Action Unit Models of Facial Expression of Emotion in the Presence of Speech

Miraj Shah[1], David G. Cooper[1], Houwei Cao[1], Ruben C. Gur[2], Ani Nenkova[3], and Ragini Verma[1]
[1]Section of Biomedical Image Analysis, Department of Radiology
[2]Department of Psychiatry
[3]Department of Computer & Information Science
University of Pennsylvania
Philadelphia, PA19104, United States
mirajs@seas.upenn.edu, david.g.cooper@uphs.upenn.edu, houwei.cao@uphs.upenn.edu,
gur@mail.med.upenn.edu, nenkova@seas.upenn.edu, ragini.verma@uphs.upenn.edu

***Abstract*—Automatic recognition of emotion using facial expressions in the presence of speech poses a unique challenge because talking reveals clues for the affective state of the speaker but distorts the canonical expression of emotion on the face. We introduce a corpus of acted emotion expression where speech is either present (talking) or absent (silent). The corpus is uniquely suited for analysis of the interplay between the two conditions. We use a multimodal decision level fusion classifier to combine models of emotion from talking and silent faces as well as from audio to recognize five basic emotions: anger, disgust, fear, happy and sad. Our results strongly indicate that emotion prediction in the presence of speech from action unit facial features is less accurate when the person is talking. Modeling talking and silent expressions separately and fusing the two models greatly improves accuracy of prediction in the talking setting. The advantages are most pronounced when silent and talking face models are fused with predictions from audio features. In this multi-modal prediction both the combination of modalities and the separate models of talking and silent facial expression of emotion contribute to the improvement.**

Keywords—emotion; talking; silent; multimodal; face; voice.

I. INTRODUCTION

People convey their emotional state both in their facial expression and their voice, normally in a mix of talking and silence occurring intermittently. A rich tradition of research in multimodal emotion recognition has thoroughly documented the benefit of using simultaneously visual and audio modalities [1,2,3,4,5,13]. However there is little research in determining the exact effect that talking has on the recognition of facial expressions and subsequent analysis of emotion. It stands to reason that the act of talking introduces facial movement unrelated to emotion expression. Studies have shown that facial expression recognition in the presence of speech is hard [21,2] but it remains an open question to quantify and productively exploit the differences in facial expression of emotion when the person is talking or silent.

In this paper we present a comprehensive analysis of the degradation of facial emotion recognition that occurs when the subject is talking. We show that the composition of the training set in terms of talking and silent data significantly impacts prediction performance. More importantly, we find that the fusion of different models of facial expression of emotion leads to significant improvement and thus is key for properly handling facial emotion in the presence of speech. The effect of combining the two facial models leads to largest improvements when audio information is incorporated as well. As in previous studies, combining audio and facial features is beneficial but the best results come from combining both talking and silent models of the face with audio.

Our facial expression classifiers use automatically detected facial action units (AUs) [10,11]. We train and evaluate the subject-independent performance of an SVM classifier on our specially designed dataset of evoked emotion. The dataset is large, containing videos of 91 professional actors displaying various degrees of emotion in face and voice for each of the five basic emotions (anger, disgust, fear, happy or sad).

II. APPROACH

We train a linear Support Vector Machine (SVM) with the LibLinear library [22] to classify both facial and voice expression of emotion. Before giving further details of the classifiers and the fusion of modalities, we describe the dataset on which we base our experiments.

A. Data Acquisition

We created a dataset of posed anger, disgust, fear, happy, and sad emotional states. We recorded frontal video of the head and torso, as well as the voice of professional actors as they express emotions at varying degrees of intensity. An on-staff director supervised the actors and approved the renditions included in the final database. In some videos the actors express silent emotion, while in others they say a semantically neutral sentence as they act out the emotion, using both their voice and face to create the emotional impression.

We use data from 74 actors for training and data from 17 other actors for testing. We distinguish three datasets which correspond to different partitions of the collected expressions.

Silent set, in which each of the actors emote without talking. Videos in the silent set present emotional expression progressing from Neutral to Apex and back to Neutral. Each actor portrays each emotion in three levels of intensity---low, medium, and high---yielding 15 clips per actor for the five emotions of anger, disgust, fear, happy and sad. In total there are 1,062 videos in the silent training set: 213 for anger and fear, 211 for happy and sad and 214 for disgust. There are about 10 recordings for each emotion which could not be processed as part of the dataset because of failure in facial

features detection by CERT, which is the software we use to extract facial features.

