SPEECH-BASED EMOTION CLASSIFICATION USING MULTICLASS SVM WITH HYBRID KERNEL AND THRESHOLDING FUSION

N. Yang*, *R. Muraleedharan*$^{\oplus}$, *J. Kohl**, *I. Demirkol*†, *W. Heinzelman**, *M. Sturge-Apple*$^{\wedge}$

* Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA
$^{\oplus}$ Department of Electrical and Computer Engineering, Rowan University, Glassboro, NJ, USA
† Department of Telematics Engineering, Universitat Politecnica de Catalunya, Barcelona, Spain
$^{\wedge}$ Department of Clinical and Social Sciences in Psychology, University of Rochester, Rochester, NY, USA

ABSTRACT

Emotion classification is essential for understanding human interactions and hence is a vital component of behavioral studies. Although numerous algorithms have been developed, the emotion classification accuracy is still short of what is desired for the algorithms to be used in real systems. In this paper, we evaluate an approach where basic acoustic features are extracted from speech samples, and the One-Against-All (OAA) Support Vector Machine (SVM) learning algorithm is used. We use a novel hybrid kernel, where we choose the optimal kernel functions for the individual OAA classifiers. Outputs from the OAA classifiers are normalized and combined using a thresholding fusion mechanism to finally classify the emotion. Samples with low 'relative confidence' are left as 'unclassified' to further improve the classification accuracy. Results show that the decision-level recall of our approach for six-class emotion classification is 80.5%, outperforming a state-of-the-art approach that uses the same dataset.

Index Terms— Emotion classification, support vector machine, speaker independent, hybrid kernel, thresholding fusion.

1. INTRODUCTION

Emotion is an essential organizing force of human communication, directing non-linguistic social signals such as body language, facial expression, and prosodic features in the service of communicating wants, needs, and desires. Existing methodologies for assessing behavioral data for emotions are subjective and error-prone. In particular, traditional methods require the use of trained observational coders who manually decode the different parameters in the signal. Such procedures are costly from a time and financial standpoint. While automatic emotion classification has been recently introduced, the classification accuracy is still not adequate. Furthermore, while prosodic features are easy to capture, and thus have been widely used in automatic emotion classification, mining useful emotion information solely from prosodic features is still a challenging task. Therefore, improved emotion classification methods are needed.

There are a variety of applications that can benefit from improved emotion classification. In the healthcare field, emotion classification can be used by clinicians for online assessment of psychological disorders arising from emotional difficulties. Emotion classification may also be an entry point for elaborate context-aware systems for future consumer electronics or services. For example, since smartphones interface heavily with voice, they may be customized to automatically choose songs based on the user's current emotion.

To address the accuracy gap in existing emotion classification approaches, we propose an emotion classification method based on speech prosodic features and an enhanced Support Vector Machine (SVM) with a novel hybrid kernel function and a fusion mechanism with a relative confidence classification threshold. The proposed emotion classification solution extracts the speech signal's pitch, energy and other acoustic features, and the widely employed SVM learning algorithm is used for individual one-against-all (OAA) emotion classification, with a novel hybrid approach for the individual kernel selections. These individual OAA classifications are then normalized and combined using a thresholding fusion algorithm to improve the classification accuracy.

Thus, the contributions of this work are: 1) we optimize the SVM kernel functions for each OAA classifier to achieve the optimal performance; 2) we apply the z-score normalization and a fusion approach to combine the individual OAA classifications; and 3) we introduce a threshold parameter in our fusion mechanism, such that only classifications with high enough relative confidence are considered, and those without high relative confidence are left as 'unclassified'. We apply our approach to emotion classification to the LDC database [1]. The average classifier-level accuracy of previous work [2] that also uses the LDC dataset is only 55.8% for six-class emotion classification. Our method has improved

[1]This research was supported by funding from National Institute of Health NICHD (Grant R01 HD060789).

the accuracies to 96.4%, with an overall decision-level recall of 80.5%. The emotion classification performance is further improved when we increase the relative confidence threshold.

