Presented at the DLSU Research Congress 2015
De La Salle University, Manila, Philippines
March 2-4, 2015

Modeling Blended Emotions in Spontaneous Filipino Laughter through Facial Expression Analysis

Katrina Ysabel Solomon[1,*], Jocelynn Cu[1] and Merlin Teodosia Suarez[1]
[1]*Center for Empathic Human-Computer Interactions, College of Computer Studies, De La Salle University*
**katrina.solomon@dlsu.edu.ph*

Abstract: Most of the existing studies that concentrate on emotion recognition are limited to identifying a single emotional state at a given time. However, there are cases in which these states tend to blend. Such event occurs when a person exhibits two emotional states simultaneously and has been observed in real-life settings such as call centers. The concept has not been thoroughly explored especially when it comes to the recognition of emotion in laughter in a Filipino context. This paper aims to identify the different emotional states in Filipino laughter that tend to blend. To accomplish this, the modality that will be used is the face and it will be divided into two regions which are the upper facial region and the lower facial region. The division was done as such because the emotional states are easily distinguishable when observed from a region. The upper facial region favors the display of negative emotions while the lower facial region is leans more on positive emotions. The movement of the facial features will be tracked using the Active Appearance Model algorithm. Given this data, this paper will also ascertain which facial features are significant for each type of emotion so feature selection can be performed without compromising the performance of the classifier. In building the models, the algorithms to use are the Multilayer Perceptron (MLP) and Support Vector Machines (SVM). The accuracy rates of gained by using MLP are 92.59%, 91.45%, and 90.79% for the whole, lower and upper facial regions respectively. Meanwhile, SVM yielded 93.50%, 88.81%, and 91.46% accuracy rates.

Key Words: facial recognition; laughter detection; emotion recognition; blended emotions

1. INTRODUCTION

Laughter is considered as a non-verbal phonetic activity that occurs during interactions in conversations (Truong and Trouvain, 2012). It is characterized by rapid rhythmic shoulders and torso movements, visible inhalations and various facial expressions often accompanied by rhythmic and communicative gestures (Ruch and Ekman, 2001). Laskowski and Burger (2007) observed that laughter is the most frequently annotated acoustic non-verbal behavior in model building.

According to Ruch and Ekman (2001), laughter has two variations: spontaneous or natural and voluntary or fake. In spontaneous laughter, one follows the urge to laugh without holding back anything. There is little or no attempt to suppress the reaction. The opposite can be said about voluntary laughter, in which a person attempts to produce a sound that would resemble natural

laughter. While spontaneous laughter is rich in emotions, voluntary laughter is lacking.

Laughter is a common signal for positive emotions such as happiness and joy, making it a powerful social signal (Petridis and Pantic, 2011). In some cases, however, it has also been observed that laughter is evident in negative emotions such as anger, shame and nervousness (Owren and Bachorowski, 2003). This circumstance is one of the many considered as an occurrence of mixed emotions.

Buisine et al. (2006) enumerated three types of mixed emotions namely: sequenced, masked and blended. Sequenced emotions are characterized by a quick succession of different emotions. Masked emotions involve suppressing or overacting of another emotion in order to conceal the real one. Blended emotions are the result of a superposition of two emotions, one emotion evident on the upper facial region while the other on the lower facial region.

There are various studies on mixed emotions but are limited. Among them include the work of Buisine et al. (2006) that concentrates on using Embodied Conversational Agents (ECAs) to simulate mixed emotions. Douglas-Cowie et al. (2005) focused on annotating said emotions. Devillers at al. (2005) and Vidrascu and Devillers (2005) highlighted the occurrence of mixed emotions in call center recordings. Soladie et al. (2012) proposed an innovative method of modeling facial expressions that exhibit blended emotions.

The objective of this study is to build an affect model for spontaneous Filipino laughter, specifically those exhibiting blended emotions. To achieve this, the focus will be on the occurrences of blended emotions through the tracking of a person's facial expression. As explained by Ruch and Ekman (2001), laughter can also be reviewed and identified by observing the movements of the facial muscles. Also, Buisine et al. (2006) proposed a method of simulating blended emotions by using facial expressions over voice. Analysis of audio will not be part of the study. Though there are several ways of labeling laughter, the classification proposed by Suarez et al. (2012) will be followed.

2. METHODOLOGY

2.1 Data Annotation

The data utilized in this study were collected by Galvan et al. (2011). There are 270 video clips in all. The resolution is set to 720 pixels by 480 pixels. The frame rate is 25 frames per second. The average length is approximately 6.5 seconds. The subjects are all students of De La Salle University. Of the four, two are male and the other two are female.