Talking set, in which the actors express each emotion while uttering 12 different sentences rated as semantically neutral in a prior study [15].

Don't forget a jacket.
I'm on my way to the meeting.
I think I have a doctor's appointment.
I think I've seen this before.
I wonder what this is about.
I would like a new alarm clock.
Maybe tomorrow it will be cold.
The airplane is almost full.
We'll stop in a couple of minutes.
That is exactly what happened.
The surface is slick.
It's eleven o'clock.

For all but the last sentence, the target intensity of the emotion was not specified to the actors. They were asked to display three levels of emotion for the last sentence. In total, there are 5,180 talking expressions in the dataset, with 1,036 instances for each of the five emotions. Each actor recorded both the silent and the talking condition for each emotion.

Combined set, which is the union of silent and taking instances. The talking set is much larger than the silent one, so we create two types of combined sets: the **full** combined set in which the proportion of talking and silent instances is unbalanced and the **balanced** combined set where talking instances are randomly downsampled so that there is equal number of talking and silent instances in the dataset. The random selection of downsampled balanced set was performed ten times. Ten classifiers were trained on the different samples. To produce a single prediction in testing, the probability of each emotion class from the ten classifiers was averaged and the emotion with highest average probability was predicted to be the label emotion of that clip.

B. Video-Based Classification

We use Action Units (AUs) as features for the face classifier. We extract AUs using CERT [11].

1) Action Unit Detection: Facial Action Units (AU) are commonly used to determine the emotional expression of a person. They are often detected by combining facial feature point tracking and texture features in particular regions of interest. For each AU, CERT outputs a continuous value at each frame, which is the distance of the feature point from the seperating hyperplane of the classifier for detecting that AU. These values are highly correlated with expert annotation of action unit intensity in posed emotions. The correlation is moderate for spontaneous emotion expressions in the presence of speech [21].

We use 15 action units that are detected with higher precision [10] to represent each video frame; the intensity of these action units is used as the feature representation for automatic prediction. These include seven features from the upper part of the face (eyes and cheeks) and 8 features on the lower part of the face (lips and chin). These features are inner brow raiser (AU1), outer brow raiser (AU2), brow lowerer (AU4), upper lid raiser (AU5), cheek raiser (AU6), lid tightener (AU7), nose wrinkler (AU9), upper lip raiser (AU10), lip corner puller (AU12), lip corner depressor (AU15), chin raiser (AU17), lip pucker (AU18), lip stretcher (AU20) , lip tightener (AU23) and lips part (AU25).

2) Multi-class classification: We construct a linear Support Vector Machine (SVM) using ℓ-2 regularization and ℓ-2 loss for video classification. We first perform frame-level prediction, then derive a prediction for the full video. We first subsample the video clip, selecting every fifth frame in the sequence to reduce redundancy. Then we automatically select the k most expressive frames for each video clip. To do that, we find the frame in the video for which the sum of AU intensity scores predicted by CERT is lowest. We hypothesize that a frame with low scores for an AU would correspond to a neutral face and search for the frames that differ most from that neutral one. All other frames are scored according to the ℓ-1 norm between their AU representation and that of the neutral frame, and the k frames with the highest score are chosen. We searched for the best value for k for values between 3 and 9 with 10 fold cross-validation on the training set. Some short videos consist of only 9 frames, so we did not test higher number of frames. The best cross-validation results were for k=9 on both the silent and talking training sets.

A multi-class SVM classifier is then trained on the n training instances for a total of $n \times k$ training frames for the classifier. Each frame was labeled as the emotion that the actor was expressing in that video.

For testing, the SVM classifier outputs predictions for the k best frames of each testing clip. We compute the overall probability for each emotion class e_i, given the AU feature vector V^j for each clip j, $p(e_i|V^j)$ using Equation (1). The emotion class yielding the highest probability is chosen as the final prediction:

$$p(e_i|V^j) = \frac{\sum_{f=1}^{k} p(e_i|V_f^j)}{\sum_{i=1}^{5}\sum_{f=1}^{k} p(e_i|V_f^j)}, \; i=1,\ldots.5 \qquad (1)$$

C. Audio-Based Classification

For the audio based classification, we also train a linear SVM using ℓ-2 regularization and ℓ-2 loss classifier using the audio from the talking data set.