2. RELATED WORK

Speech-based emotion classification applications have already been used to facilitate interactions in our daily lives. For example, call centers apply emotion classification to prioritize impatient customers [3]. As another example, a warning system has been developed to detect if a driver exhibits anger or aggressive emotions [4]. Emotion sensing has also been used in behavior studies [5][6].

We survey existing emotion classification techniques according to their $acoustic\ features$ and $classifier(s)$. As discussed in [7], acoustic features have been extensively explored in both the time domain (energy, speaking rate, duration of voiced segments, zero crossing rate, etc.) and the frequency domain (pitch, formant, Mel-frequency cepstral coefficients, etc.). In our work, we only choose the most basic features: pitch, energy, and formants. This reduces the computational complexity of the approach and can lead to both energy and bandwidth savings when the voice is captured on mobile devices.

Commonly used classifiers for emotion classification include Support Vector Machines (SVM) [2], Gaussian Mixture Models (GMM) [8], and k Nearest Neighbors (kNN) [9]. We choose SVM as our basic classifier due to its ease of training and its ability to work with any number of attributes. The One-Against-All (OAA) approach is one of the approaches used for multi-class SVM. This strategy consists of constructing one SVM per class, with each SVM trained to distinguish the samples of one class from the samples of all remaining classes. Conventional OAA approaches combine these binary decisions from multiple OAA classification models in different ways, such as anding binary decisions [10], or choosing the class with the largest confidence value [11]. We use an approach that considers the relative confidence level of the two classes with the highest confidence values, and considers an emotion 'unclassified' if this relative confidence is not higher than a pre-set threshold. Relative confidence has been used in standard SVM approaches [12] as well as in classification for biomedical applications [13]. However, to the best of our knowledge, this is the first study that uses relative confidences for emotion classification.

For problems solved by using SVM, kernel functions are used to map data to a higher dimensional feature space without losing the originality. The conventional method of using kernel functions in SVM is to run simulations on training sets and find the kernel function that attains the highest averaged classification accuracy for the given problem. The most commonly used kernel function for SVM is radial basis function (RBF) [14]. Recently, researchers have concluded that no single kernel function provides an optimal solution. Thus, Multiple Kernel Learning (MKL) was proposed in [15], where the decision rule is a weighted linear combination of multiple single kernel function outputs. However, all the previous work only uses one type of kernel for all OAA classifiers, either a single kernel or a weighted kernel. In our approach, on the other hand, we choose the best kernel function for each individual OAA classifier in a multiclass classification method. In this paper, we also use an SVM normalization method called z-score [16] in the emotion classification problem to transform individual OAA classifier outputs to comparable values, which further improves the speech-based emotion classification accuracy.

3. ACOUSTIC FEATURES EVALUATED

We classify the emotion of each speech utterance in the LDC database [1]. Each utterance is between one and two seconds in length. We separate each utterance into 60 ms segments with 10 ms time shifts. Since human speech consists of audible and inaudible segments, we only analyze the acoustic features for the audible segments, and we ignore the inaudible ones. A segment is selected to be an audible segment if a certain percentage of the samples' absolute amplitudes in that segment are above a certain threshold.

Digital signal processing methods are applied on each audible segment to extract the acoustic features. In order to reduce the computational complexity of the approach, we only choose the most basic features: pitch, energy, and formant (frequency and bandwidth for the first four formants). Since the change in acoustic features is also related to emotional states, we include the pitch difference and energy difference as additional features. The 12 acoustic features employed are as follows:

- $Pitch$: pitch is defined as the relative highness or lowness of a tone as perceived by the ear. It depends on the number of vibrations per second produced by the vocal cords. We use cepstrum in the frequency domain to extract pitch values.
- $Energy$: energy represents the loudness of the speech. We calculate the energy for each segment by taking the summation of all the squared values of the samples' amplitudes.
- $Pitch\ difference$ and $energy\ difference$: the difference of pitch or energy values between two neighboring segments. More fluctuations may indicate active emotions, such as happiness and anger.
- $Formant$ (frequency and bandwidth for the first four formants, thus eight features): formants are determined by the shape of the vocal tract, and are influenced by different emotions. For example, high arousal results in higher mean values of the first formant frequency

