

# The Investigation the Effects of the Performance of an Independent Emotion Recognition of Model Used in the Dimensioning of Emotions

Turgut Özseven

Turhal Vocational School, Gaziosmanpaşa University, Tokat, Turkey

Muharrem Düğenci

Faculty of Engineering, Department of Industrial Engineering, Karabük University, Karabük, Turkey

## Abstract

Emotion recognition aims at determining the state of emotion which is included in the speech or mimics of a person. Emotion recognition from speech is an area related to signal processing and psychology. Acoustic parameters obtained from speech signals through acoustic analysis, which is one of objective evaluation methods, is intensively used in emotion recognition studies. In this paper success in emotion recognition is examined in categorical aspects and the impact of dimensional model on independent emotion recognition success is investigated. Acoustic parameters were subjected to classification with Support Vector Machine in order to determine emotion recognition success. According to the obtained findings, emotion recognition success in categorical structure, dimensional structure and categorical-dimensional structure were 69.5%, 73.3% and 87.1%, respectively. Even if dimensional structure is higher in arousal than in valence, when emotion recognition success is examined on each emotion dimension, valence provided higher success.

**Keywords:** *Dimensional Emotion, Emotion Recognition, Acoustic Analysis, SVM*

## Introduction

The emotional state of persons causes changes in the physiological structure of persons, which naturally affects voice. Objective and subjective methods are used for emotion recognition from speech. Perceptual recognition, which is a subjective method, is performed by listening by an expert and can show difference depending on the experience and probation of the expert. In order to overcome this problem of subjective evaluation, objective evaluation methods are employed. Acoustic analysis is used with the purpose of objective evaluation of voice disorders by obtaining several acoustic parameters from voice.

Features of voice signal and voice path change depending on the person, age, sex, sound path length, height, weight and emotional state. Therefore the emotional state of a person can be identified depending on feature vectors that will be obtained with sound analysis methods. Speech Emotion Recognition is not a new area of study; it was first developed in mid 1980s using statistical features of some acoustic parameters. In following years, with the development of computer architecture, more complicated emotion recognition algorithms began to be used (Ververidis & Kotropoulos, 2006). With acoustic parameters, language and discourse information were also used for emotion recognition (Chul Min Lee & Narayanan, 2005).

The most important limitation of Speech Emotion Recognition is about obtaining data whose validity is accepted and which can reflect emotions fully. With this purpose, there are ready datasets in the literature (EMO-DB, EMOVA, SAVEE, eINTERFACE, SUSAS) used in several studies whose validity is recognized; in addition, some researchers aimed at gathering their own datasets.

The basic purpose of the process in most emotion recognition studies is classification, and in most of these studies classification of emotional states is performed. Traditional classification techniques have been used in almost all of the proposed emotion recognition systems. Recent studies are focusing on the impact of hybrid classifiers and classifiers on acoustic parameters. The most widely used classifiers in the literature are Support Vector Machine (SVM), Gaussian Mixture Model (GMM) and Hidden Markov Model (HMM).

There are three basic approaches in modelling of emotions in psychology studies (Sezgin, Gunsel, & Kurt, 2012): categorical, dimensional, and appraisal-based approach. Since the appraisal-based approach is not prevalently used because of its complex and sophisticated measurements of change (Gunes, Schuller, Pantic, & Cowie, 2011).

Categorical approach is independent expression of basic emotions. Ekman (Ekman, 1992) defined emotions in 6 basic categories: happiness, sadness, anger, fear, surprise, and disgust. However, categorical approach can fail to reflect the real situation when people are not in a single emotion but in complicated emotions (Gunes et al., 2011).

Due to the deficiencies of categorical approach in reflecting complicated emotions, dimensional approach is employed. Dimensional approach states that emotional state of a person is not depending on a single category but on multiple categories. There are several emotional dimensioning models presented for this purpose (Sezgin et al., 2012). A three-dimensional emotion space is proposed: arousal (activation), potency (power), and valence (pleasure) evaluation (Schlosberg, 1954). Another alternative is simpler two-dimensional emotion space: arousal and valence. Yet, the most widely used dimensional model is based on the assumption of Russell (Russel, 1980) that each basic emotion is represented by a bipolar entity being a part of the same emotional dimension in two-dimensional emotion space.