We use the openSMILE feature extraction library [17] to obtain a comprehensive set of standard acoustic features to characterize each utterance. The openSMILE library extracts 26 low-level descriptors including intensity, loudness, F0, F0 envelope, probability of voicing, zero-crossing rate, 12 MFCCs, and 8 LSFs. We also use the first order delta coefficients for these features, as well as 19 summary functions for a total of 988 features.

D. Fusion

Fusion of different modalities is done at the decision level by finding the emotion that has highest sum of probabilities from all modalities, according to Equation (2). Here *m* is the number of classifiers that are being combined. The probabilities that are added are those described in Equation (1). We first get the probability of each emotion class from the single-modality classifiers, and then we combine them together without weights.

.

$$L = \quad argmax_i \sum_{j=1}^{m} p(e_i)_j \ , i=1,....5 \qquad (2)$$

where classes e_i, i=1..5 correspond to anger, fear, disgust, happy, and sad, *m* is number of classifiers being combined, $p(e_i)_j$ denotes the posterior probability of i^{th} emotion for the j^{th} classifier.

We use SVM classifiers trained on talking, silent and the combined training set for the fusion experiments, as well as a classifier based on audio features, for a total of five different classifiers. In addition, we evaluate the individual performance of facial models.

III. EXPERIMENTS

Data from 74 actors are used for training, as well as for parameter tuning via cross-validation. Data from the remaining 17 actors form the test sets.

The talking test set consists of 1,190 videos, 238 videos for each emotion. The silent test set has 242 videos: 48 for each of anger, disgust, happy and sad, and 50 for disgust. The parameters for classifier (cost and epsilon values) are optimized with a grid search on the training data using 10-fold actor-independent cross-validation.

We analyze the difference in performance of the different models on the silent and talking test set separately. We also perform a paired Sign-Test to check statistical significance of these differences.

IV. RESULTS

Table I shows the overall accuracy obtained on the talking and silent test set when trained on different datasets indicated in the first column. X+Y represents the fusion classifier given by Equation (2) using posterior probabilities of unimodal multiclass SVM classifiers X and Y e.g. A+S+T represents fusion of audio, silent and talking modalities.

We test for statistical significance between all classifiers and the classifier that explicitly combines silent and talking models of emotion expression: S+T for face only prediction and A+S+T for prediction that relies on acoustic features as well. The goal is to find out in which situations the combination that explicitly models emotion on the talking and silent face separately behaves differently from the alternatives.

First we discuss results that rely only on facial features. It is immediately clear in this case, that the accuracy of emotion prediction in the presence of speech is worse than that on a silent face, regardless of the type of training set and fusion. The best overall prediction on silent faces, where the actor does not speak, is achieved by the classifier trained on silent video. The model trained on the talking training set is almost 3% worse. Interestingly the second-best model, just 0.18% worse in absolute accuracy, is the one trained on the balanced combined dataset. This balanced set provides the easiest way of incorporating of information about the distortion of facial expression during speech, by simply providing a training set with more variation. Even in this case, the mix of silent and talking videos affects performance. The classifier trained on the full unbalanced set is almost 2% less accurate than the one trained on the balanced one. Here, fusion of different facial models does not lead to improvements over single types of training data for prediction in the absence of speech. None of the differences with the fusion of talking and silent classifier (S+T) is statistically significant.

For emotion prediction in the presence of speech from video features, the best overall performance comes from the fusion of models of silent, talking and combined faces. The improvement that combination gives over the fusion of silent and talking only tends towards significance. The classifier trained only on silent data is significantly worse than the fusion of silent and talking prediction. Again, the model trained on balanced combined dataset is the one with highest accuracy when a single classifier is used.

For all classifiers, the performance on silent faces is higher than that on videos where speech is present. For a classifier trained on silent videos the degradation is worse, close to 7%. The classifier that is most robust to the change of training set is the fusion of silent, talking and combined facial expression. Its accuracy on the talking test set is only about 2% less than that on the silent test set.