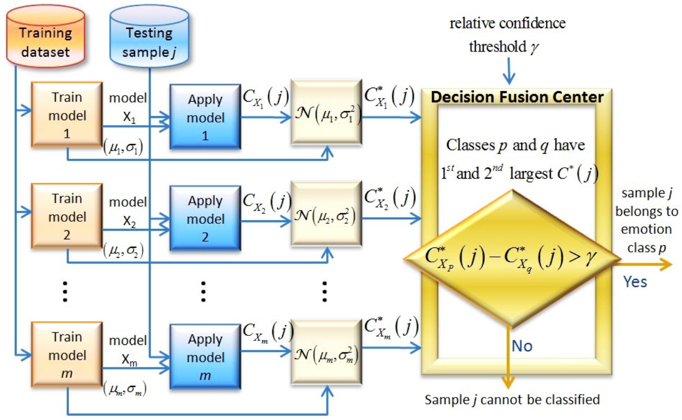

Fig. 1. The proposed emotion classification approach using OAA SVM with hybrid kernel and thresholding fusion.

in all vowels, whereas positive valence results in higher mean values for the second formant frequency [17]. We use the popular linear predictive coding method [18] for formant calculation.

We find these 12 features for each 60 ms segment of the speech sample, and then we calculate the mean, the maximum, the minimum, the range, and the standard deviation for each feature, resulting in $12 \times 5 = 60$ attributes that are sent to the classifier.

4. EMOTION CLASSIFICATION USING MULTICLASS SVM WITH HYBRID KERNEL AND THRESHOLDING FUSION

In this section, we present an enhanced one-against-all (OAA) support vector machine (SVM) classification method with a hybrid kernel and a thresholding fusion method, shown in Figure 1. The acoustic attributes described in Section 3 are used as inputs (training dataset and testing samples) to the SVM classifier.

4.1. Support vector machine

SVM uses binary classification based on statistical learning theory [10], where data are represented in a higher dimensional space using kernel functions to achieve a maximum margin hyperplane. Although many kernel functions exist, only one function can satisfy the maximum margin condition with minimal classification error. SVM uses the structural risk minimization (SRM) principle [10], which avoids overfitting of training models. Therefore SVM has the ability to generalize new data accurately using trained models designed in the learning phase.

4.2. Classification method: One-against-all (OAA)

For multiclass classification problems, methods such as one-against-all, multi-class ranking and pairwise SVM approaches can be applied. A one-against-all (OAA) SVM was proposed in [10], where a data point can be classified only if one of the SVM classes accepts the data point while all other SVMs reject it at the same time, thus making a unanimous decision. In order to reduce the amount of unclassified data in the OAA approach, the author in [11] chooses the class with the largest confidence value.

4.3. Our approach: OAA SVM with hybrid kernel and thresholding fusion

Using the ‘largest confidence value’ fusion approach, the classification accuracy in OAA SVM is dependent on the highest confidence value among the SVM classes, and can therefore increase the misclassification rate when two or more SVM classes have relatively similar confidence values. For example, in an emotion classification application, opposing emotions such as sadness and happiness can be easily classified, while similar emotions such as anger and disgust can be misclassified using the OAA approach. Thus ‘relative confidence’ needs to be considered to evaluate the level of confidence in classifying a data point.

Figure 1 illustrates our proposed OAA SVM classification with fusion, which comprises of learning and testing phases. In the learning phase, for each utterance, the 60 attributes and the emotion labels are used to train the OAA SVM models X_i, where i = *1,2,...,m*, and *m* denotes the number of emotion classes. The ‘best’ kernel function among linear, radial basis function (RBF), quadratic, polynomial, and multi-layer perceptron (MLP) is found for each training model, resulting in a hybrid kernel for the classifiers. The statistical values including the mean μ_i and the standard deviation σ_i for the confidence values of the training data are calculated.