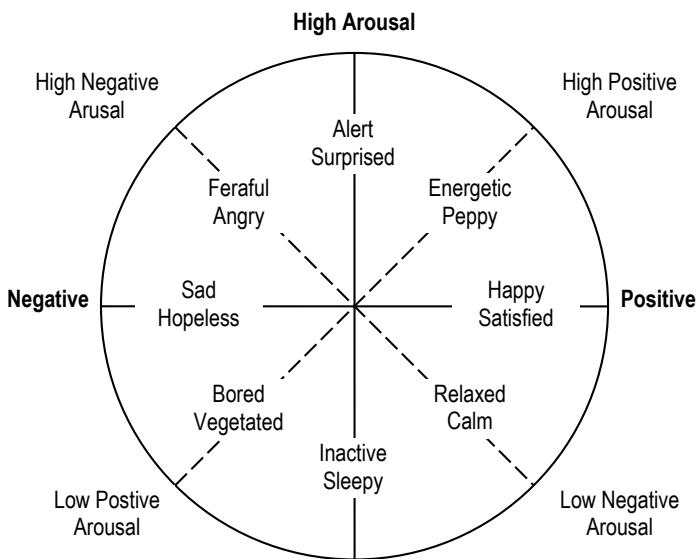


Figure 1. The two-dimensional structure of affect, showing the dimensions of Positive Activation and Negative Activation on the diagonals and their relationship to the alternative dimensions of Valence (horizontal) and Arousal (vertical), as well as the 16 affect terms used to measure current mood (Kallinen, 2006). Adapted from Larsen and Diener (Larsen & Diener, 1992) and Watson et al. (Watson, Wiese, Vaidya, & Tellegen, 1999).

In our study EMO-DB dataset is used; therefore, the studies which consist of categorical and dimensional emotion recognition using EMO-DB in the literature were examined. A summary of the literature search is given in Table 1.

Table 1. Related works

Emotions/Dimension	Class.	Par.Count	Acc.	Ref.
Arousal-Negative (boredom, disgust, neutral, sadness)			96.8%	
Arousal-Positive (anger, fear, happiness)				
Valence-Negative (anger, boredom, disgust, fear, sadness)	SVM	6552	87.0%	(Schuller, Vlasenko, Eyben, Rigoll, & Wendemuth, 2009)
Valence-Positive (happiness, neutral)				
Categorical			84.0%	
High arousal (anger, happiness, anxiety/fear), Low arousal (sadness, boredom, disgust, neutral)	ANN	133	83.7%	(Iliou & Anagnostopoulos, 2009)
Group 1...Group 8 (Watson and Tellegen's model)	HMM	39	81.2%	(Yun & Yoo, 2009)
High Arousal Low Arousal	ANN	24	47.0%	(Anagnostopoulos & Vovoli, 2009)
Anger, Disgust, Fear, Happy, Neutral, Sadness	SVM	121	78.3%	(Luengo, Navas, & Hernaez, 2010)
Anger, Fear, Happy, Neutral, Sadness, Boredom	SVM	~80	79.7%	(Giannoulis & Potamianos, 2012)
Arousal, Valence	SVM	-	86.3%	(Sezgin et al., 2012)
Anger, Disgust, Fear, Happy, Neutral, Sadness, Boredom	SVM	4368	93.8%	(Ivanov & Riccardi, 2012)
Anger, Disgust, Fear, Happy, Neutral, Sadness, Boredom	SVM	39	67.0%	(Vasuki & Aravindan, 2012)
Anger, Disgust, Fear, Happy, Neutral, Sadness, Boredom	SVM	95	79.5%	(Hassan & Damper, 2012)
Arousal, Valence	SVM	~70	90.5%	(Ramakrishnan & El Emary, 2013)
Anger, Disgust, Fear, Happy, Neutral, Sadness, Boredom	SVM	32	80.6%	(Garg, Kumar, & Sinha, 2013)
Anger, Disgust, Fear, Happy, Neutral, Sadness, Boredom	SVM	6552	85.2%	(Chiou & Chen, 2013)
Anger, Disgust, Fear, Happy, Neutral, Sadness, Boredom	SVM	90	80.4%	(Zhao, Zhang, & Lei, 2014)