Table I. % Accuracy of different facial modalities **with/without** audio on talking and silent test sets.

Experiments **without fusing audio model**(*significance paired signed test compared with S+T)			Experiments **with fusing audio** (*significance paired signed test compared with A+S+T)	
Type of training set	Accuracy (%) tested on silent	Accuracy (%) tested on talking	Type of training set	Accuracy (%) tested on talking
-	-	-	Audio(A)	60.08
Talking(T)	57.44	53.74	A+T	70.75***
Silent(S)	60.33	52.98**	A+S	71.93***
CFull	58.26	54.29	A+CFull	71.59***
CBal	60.15	56.05	A+CBal	71.79***
S+T	**57.43**	**54.95**	**A+S+T**	**75.04**
S+T+CFull	57.44	56.05	A+S+T+CFull	75.63
S+T+CBal	58.26	56.13*	A+S+T+CBal	75.21

(T: Talking, S:Silent, CFull : Combined Full CBal: Combined Balanced A:Audio) significant p-value for difference with S+T/A+S+T for each modality is indicated by '*' character (***($p<0.001$),**($p<0.05$),*($p<0.1$))

Finally, we discuss the performance of prediction that incorporates facial and audio features, shown in Column 2 of Table I. The prediction is performed only on the talking test set because actors do not speak in the silent videos. The accuracy of only the acoustic classifier is close to 60% and adding any of the facial models leads to at least 10% absolute improvement. More pertinent to our investigation however is the fact that different face models bring in different complementary information about emotional state of the speaker. Contrary to expectation, fusion of audio and the silent facial expression (A+S) leads to better accuracy than fusion of audio and talking face features. However, fusion of acoustic prediction with both silent and talking face models (A+S+T) gives statistically significant improvements of almost 4% over that. As in the case when audio is not involved, the accuracy improves somewhat when the combined model is added to the talking and silent ones, however the difference is not statistically significant.

We again observe that adding silent (A+S+T, A+S+T+CFull, A+S+T+CBal) modality gives significant improvement of approximately 5% over using audio and talking (A+T) alone.

Table II. %Accuracy of individual emotions of different modalities tested on **Talking set**

Type of Training Set	Accuracy(%)					
	Anger	Disgust	Fear	Happy	Sad	Overall
T	7.56	52.94	68.90	93.27	46.02	53.74
S	20.16	62.18	55.88	90.33	36.40	52.98
S+T	16.80	60.00	63.86	92.85	41.17	54.95
A	76.89	51.68	51.26	50.00	70.58	60.08
A+T	79.41	66.8	53.78	78.57	75.21	70.75
A+S	76.05	71.00	59.66	82.35	70.58	71.93
A+S+T	75.21	71.00	66.38	92.85	69.74	75.04

Table III. %Accuracy of individual emotions of different modalities tested on **Silent set**

Type of Training Set	Accuracy(%)					
	Anger	Disgust	Fear	Happy	Sad	Overall
T	14.58	60.41	68.00	95.83	47.91	57.44
S	29.16	60.41	66.00	93.75	52.08	60.33
S+T	14.58	60.41	68.00	95.83	47.91	57.43

Table II shows individual emotion accuracy of the different models involving (A, S, T) predicted on the **talking** test set. Anger and disgust are poorly classified when training the facial models only on talking videos (T). This indicates that the activation of AUs describing the lower part of the face, which is related to the speech articulation, negatively affects the prediction of anger and disgust emotion while talking. However fear, happy and sad are better predicted while talking. Thus S and T give complementary information leading to the fusion (S+T) giving better accuracy. We also see that audio features lead to substantial improvement for the prediction of anger and sad. Thus fusing audio with facial classifiers improves the overall accuracy by more than 10% over individual modalities.

Table III shows the individual emotion accuracy for S and T modalities on the **silent** test set. We observe that using Talking (T) modality does not drastically improve the performance of any emotion. Thus it is preferable to use only silent modality to predict emotion during absence of speech.