During the data collection process, the subjects were grouped into pairs. They were placed in separate rooms and conversed with each other through the use of the software Skype as they were being recorded. After the data collection, the clips were segmented and only those clips with laughter segments were retained. The clips were then given a single label according to the classification presented by Suarez et al. (2012). These labels are: *kinikilig, mapanakit, nahihiya, nasasabik and natutuwa*.

To meet the objectives of this research, the date were re-annotated. Instead of each clip having one label, the volunteer annotator provided three labels. One label is the perceived emotion given the whole facial region. Another is the perceived emotion exhibited based from the upper facial region. The last label is the perceived emotion as seen on the lower facial region. The annotator would have to watch a certain clip three times, as seen in Figure 1. For every iteration, one of three facial regions involved in the research is displayed. This allows the annotator to provide the most accurate label for each region without being influenced by the overall appearance of the face. The audio tracks were removed to also prevent additional factors that may influence the labeling process.

Fig. 1. The same video clip viewed by the annotator three times with only the region to be labeled seen in every iteration

After the re-annotation process, clips with the same labels for all three regions or both the upper and lower facial region were identified as exhibiting single state emotions and not blended emotions. Only those with different labels for the upper and lower facial region were considered. The labels used were *kinikilig, mapanakit, nahihiya, nasasabik and natutuwa*.

2.2 Feature Extraction

For the tracking of the facial expression of the subjects, the Active Appearance Model (AAM) algorithm was used (Cootes et al., 2001). The first step to determining the facial expression is to track

the facial points. AAM can track a total of 68 facial points as seen in Figure 2.

Fig. 2. 68 facial points identified by AAM

Once the facial points have been identified, the algorithm will derive 170 facial point distances as presented in Figure 3. These distances will be the features to be used in making the data sets.

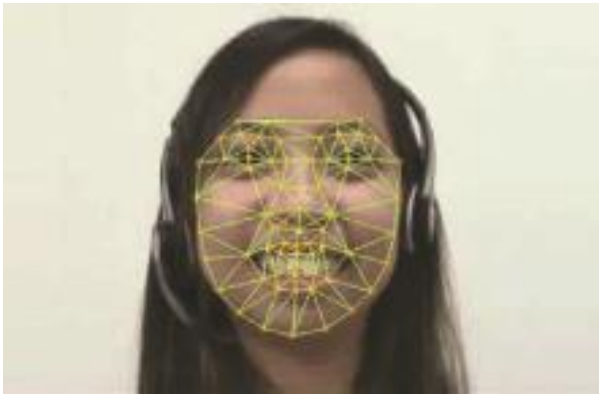

Fig. 3. 170 facial point distances computed by AAM

However, not all 170 facial point distances will be relevant to the proposed approach in this study. As stated by Buisine et al. (2006), positive emotions are predominant on the lower facial region while negative emotions are more evident on the upper facial region. Only the relevant facial point distances suggested by Luo et al. (2011) were retained.

Table 1. Relevant facial point distances markers

Upper Facial Region (44 distances)	Lower Facial Region (53 distances)
44-56	108-117
58-72	120-123
75-84	132-170
89-94	

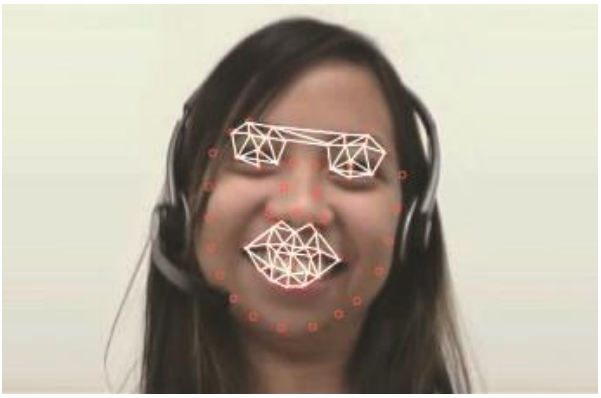

Fig. 4. Relevant facial point distances location

2.3 Model Building

Machine learning algorithms were performed to classify the laughter emotions exhibited by the facial regions. The algorithms tested were: Multilayer Perceptron and Support Vector Machine. Cross validation was done 10-fold.