In the testing phase, attributes of the testing utterance *j* are sent to each trained model X_i, resulting in confidence measure $C_{X_i}(j)$, where *j* = *1,2,...,n*, and *n* denotes the number of testing utterances. The confidence measure $C_{X_i}(j)$ is normalized using the z-score normalization $N(\mu_i, \sigma_i^2)$ [16], resulting in the normalized confidence:

$$C^*_{X_i}(j) = \frac{C_{X_i}(j) - \mu_i}{\sigma_i} \qquad (1)$$

The normalized confidence measure $C^*_{X_i}(j)$ is then sent to the fusion center, where the values from each model X_i are sorted in a descending order, and their relative confidence measure *R(j)* is calculated as $C^*_{X_p}(j)$ - $C^*_{X_q}(j)$, where models X_p and X_q yield the first and the second highest normalized confidence $C^*_{X_i}(j)$, respectively. *R(j)* is then compared against the user-controlled relative confidence threshold γ to attain minimal misclassification error. For example, opposing emotions tend to have data points that can be easily separable using a hyperplane, whereas similar emotions such as anger and disgust are more likely to be misclassified, hence setting γ above 0.1 can lead to fewer type I errors (false positives) at the expense of an increased amount of unclassified data. If the relative confidence threshold γ is set to 0, our thresholding fusion mechanism is the same as the 'largest confidence value' fusion mechanism proposed in [11].

5. EVALUATION

5.1. Speech prosody database

The LDC speech database [1] is a standard benchmark library for emotion and speech processing research. This corpus contains speech performed by professional actors and actresses speaking semantically neutral number and date utterances (e.g., 2001, December 8) with different emotions. There are three male speakers and four female speakers in the LDC database. For every emotion category, there are about 15-25 utterances spoken by each speaker.

The advantage of using this library is that emotional utterances generated by professionals are more easily recognized. An alternative is to use speech material from movies or recordings of everyday life. However, it is difficult to determine appropriate reference labels, since many natural utterances are emotionally ambiguous.

5.2. Experimental results and analysis

In our experiments, the SVM toolbox in MATLAB with built-in kernel functions is used. The optimized parameter setting for SVM is based on sequential minimization optimization (SMO) applied on the training set.

5.2.1. Hybrid kernel selection

In our first set of experiments, we train classifiers for the following emotions: disgust, happiness, anger, sadness, neutral and fear. In this set of experiments, the threshold parameter γ is set to 0. We run 7-fold cross-validation tests on the entire dataset for all of the built-in kernels (kernel parameters in the parentheses): linear, quadratic, polynomial (polyorder=3), MLP (default scale [1 -1]), and RBF (sigma=200). We call the recall and accuracy derived from the binary decisions of OAA classifiers as 'classifier-level recall (CL-recall)' and 'classifier-level accuracy (CL-accuracy)', respectively. Note that these values are not used for the final emotion classifications, which are derived after fusing the OAA binary decisions. The 'CL-recall' and 'CL-accuracy' for classifier X_i, i.e. for the classification of an instance as 'Emotion i or Not' are defined as:

$$CL\text{-}recall_{X_i} = \frac{Ctp_i}{Ctp_i + Cfn_i}, \tag{2}$$

and

$$CL\text{-}accuracy_{X_i} = \frac{Ctp_i + Ctn_i}{Ctp_i + Ctn_i + Cfp_i + Cfn_i}, \tag{3}$$

where Ctp_i, Ctn_i, Cfp_i, Cfn_i denote the number of classifier-level true positive, true negative, false positive, and false negative utterances, respectively. We use $CL\text{-}recall_{X_i}$ to select the 'best' kernel for classifier X_i.