Arousal, Valence	ANFC	41	91.9%	(Lika, Seldon, & Kiong, 2014)
Anger, Disgust, Fear, Happy, Neutral, Sadness, Boredom	SVM	204	78.8%	(Zhao & Zhang, 2015)
Anger, Disgust, Fear, Happy, Neutral, Sadness, Boredom	SVM	2641	85.9%	(Shahzadi, Ahmadyfard, Harimi, & Yaghmaie, 2015)
High Activation, Low Activation	SVM	-	90.3%	(Alonso, Cabrera, Medina, & Travieso, 2015)

When the literature search given in table 1 is examined, it can be seen that the highest success obtained with SVM classifier in categorical classification was 93.8 percent. In studies which use the same data, the reason of obtaining different classification successes is the difference in the number and types of acoustic parameters obtained from data. Studies which examine emotional dimensions looked in arousal-valence dimension and high success was obtained with 96.8 percent. In studies found in the literature, emotion recognition is examined in categorical and dimensional aspects and no studies are found which evaluate each dimension categorically within itself.

The purpose of this study is to examine with SVM classifier the emotion recognition success in categorical and dimensional aspect of the seven emotions (anger, boredom, disgust, anxiety/fear, happiness, sadness, neutral) found in the Berlin Database of Emotional Speech (Emo-DB) database. In addition, the emotions in every emotion dimension are evaluated categorically.

## Materials and Method

In our study EMO-DB is used. Emo-DB includes voice recordings consisting of 7 different emotions (anger, boredom, disgust, anxiety/fear, happiness, sadness, neutral) spoken by actors within the project developed by department of communication science at Institute of Speech and Communication of Technical University of Berlin. The database was created by emotional expressions of 10 different sentences by 10 actors in 20-35 age interval. Voice records have 16 kHz sampling frequency and 16 bit mono features (Burkhardt, Paeschke, Rolfes, Sendlmeier, & Weiss, 2005).

In the study 535 voice records are examined, the distribution of data-specific features of which are given in Table 2.

Table 2. The features belonging to the data used in the study

Emotion	Male	Female	Age1 (30-35)	Age2 (20-29)	Total
Anger	60	67	90	37	127
Fear	34	33	50	17	67
Boredom	35	46	60	21	81
Disgust	11	35	31	15	46
Happiness	27	44	53	18	71
Sadness	25	37	47	15	62
Neutral	39	40	50	29	79

<b>Total</b>	233	302	381	154	<b>535</b>
--------------	-----	-----	-----	-----	------------

Acoustic analysis of speech records in both databases was performed with Praat (Boersma & Weenink, 2002) software and 15 acoustic parameters were obtained from each speech record. In the pre-processing of speech records, frame size 25ms, 50% overlap and Hamming windowing were used. Acoustic parameters obtained with acoustic analysis are given in Table 3.

Table 3. The acoustic parameters used in this paper

<b>Acoustic Parameters</b>	<b>Descriptive Statistics</b>	
	<b>Mean</b>	<b>Std.Dev.</b>
Fundamental Frequency (F0)	✓	✓
Formant Frequency (F1, F2, F3)	✓	✓
Jitter (Local, Rap)/ Shimmer (Local, apq3)	✓	
Unvoiced Frame	✓	
Intensity	✓	✓

Fundamental frequency (F0), reflects vibration of speed of vocal fold and determines the individual's sound (Zupan, Neumann, Babbage, & Willer, 2009). Formant is resonant on the sound path. There are an infinite number of format theoretically. but in practice, only the first 3 or 4 contains important information. Formantlar are defined with formant numbers as F1, F2 ve F3 (Rezaei & Salehi, 2006). Jitter, indicates the change between periods and contains the resulting involuntary irregularities. The periodic variation between amplitude peak is called as shimmer. Unvoiced frame shows the ratio of pauses during speech. Intensity shows the energy created as a result of the amplitude change in sound signal (Zupan et al., 2009).

In the evaluation of emotion recognition success, Support Vector Machine (SVM) classifier and WEKA (Hall et al., 2009) package program were used. SVM is a method based on statistical learning theory. The basic purpose is based on the principle of defining the function which best distinguished classes, in other words, the hyper-plane. It is one of the most widely used classifiers in SER systems in the literature.