3. RESULTS AND DISCUSSION

3.1 Summary of Data

Table 2. Number of instances for all data sets

Data Set	Emotion	Number of Instances	Total
1	kinikilig	678	3616
	mapanakit	633	
	nahihiya	870	
	nasasabik	326	
	natutuwa	1109	
2	kinikilig	218	2994
	mapanakit	740	
	nahihiya	441	
	nasasabik	497	
	natutuwa	1098	
3	kinikilig	433	2236
	mapanakit	378	
	nahihiya	512	
	nasasabik	294	
	natutuwa	619	
4	kinikilig	350	1617
	mapanakit	338	
	nahihiya	322	
	nasasabik	302	
	natutuwa	305	
5	kinikilig	299	1213
	mapanakit	222	
	nahihiya	267	
	nasasabik	182	
	natutuwa	243	

Table 2. Number of instances for all data sets (cont.)

Data Set	Emotion	Number of Instances	Total
6	kinikilig	167	1090
	mapanakit	227	
	nahihiya	271	
	nasasabik	198	
	natutuwa	227	

Six data sets were created, each representing a particular configuration and having varying number of instances. The first three data sets emphasized on the regions individually. The remaining three sets all combines the facial regions, but differ in labels. The purpose of these three is to examine if it is possible to derive a consolidated label by just looking at either the lower or upper facial region.

Data set number one contains 3616 instances with 170 attributes, representing the 170 facial point distances. This set is for analyzing the whole facial region.

Data set number two contains 2994 instances with 53 attributes, representing the lower facial region.

Data set number three has 2236 instances with 44 attributes which are the facial point distances found on the upper facial region.

Data set number four contains 1617 instances with 97 attributes, which are the relevant facial point distances found in both the lower and upper facial regions. The instances in this set were identified to have a single state emotion regardless of the region in question.

Data set number five contains 1213 instances with 97 attributes. The labels of the instances in this set are the same for the whole and lower facial region but differ from the upper facial region.

Data set number six has 1090 instances with 97 attributes. It is similar to the previous two sets with the difference being the whole and the upper facial regions are the ones that have the same label while they differ from the lower facial region.

3.2 Performance Test Results

Numerous metrics were looked at to determine the performance of the classifiers. These metrics are: accuracy, confusion matrix, true positive (TP) rate, false positive (FP) rate, precision and F-measure.

When the MLP algorithm was used, different numbers of hidden nodes were experimented upon. The setting that returned the best performance was the sum of the number of attributes and classes.

Table 3. Test results for Multilayer Perceptron (MLP) with *hidden nodes = attributes + classes*

	Data Set Number					
	1	2	3	4	5	6
Accuracy (%)	92.59	91.45	90.79	96.66	95.05	95.23
Kappa Statistic	0.90	0.89	0.88	0.96	0.94	0.94
Ave. TP	0.92	0.92	0.92	0.97	0.95	0.95
Ave. FP	0.02	0.02	0.02	0.01	0.02	0.02
Ave. Precision	0.92	0.92	0.91	0.97	0.95	0.95
Ave. F-Measure	0.92	0.92	0.91	0.97	0.95	0.95

Out of all the kernels tested for SVM, the Pearson VII Universal Function kernel (PUK) produced the best performance results. This kernel is known for its robustness and is considered as the generic universal kernel. PUK can actually work as any other kernel such as linear, polynomial and radial basis function. This allows model building to finish at a shorter amount of time.

Table 4. Test results for Support Vector Machine (SVM) with PUK kernel

	Data Set Number					
	1	2	3	4	5	6
Accuracy (%)	93.50	88.81	91.46	96.35	96.21	95.41
Kappa Statistic	0.92	0.85	0.89	0.95	0.95	0.94
Ave. TP	0.93	0.89	0.92	0.96	0.96	0.95
Ave. FP	0.02	0.03	0.02	0.01	0.01	0.01
Ave. Precision	0.93	0.90	0.92	0.96	0.96	0.95
Ave. F-Measure	0.93	0.89	0.92	0.96	0.96	0.95

Both MLP and SVM yielded good performance rates and are almost the same if the results are to be averaged. However, SVM surpasses MLP in terms of speed.

A generalization experiment was tried to determine if a person's overall emotion can be determined by just looking at a particular facial region. There were a total of 49 cases, 10 cases per emotion. The test was done in such a way that the emotion will be determined for the entire facial region then it will be compared to whatever emotion is identified from the upper or lower facial region only.