Table 1 shows the CL-recall values for the different kernels. For each OAA classifier, we choose the kernel with the highest CL-recall (numbers in bold in Table 1). For example, we choose the polynomial kernel for the 'Happy or Not' OAA classifier in our hybrid kernel approach. Note that when training individual SVM OAA classifiers, we make sure that there are a comparable number of training utterances for both classes, otherwise, the trained model will provide biased classification results. For example, when training the 'Happy or Not' OAA classifier, 'not happy' utterances consist of all of the other five emotions. Therefore, we need to sample utterances from the five emotions with the same number as that of happy utterances. The selected kernel functions for the individual OAA classifiers are shown in Table 1.

Kernels	Anger	Sadness	Disgust	Neutral	Happiness	Fear
Linear	76.9	95.4	98.7	**100**	70.5	73.0
Quad.	**96.2**	98.2	95.5	98.0	96.2	**93.1**
Poly.	94.0	**99.4**	**98.7**	100	**98.1**	92.0
MLP	54.4	53.7	69.5	77.3	43.8	39.8
RBF	69.8	86.7	84.4	82.6	75.2	32.8
Hybrid	Quad.	Poly.	Poly.	Linear	Poly.	Quad.

Table 1. Hybrid kernel selection based on classifier-level recall (%).

5.2.2. One-against-all for six emotion categories

To investigate the effect of our approach of multiclass SVM with hybrid kernel functions, output normalization and thresholding fusion, we compare the performance of our approach with the results in [2], which also uses the OAA multiclass SVM approach and the same LDC dataset. In this set of experiments, the threshold parameter γ is set to 0. We compare our approach with the results when using utterance-level prosodic features from [2]. The performance metric used in [2] is the classifier-level accuracy defined in Eq. (3). Therefore, we also show our results in terms of CL-accuracy. Also, 7-fold cross-validation was used in [2], which is the same as ours. As shown in Table 2, the proposed approach improves the mean classifier-level accuracy by 73%.

Compared with classifier-level accuracy or recall, we are in general more concerned with the decision-level results, which are derived after fusing the decisions from the individual classifiers. Therefore, we use the metric 'decision-level recall (DL-recall)', which is defined as:

$$DL\text{-}recall_{Emotion_i} = \frac{Dtp_i}{Dtp_i + Dfn_i}, \tag{4}$$

where Dtp_i and Dfn_i denotes the number of decision-level true positive and false negative utterances, respectively. The DL-recall of the proposed approach is also listed along in Table 2.

Emotions	Reference[2]	Proposed
Anger	63.6	96.2
Sadness	53.2	99.4
Disgust	51.6	97.4
Neutral	53.5	97.1
Happiness	56.7	95.2
Fear	55.9	93.1
Average	**55.8**	**96.4**

(a)

Emotions	DL-recall
Anger	80.8
Sadness	79.6
Disgust	71.4
Neutral	81.5
Happiness	83.8
Fear	86.1
Average	**80.5**

(b)

Table 2. a) Comparison of classifier-level accuracy (%) of the results in [2] and the proposed approach; b) decision-level recall (%) of the proposed approach.

5.2.3. Tradeoff between emotion classification accuracy and rejection rate

In the previous experiments, we set the relative confidence threshold γ to 0 such that the classifier with the highest normalized confidence was selected as the classified emotion, and there were no 'unclassified' utterances. As the relative confidence threshold γ increases, more testing instances with low relative confidence values are considered to be 'unclassified'. We define the ratio of unclassified instances in the test set as $rejection\ rate$. To measure the average classification performance for all emotions, we define the metric 'decision-level % of correct classification' as:

$$DL\text{-}\%correct = \frac{\sum_{i=1}^{m} Dtp_i}{\sum_{i=1}^{m} (Dtp_i + Dfn_i)}, \qquad (5)$$

where m denotes the number of emotion classes.

