreu1, W., Pietroni, D., Manstead, A.S.D. (2006), *Power and emotion in negotiation: Power moderates the interpersonal effects of anger and happiness on concession making*, European Journal of Social Psychology, Vol. 36, pp. 557–581, DOI: 10.1002/ejsp.320.

Ververidis, D. and Kotropoulos, C. (2003) *A review of emotional speech databases,* Proceedings of the 9th Panhellenic Conference in Informatics 2003.

Ververidis, D., Kotropoulos, C., Pitas, I. (2004), *Automatic emotional speech classification*, in Proc. 2004 IEEE Int. Conf. Acoustics, Speech and Signal Processing, Vol. 1, pp. 593-596.

Ververidis, D. and Kotropoulos, C. (2006), *Emotional speech recognition: Resources, features, and methods*, Speech Communication Journal, Vol. 48, No. 9, 1162-1181.

Vlasenko, B., Schuller, B. Wendemuth, A., Rigoll, G. (2007), *Combining Frame and Turn-Level Information for Robust Recognition of Emotions within Speech*. InterSpeech conference 2007, pp. pp. 2225–2228.

Wessel1, E.; Guri, C.; Drevland1, C.B., Eilertsen1, D.E., Magnussen, S. (2006), *Credibility of the Emotional Witness: A Study of Ratings by Court Judges, Law and Human Behavior*, Vol. 30: pp. 221–230, DOI 10.1007/s10979-006-9024-1

Witten I. and Frank, E. (2000), *Data mining*, Los Altos, CA: Morgan Kaufflan.

Yang, Ch. and Pu, Y. (2008), *Efficient speech emotion recognition based on multisurface proximal support vector machine*, 2008 IEEE Conference on Robotics, Automation and Mechatronics, pp. 55-60.