Two groupings were made in order to dimensionally evaluate emotions, which are;

- Arousal-Negative (boredom, disgust, sadness, neutral) ve Arousal-Positive (anger, fear, happiness)
- Valence-Negative (anger, boredom, disgust, fear, sadness) ve Valence-Positive (happiness, neutral)

## Results

Emotion recognition successes are realized in three different patterns, namely categorical, dimensional and categorical for each dimension. The categorical emotion recognition successes are given in Table 4.

Table 4. Categorical emotion recognition

	Happiness	Neutral	Anger	Sadness	Fear	Boredom	Disgust	AVG
Accuracy	56.3%	64.6%	84.3%	85.5%	60.9%	76.5%	37.0%	69.5%

The highest success in categorical emotion recognition is sadness and the lowest success is obtained in the emotion of disgust. The reason for such low success in disgust emotion is that its number in the entire data is few. Classification successes obtained for arousal and valence dimensions is given in Table 5.

Table 5. Dimensional emotion recognition

	Arousal Negative	Arousal Positive	Arousal AVG	Valence Positive	Valence Negative	Valence AVG
Accuracy	90.3%	92.9%	91.6%	12.7%	97.9%	74.0%

When the results given in table 5 are examined, it can be seen that the success of arousal dimensioning is higher compared to valence. The very low success in valence positive dimension is due to the fact that it consists of two emotions and the neutral emotion does not lead to essential change on the parameters. Table 6 gives the results obtained when emotions in each dimension of the dimensional structure are categorically classified.

Table 6. The categorical classification of emotions contained within the dimensional structure

	Arousal Positive			Arousal Negative			
	Happiness	Anger	Fear	Neutral	Sadness	Boredom	Disgust
Accuracy	56.3%	81.9%	69.6%	67.1%	87.1%	74.1%	71.7%

	Valence Positive		Valence Negative				
	Happiness	Neutral	Anger	Sadness	Fear	Boredom	Disgust
Accuracy	95.8%	93.7%	92.1%	90.3%	59.4%	86.4%	47.8%

When the results given in Table 6 are examined, even if the emotion recognition success is higher in arousal dimensional structure, valence is more successful in categorical classification of the dimensional structure. Table 7 gives the emotion recognition success rates according to the categorical and dimension-based categorical result.

Table 7. Emoton recognition accuracy according to the results categorical based on dimension and categorical

Emotions	Accuracy

	Categorical	Arousal-Categorical	Valence-Categorical
Anger	84.3%	81.9% ↓	92.1% ↑
Fear	60.9%	69.6% ↑	59.4% ↓
Boredom	76.5%	74.1% ↓	86.4% ↑
Disgust	37.0%	71.7% ↑	47.8% ↑
Happiness	56.3%	56.3% ↔	95.8% ↑
Sadness	85.5%	87.1% ↑	90.3% ↑
Neutral	64.6%	67.1% ↑	93.7% ↑
AVG	69.5%	73.3% ↑	87.1% ↑

According to Table 7, recognition success in some emotions is increasing in arousal-categorical structure whereas some others witnessed decline. In valence-categorical structure, only recognition success in the emotion of fear declines slightly whereas recognition success in other emotions increased considerably. When average success rates are examined, arousal dimensioning increased dimension-based categorical success slightly and valence dimensioning increased dimension-based categorical success considerably. These results show that usage of dimensional structure in emotion recognition provides considerable increase in the success of emotion recognition.

## Conclusion

The emotional state of persons can cause changes in their physiological structures, which, therefore, is reflected on their voice. Sometimes the feelings of a person do not consist of a single emotion but a variety of emotions. In such cases effort is paid to perform dimensional analysis of the emotion and to increase the success rate of emotion recognition. Emotion recognition aims at determining the emotional state of a person by examining their speeches through signal processing methods. In this paper, categorical, dimensional and dimensional-categorical emotion recognition success rates in emotion recognition are examined. According to the obtained results, dimensional structure provided higher success compared to categorical structure. Dimensional-categorical structure increased the success rate obtained through each emotion considerably. In the light of these findings, higher success rates will be achieved by using valence dimensional structure followed by categorical structure for each dimension in emotion recognition.