We further analyze the differences in the activation of AU in silent and talking dataset to better understand the differences in the presence and absence of speech. The AU values of only key frames (as described earlier in section III.B) are considered. We first calculate the Z-score [20] for each action unit according to (3), in the silent and talking dataset to identify the action units most descriptive for each emotion.

$$Z\text{-}score_i(j) = \frac{\mu_{ij}}{\sigma_j} - \frac{\mu_j}{\sigma_j} \qquad i=1..5,\ j=1...15 \qquad (3)$$

where μ_{ij} denotes mean of j^{th} AU for i^{th} emotion, μ_j denotes mean of j^{th} AU over all emotions and σ_j denotes standard deviation of j^{th} AU over all emotions. The higher the value of |Z-score|, the more extreme---high or low---is the activation of an AU in an emotion. We select the top 5 AUs per emotion for the silent and talking datasets respectively, based on maximum |Zscore| value. Fig 1.A indicates which AUs are selected for talking and silent dataset for each emotion. For three emotions---anger, happy and fear---the top five action units that characterize the emotion differ for the silent and the talking datasets. Many of these differences are related to the presence of speech. For anger, AU 18, lip pucker, is among the best descriptors of the emotion on silent face, but obviously this feature loses its power when the person is talking which renders this movement impossible. Instead, AU4, brow lower, characterizes anger. The differences of descriptive action units for the other two emotions are not related to lip movement. It is possible however that facial muscle coordination during speaking may change the possibility to activate even nose or eye-related action units.

Figure 1.B-F show box-plots of the distribution of activation of the union of the top selected AUs for silent and talking dataset. Red plots are for talking and blue plots are for silent dataset. They are placed in a decreasing order of |Zscore| value in the silent dataset. The variance of activation is also displayed on the plots.

We also perform a non-parametric two sided Wilcoxon rank-sum significance test [19] of the null hypothesis that talking and silent activations are independent samples from identical continuous distributions with equal medians, against the alternative that they do not have equal medians. In many cases, the differences in activation were significant. For all emotions but disgust the activation of AU12, lip raiser, and AU6, check raiser, is distributed differently in talking than silent videos, which is not surprising, as these AUs are affected as a result of speech. With very few exceptions, the median activation is significantly different for the two datasets. The variance in the talking dataset is also consistently higher for all action units.

EMOTION	*SELECTED ACTION UNIT*
ANGER	12***,6***,18***,7,1***,4***
DISGUST	4***,9,7**,10***,2***
FEAR	5***,6***,7***,12***,2***,9***
HAPPY	12***,6***,18***,25,4,2***
SAD	25***,10***,6***,12***,9***

Figure 1.A Selection of AU based on zscore value. red indicates AU selected on silent dataset,black indicates selected on talking and green indicates common to both. significant p-value is indicated by '*' (***(p<0.001),**(p<0.01),*(p<0.1))

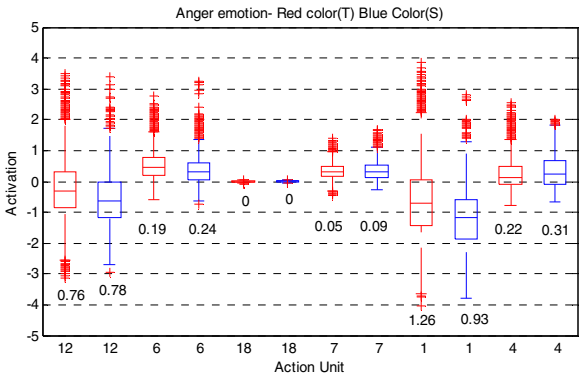

Figure 1.B Emotion Anger: Selected AUs: lip corner puller (AU12), cheek raiser (AU6), lip pucker (AU18), lid tightener (AU7), inner brow raiser (AU1), brow lowerer (AU4).

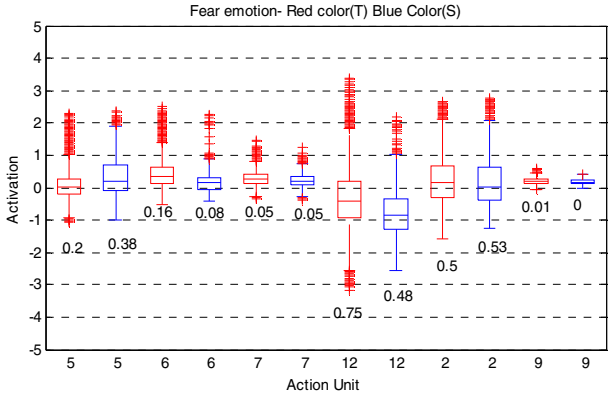

Figure 1.C Emotion Fear: Selected AUs: upper lid raiser (AU5), cheek raiser (AU6), lid tightener (AU7), lip corner puller (AU12), outer brow raiser (AU2), nose wrinkler (AU9).