With a higher rejection rate, the emotion classification performance can be further improved. To show this, we run 3-fold cross-validation on all utterances from the 6 emotion classes, and determine the decision-level % of correct classification as we vary the threshold γ. Note that we use 3-fold instead of 7-fold cross-validation because that as the rejection rate increases, less data is classified. Thus, there is not enough data for testing if we use the 7-fold cross-validation. As shown in Figure 2, for small relative confidence threshold values, the decision-level % of correct classification increases as the threshold increases, but more instances are rejected. However, for large relative confidence threshold values, very few instances remain. Thus, any misclassification will greatly reduce the % of correct classification value. Therefore, there is a maximum value beyond which γ should not be increased. In Figure 2, if we set the threshold γ to 0.7, we achieve 87.4% decision-level correct classification, an increase from 80.5% when γ is set to 0. However, this increase is achieved by

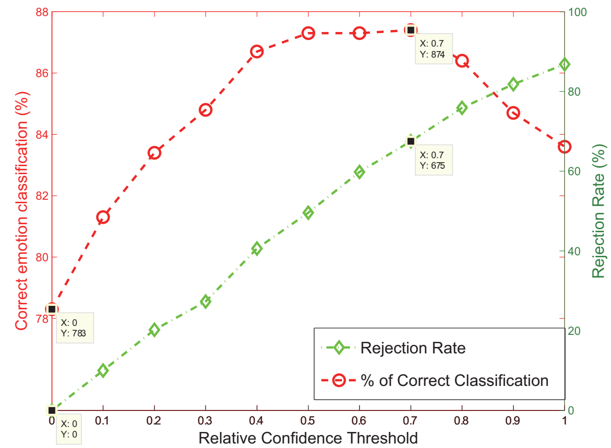

Fig. 2. Rejection rate and decision-level % of correct classification vs. relative confidence threshold γ.

leaving 67.5% of the instances unclassified. Therefore, the thresholding enables the users to exploit the trade-off between the decision-level % of correct classification and the rejection rate.

5.2.4. Speaker-independent emotion classification

To test the performance of the proposed emotion classification method for any new speakers, we run a speaker-independent test (leave-one-subject-out), in which data from the target speaker is unseen in the training phase. We run the test in a Round Robin fashion to test on each speaker with the model trained on all the other six speakers using 3-fold cross-validation. Therefore, we obtain different hybrid kernels for each round of the tests. Finally, the accuracy of each emotion is averaged over all seven speakers. In this set of experiments, the relative confidence threshold parameter γ is set to 0. Additionally, we run cross-validation on all the seven speakers. Both speaker-independent and speaker-dependent DL-recall values are presented in Table 3. Note that the speaker-dependent DL-recall values here are slightly different from those in Table 2(b), which are obtained from 7-fold cross-validation. We can see that speaker-independent DL-recall values are only less than half of the speaker-dependent DL-recall values. It has been proposed in [8] that we should use different feature sets to improve the accuracy of speaker-independent emotion classification.

Emotions	Speaker-independent	Speaker-dependent
Anger	25.9	77.5
Sadness	29.7	79.9
Disgust	27.8	75.9
Neutral	49.6	76.2
Happiness	22.0	79.3
Fear	25.4	81.0
Average	**30.1**	**78.3**

Table 3. Comparison of speaker-independent and speaker-dependent decision-level recall (%) of the proposed approach.

6. CONCLUSIONS

In this paper, we present a speech-based emotion classification method using multiclass SVM. In order to achieve a high emotion classification accuracy, we propose a novel hybrid kernel and we apply a thresholding fusion mechanism. Additionally, SVM outputs are normalized using z-score. The averaged emotion classification accuracy achieved by the hybrid kernel is higher than any of the single built-in kernel functions in the MATLAB SVM toolbox. By comparing with related work that also uses the same LDC dataset for six-class emotion classification, we see that our method has improved the classifier-level accuracy from 55.8% to 96.4%, and achieved a decision-level recall of 80.5%. We also show that by increasing the relative confidence threshold, we can further improve the decision-level % of correct classification to as high as 87.4% for six class emotion classification, at the expense of 67.5% of the data being left unclassified. Finally, we show the benefit of using speaker-dependent emotion classification.