## References

- [1] Alonso, J. B., Cabrera, J., Medina, M., & Travieso, C. M. (2015). New approach in quantification of emotional intensity from the speech signal: emotional temperature. *Expert Systems with Applications*, 42(24), 9554–9564. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0957417415005229>
- [2] Anagnostopoulos, C. N., & Vovoli, E. (2009). Sound processing features for speaker-dependent and phrase-independent emotion recognition in Berlin Database. In *Information systems development* (pp. 413–421). Springer. Retrieved from [http://link.springer.com/10.1007/b137171\\_43](http://link.springer.com/10.1007/b137171_43)
- [3] Boersma, P., & Weenink, D. (2002). Praat, a system for doing phonetics by computer. *Glott International*, 5(9), 341–345.

- [4] Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., & Weiss, B. (2005). A database of German emotional speech. In *Interspeech* (Vol. 5, pp. 1517–1520). Retrieved from [https://www.doc.ic.ac.uk/project/2013/163/g1316307/web/pdfs/database\\_of\\_German\\_Emotional\\_Speech.pdf](https://www.doc.ic.ac.uk/project/2013/163/g1316307/web/pdfs/database_of_German_Emotional_Speech.pdf)
- [5] Chiou, B.-C., & Chen, C.-P. (2013). Feature space dimension reduction in speech emotion recognition using support vector machine. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific* (pp. 1–6). IEEE. Retrieved from [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6694251](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6694251)
- [6] Chul Min Lee, & Narayanan, S. S. (2005). Toward detecting emotions in spoken dialogs. *IEEE Transactions on Speech and Audio Processing*, 13(2), 293–303. <http://doi.org/10.1109/TSA.2004.838534>
- [7] Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3–4), 169–200. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/02699939208411068>
- [8] Garg, V., Kumar, H., & Sinha, R. (2013). Speech based Emotion Recognition based on hierarchical decision tree with SVM, BLG and SVR classifiers. In *Communications (NCC), 2013 National Conference on* (pp. 1–5). IEEE. Retrieved from [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6487987](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6487987)
- [9] Giannoulis, P., & Potamianos, G. (2012). A hierarchical approach with feature selection for emotion recognition from speech. In *LREC* (pp. 1203–1206). Retrieved from [http://www.lrec-conf.org/proceedings/lrec2012/pdf/917\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2012/pdf/917_Paper.pdf)
- [10] Gunes, H., Schuller, B., Pantic, M., & Cowie, R. (2011). Emotion representation, analysis and synthesis in continuous space: A survey. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on* (pp. 827–834). IEEE. Retrieved from [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5771357](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5771357)
- [11] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter*, 11(1), 10–18. Retrieved from <http://dl.acm.org/citation.cfm?id=1656278>
- [12] Hassan, A., & Damper, R. I. (2012). Classification of emotional speech using 3DEC hierarchical classifier. *Speech Communication*, 54(7), 903–916. <http://doi.org/10.1016/j.specom.2012.03.003>
- [13] Iliou, T., & Anagnostopoulos, C.-N. (2009). Comparison of different classifiers for emotion recognition. In *Informatics, 2009. PCI'09. 13th Panhellenic Conference on* (pp. 102–106). IEEE. Retrieved from [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5298878](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5298878)
- [14] Ivanov, A., & Riccardi, G. (2012). Kolmogorov-Smirnov test for feature selection in emotion recognition from speech. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on* (pp. 5125–5128). IEEE. Retrieved from [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6289074](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6289074)
- [15] Kallinen, K. (2006). *Towards a comprehensive theory of musical emotions: a multidimensional research approach and some empirical findings*. University of Jyväskylä. Retrieved from <https://jyx.jyu.fi/dspace/handle/123456789/13410>
- [16] Larsen, R. J., & Diener, E. (1992). Promises and problems with the circumplex model of emotion. Retrieved from <http://psycnet.apa.org/psycinfo/1992-97396-002>
- [17] Lika, R. A., Seldon, H. L., & Kiong, L. C. (2014). Feature analysis of speech emotion data on arousal-valence dimension using adaptive neuro-fuzzy classifier. In *Industrial Automation, Information and Communications*