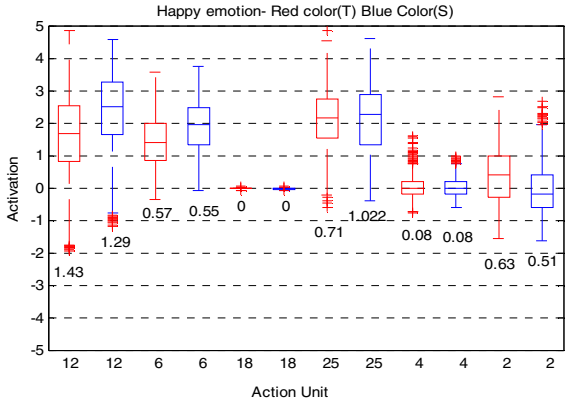

Figure 1.D Emotion Happy: Selected AUs: lip corner puller (AU12), cheek raiser (AU6), lip pucker (AU18), lips part (AU25), brow lowerer (AU4),outer brow raiser (AU2).

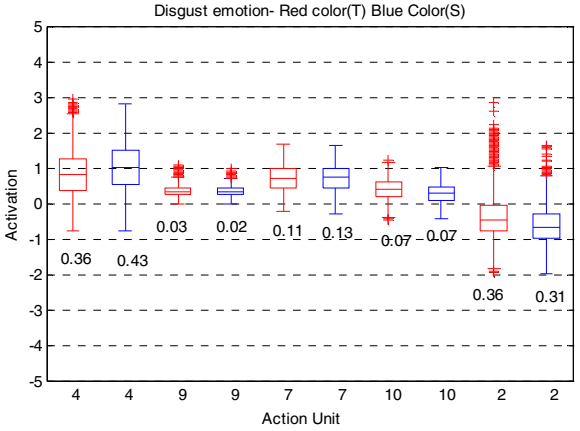

Figure 1.E Emotion Disgust: Selected AUs: brow lowerer (AU4), nose wrinkler (AU9), lid tightener (AU7), upper lip raiser (AU10), outer brow raiser (AU2).

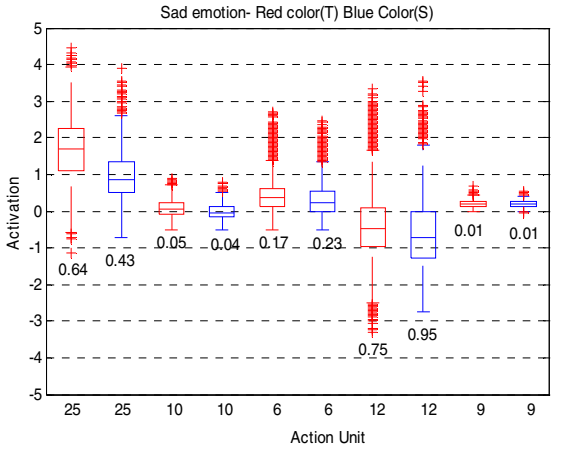

Figure 1.F Emotion Sad :Selected AUs: lips part (AU25), upper lip raiser (AU10), cheek raiser (AU6), nose wrinkler (AU9).

Figure 1. Comparison of AU values of talking (red) and silent (blue) dataset. (values below the box plots in the graph represent variance for each AU)

V. CONCLUSION

In this paper we analyzed the performance of emotion recognition when a person is talking and when a person expresses the emotion in silence, only on the face. We showed that for face analysis based on the widely used action unit coding, the data used for training influences the robustness of the resulting classifier. In the presence of speech, highest accuracy is achieved when the emotion expression on the face is modeled separately for talking, silent and combined faces and fused at the decision level. When speech is not present, the best prediction comes from a model trained on silent expressions. In either case, the model that explicitly fuses prediction from silent and talking expressions is not significantly worse than other combination, indicating that this combination is the most robust one.