7. REFERENCES

[1] "Emotional prosody speech and transcripts from the linguistic data consortium," http://www.ldc.upenn.edu/Catalog/catalogEntry.jsp?catalogId=LDC2002S28.

[2] D. Bitouk, V. Ragini, and N. Ani, "Class-level spectral features for emotion recognition," *Journal of Speech Communication*, vol. 52, no. 7-8, pp. 613–625, 2010.

[3] P. Gupta and N. Rajput, "Two-Stream Emotion Recognition for Call Center Monitoring," in *INTERSPEECH*, 2007, pp. 2241–2244.

[4] F. Al Machot, A. H. Mosa, K. Dabbour, A. Fasih, C. Schwarzlmuller, M. Ali, and K. Kyamakya, "A novel real-time emotion detection system from audio streams based on bayesian quadratic discriminate classifier for adas," in *Nonlinear Dynamics and Synchronization 16th Int'l Symposium on Theoretical Electrical Engineering, 2011 Joint 3rd Int'l Workshop on*, 2011.

[5] D. Tacconi, O. Mayora, P. Lukowicz, B. Arnrich, C. Setz, G. Troester, and C. Haring, "Activity and Emotion Recognition to Support Early Diagnosis of Psychiatric Diseases," in *Pervasive Healthcare*, 2008.

[6] R. Cowie, E. Douglas-Cowie, S. Savvidou, E. McMahon, M. Sawey, and M. Schröer, "Feeltrace: An instrument for recording perceived emotion in real time," in *Proceedings of the ISCA Workshop on Speech and Emotion*, 2000, pp. 100–102.

[1]The MATLAB code used to produce the results in this paper is available for download on the website of the Wireless Communication and Networking Group at the University of Rochester [19].

[7] D. Ververidis and C. Kotropoulos, "Emotional speech recognition: Resources, features, and methods," *Speech Communication*, vol. 48, no. 6, pp. 1162–1181, September 2006.

[8] V. Sethu, E. Ambikairajah, and J. Epps, "Empirical mode decomposition based weighted frequency feature for speech-based emotion classification," in *ICASSP*, 2008, pp. 5017–5020.

[9] M. Shamia and M. Kamel, "Segment-based approach to the recognition of emotions in speech," in *Proceedings of IEEE International Conference on Multimedia and Expo 2005*, 2005.

[10] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag New York, Inc., 1995.

[11] V. N. Vapnik, *Statistical Learning Theory*, Wiley-Interscience, 1998.

[12] A. Vlachos, "Active learning with support vector machines," Tech. Rep., School of Informatics, University of Edinburgh, 2004.

[13] A. Garg, M. Bhasin, and G. P. S. Raghava, "Support vector machine-based method for subcellular localization of human proteins using amino acid compositions, their order, and similarity search," *The Journel Of Biological Chemistry*, vol. 280, no. 15, pp. 14427–14432, 2004.

[14] M. H. Song, J. Lee, S. P. Cho, K. J. Lee, and S. K. Yoo, "Support vector machine based arrhythmia classification using reduced features," *International Journal of Control, Automation, and Systems*, vol. 3, no. 4, pp. 571–579, 2005.

[15] M. Gönen and E. Alpaydin, "Multiple kernel learning algorithms," *Journal of Machine Learning Research*, vol. 12, pp. 2211–2268, 2011.

[16] M. Farrús, P. Ejarque, A. Temko, and J. Hernando, "Histogram equalization in svm multimodal person verification," in *Proceedings of IAPR/IEEE International Conference on Biometrics*, 2007.

[17] M. Goudbeek, J. P. Goldman, and K. R. Scherer, "Emotion dimensions and formant position," in *Interspeech*, 2009.

[18] L. B. Jackson, *Digital Filters and Signal Processing*, Kluwer Academic Publishers, 1989.

[19] "Wireless Communication and Networking Group, University of Rochester," http://www.ece.rochester.edu/projects/wcng/project_bridge.html.