Table 5. Generalization test results

	Matches [correct (%)]			
	MLP		SVM	
	Upper	Lower	Upper	Lower
kinikilig	5 (50)	6 (60)	7 (70)	2 (20)
mapanakit	2 (20)	3 (30)	3 (30)	0 (0)
nahihiya	2 (20)	0 (0)	4 (40)	0 (0)
nasasabik	2 (22.22)	0 (0)	3 (33.33)	0 (0)
natutuwa	3 (30)	4 (40)	3 (30)	10 (100)
TOTAL	14 (28.57)	13 (26.53)	20 (40.82)	12 (24.49)

As seen on the table, it is not recommended to generalize one's emotion just by looking at the upper or lower facial region. Both algorithms yielded less than 50% accuracy which is not an ideal result.

4. CONCLUSIONS

Model number one is the best classifier to recognize *natutuwa* laughter since the said social signal is commonly associated with a positive emotion. One's initial perception of a person laughing, especially without context, is an expression of positive affect.

The second model recognizes *nahihiya* most while the third identifies *mapanakit*. This supports the statement of Buisine et al. (2006) that negative emotions have the tendency to be more discernable on the upper facial region. For the lower facial region, *nahihiya* is mostly noticeable because of its low intensity compared to the other laughter emotions.

Most of the classifiers have better recognition rates for the upper facial region. This may be attributed to the nature of laughter itself. The lower facial region concentrates on the mouth movement, which unfortunately, does not differ much from each laughter emotion.

Another experiment in this study showed that relying on a subregion is not recommended to determine the overall emotion of a person.

5. ACKNOWLEDGMENTS

This project is supported by the Department of Science and Technology-Engineering Research and Development for Technology (DOST-ERDT).

6. REFERENCES

Buisine, S., Abrilian, S., Niewiadomski, R., Martin, J.-C., Devillers, L., & Pelachaud, C. (2006). Perception of Blended Emotions: From Video Corpus to Expressive Agent. In J. Gratch, M. Young, R. Aylett, D. Ballin, & P. Olivier (Eds.), Intelligent Virtual Agents (Vol. 4133, p. 93-106). Springer Berlin Heidelberg.

Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001, June). Active Appearance Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(6), 681-685.

Devillers, L., Vidrascu, L., & Lamel, L. (2005). Challenges in Real-Life Emotion Annotation and Machine Learning Based Detection. Neural Networks, 18(4), 407-422.

Douglas-Cowie, E., Devillers, L., Martin, J.-C., Cowie, R., Savvidou, S., Abrilian, S., et al. (2005). Multimodal Databases of Everyday Emotion: Facing Up to Complexity. In Interspeech 2005 (pp. 813–816).

Galvan, C., Manangan, D., Sanchez, M., Wong, J., & Cu, J. (2011). Audiovisual Affect Recognition in Spontaneous Filipino Laughter. In 2011 Third International Conference on Knowledge and Systems Engineering (KSE) (p. 266-271).

Laskowski, K., & Burger, S. (2007). Analysis of the Occurrence of Laughter in Meetings. In Interspeech (p. 1258-1261).

Luo, R. C., Huang, C. Y., & Lin, P. H. (2011, July). Alignment and Tracking of Facial Features with Component-Based Active Appearance Models and Optical Flow. In 2011 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM2011) (p. 1058-1063).

Owren, M. J., & Bachorowski, J.-A. (2003). Reconsidering the Evolution of Nonlinguistic Communication: The Case of Laughter. Journal of Nonverbal Behavior, 27(3), 183-200.

Petridis, S., & Pantic, M. (2011). Audiovisual Discrimination Between Speech and Laughter: Why and When Visual Information Might Help. IEEE Transactions on Multimedia, 13(2), 216-234.

Ruch, W., & Ekman, P. (2001). The Expressive Pattern of Laughter. Emotion Qualia, And Consciousness, 426–443.

Soladie, C., Stoiber, N., & Seguier, R. (2012). A New Invariant Representation of Facial Expressions: Definition and Application to Blended Expression Recognition. In 19th IEEE International Conference on Image Processing (ICIP), 2012 (p. 2617-2620).

Suarez, M. T., Cu, J., & Maria, M. S. (2012, May). Building a Multimodal Laughter Database for Emotion Recognition. In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12). Istanbul, Turkey: European Language Resources Association (ELRA).

Truong, K. P., & Trouvain, J. (2012, May). Laughter Annotations in Conversational Speech Corpora - Possibilities and Limitations for Phonetic Analysis. In Proceedings of the 4th International Workshop on Corpora for Research on Emotion Sentiment and Social Signals (ES3 2012) (pp. 20–24). European Language Resources Association (ELRA).

Vidrascu, L., & Devillers, L. (2005). Annotation and Detection of Blended Emotions in Real Human-Human Dialogs Recorded in a Call Center. In IEEE International Conference on Multimedia and Expo, 2005 (ICME 2005).