REFERENCES

[1] K. Bousmalis, L.-P. Morency and M. Pantic, "Modeling hidden dynamics of multimodal cues for spontaneous agreement and disagreement recognition," in FG, 2011, pp. 746–752.

[2] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. L A. Kazemzadeh, S. Lee, U. Neumann, and S. S Narayanan, "Analysis of emotion recognition using facial expressions, speech and multimodal information," in Proceedings of the International Conference on Multimodal Interfaces, State Park, PA, Oct. 2004, pp. 205–211.

[3] Y. Song, L.-P. Morency, and R. Davis, "Multimodal human behavior analysis: learning correlation and interaction across modalities," in ICMI, 2012, pp. 27–30.

[4] M. Wollmer, A. Metallinou, F. Eyben, B. Schuller, and S. S. Narayanan, "Context-sensitive multimodal emotion recognition from speech and facial expression using bidirectional lstm modeling," in In Proceedings of InterSpeech, Makuhari, Japan, Sep. 2010.

[5] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 31, no. 1, pp. 39–58, 2009.

[6] H.-J. Go, K.-C. Kwak, D.-J. Lee, and M.-G. Chun, "Emotion recognition from the facial image and speech signal," in SICE 2003 Annual Conference, vol. 3, Aug. 2003, pp. 2890 –2895 Vol.3.

[7] N. Sebe, I. Cohen, T. Gevers, and T. Huang, "Emotion recognition based on joint visual and audio cues,"in 18th International Conference on Pattern Recognition (ICPR'2006), IEEE vol. 1, pp. 1136–1139.

[8] J.-C. Lin, C.-H. Wu, and W.-L. Wei, "Error weighted semi-coupled hidden markov model for audio-visual emotion recognition," Multimedia, IEEE Transactions on, vol. 14, no. 1, pp. 142 –156, Feb. 2012.

[9] Y. Wang, L. Guan, and A. Venetsanopoulos, "Kernel cross-modal factor analysis for information fusion with application to bimodal emotion recognition," Multimedia, IEEE Transactions on, vol. 14, no. 3, pp. 597–607, 2012.

[10] J. Hamm, C. Kohler, R. Gur, and R. Verma, "Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders," Journal of Neuroscience Methods, 2011.

[11] Littlewort G, Whitehill J, Wu T, Fasel I, Frank M, Movellan J, and Bartlett M (2011) The Computer Expression Recognition Toolbox (CERT). Proc. IEEE International Conference on Automatic Face and Gesture Recognition, 2011 pp. 298–305.

[13] A. Metallinou, S. Lee, and S. S. Narayanan, "Decision level combination of multiple modalities for recognition and analysis of emotional expression," in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Dallas, Texas, 2010.

[14] T. Wu, N. Butko, P. Ruvolo, J. Whitehill, M. Bartlett, and J. Movellan, "Action unit recognition transfer across datasets," in Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, Mar 2011, pp. 889 –896.

[15] J. B. Russ, R. C. Gur, and W. B. Bilker, " Validation of affective and neutral sentence content for prosodic testing." Behav Res Methods, vol. 40, no. 4, pp. 935–939, Nov 2008.

[16] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 20, no. 3, pp. 226 –239, Mar 1998.

[17] F. Eyben, M. Wöllmer, and B. Schuller, "openSMILE: The Munich versatile and fast open-source audio feature extractor," in Proceedings of the International Conference on Multimedia, MM '10. ACM, 2010, pp. 1459–1462.

[18] A. Austermann, N. Esau, L. Kleinjohann, and B. Kleinjohann, "Prosody based emotion recognition for mexi," in, International Conference on Intelligent Robots and Systems(IROS 2005), pp. 1138–1144.

[19] Frank Wilcoxon," Individual Comparisons by Ranking Methods," *Biometrics Bulletin* , Vol. 1, No. 6 (Dec., 1945), pp. 80-83

[20] Altman, Edward I."Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy"". Journal of Finance: 189–209. Sept 1968.

[21] Bartlett, M., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., & Movellan, J. (2006). Automatic Recognition of Facial Actions in Spontaneous Expressions. *Journal Of Multimedia, 1*(6), 22-35.

[22] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. LIBLINEAR: A Library for Large Linear Classification, Journal of Machine Learning Research 9(2008), 1871-1874