- Technology (IAICT), 2014 International Conference on* (pp. 104–110). IEEE. Retrieved from [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6922106](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6922106)
- [18] Luengo, I., Navas, E., & Hernaez, I. (2010). Feature Analysis and Evaluation for Automatic Emotion Identification in Speech. *IEEE Transactions on Multimedia*, 12(6), 490–501. <http://doi.org/10.1109/TMM.2010.2051872>
- [19] Ramakrishnan, S., & El Emary, I. M. M. (2013). Speech emotion recognition approaches in human computer interaction. *Telecommunication Systems*, 52(3), 1467–1478. <http://doi.org/10.1007/s11235-011-9624-z>
- [20] Rezaei, N., & Salehi, A. (2006). An Introduction to Speech Sciences (Acoustic Analysis of Speech). *Iranian Rehabilitation Journal*, 4(4), 5–14. Retrieved from [http://irj.uswr.ac.ir/files/site1/user\\_files\\_055690/admin-A-10-1-3-3c7ebda.pdf](http://irj.uswr.ac.ir/files/site1/user_files_055690/admin-A-10-1-3-3c7ebda.pdf)
- [21] Russel, J. A. (1980). A circumplex model of affect. *J Personal Soc Psychol*, 39, 1161–1178. Retrieved from <https://www2.bc.edu/~russeljm/publications/Russell1980.pdf>
- [22] Schlosberg, H. (1954). Three dimensions of emotion. *Psychological Review*, 61(2), 81. Retrieved from <http://psycnet.apa.org/journals/rev/61/2/81/>
- [23] Schuller, B., Vlasenko, B., Eyben, F., Rigoll, G., & Wendemuth, A. (2009). Acoustic emotion recognition: A benchmark comparison of performances. In *Automatic Speech Recognition & Understanding, 2009. ASRU 2009. IEEE Workshop on* (pp. 552–557). IEEE. Retrieved from [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5372886](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5372886)
- [24] Sezgin, M. C., Gonsel, B., & Kurt, G. K. (2012). Perceptual audio features for emotion detection. *EURASIP Journal on Audio, Speech, and Music Processing*, 2012(1), 1–21. Retrieved from <http://link.springer.com/article/10.1186/1687-4722-2012-16>
- [25] Shahzadi, A., Ahmadyfard, A., Harimi, A., & Yaghmaie, K. (2015). Speech emotion recognition using nonlinear dynamics features. *TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCES*, 23, 2056–2073. <http://doi.org/10.3906/elk-1302-90>
- [26] Vasuki, P., & Aravindan, C. (2012). Improving emotion recognition from speech using sensor fusion techniques. In *TENCON 2012-2012 IEEE Region 10 Conference* (pp. 1–6). IEEE. Retrieved from [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6412330](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6412330)
- [27] Ververidis, D., & Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech Communication*, 48(9), 1162–1181. <http://doi.org/10.1016/j.specom.2006.04.003>
- [28] Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychological evidence. *Journal of Personality and Social Psychology*, 76(5), 820. Retrieved from <http://psycnet.apa.org/journals/psp/76/5/820/>
- [29] Yun, S., & Yoo, C. D. (2009). Speech emotion recognition via a max-margin framework incorporating a loss function based on the Watson and Tellegen's emotion model. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on* (pp. 4169–4172). IEEE. Retrieved from [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=4960547](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4960547)
- [30] Zhao, X., & Zhang, S. (2015). Spoken emotion recognition via locality-constrained kernel sparse representation. *Neural Computing and Applications*, 26(3), 735–744. <http://doi.org/10.1007/s00521-014-1755-1>
- [31] Zhao, X., Zhang, S., & Lei, B. (2014). Robust emotion recognition in noisy speech via sparse representation. *Neural Computing and Applications*, 24(7–8), 1539–1553. <http://doi.org/10.1007/s00521-013-1377-z>

- [32] Zupan, B., Neumann, D., Babbage, D. R., & Willer, B. (2009). The importance of vocal affect to bimodal processing of emotion: implications for individuals with traumatic brain injury. *Journal of Communication Disorders*, 42(1), 1–17. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0021992408000506>