Automatic Recognition of Affective Laughter from Body Movements

Ma. Beatrice Luz[1], McAnjelo Nocum[2], Timothy Jasper Purganan[3],
Wing San Wong[4] and Jocelynn Cu[5]
Center for Empathic Human-Computer Interaction (CEHCI), College of Computer Studies (CCS),
De La Salle University – Manila
[1] ma_beatrice_luz@dlsu.edu.ph
[2] mcanjelo_nocum@dlsu.edu.ph
[3] tj_da_skolar@dlsu.edu.ph
[4] wing_wong@dlsu.edu.ph
[5] jiji.cu@delasalle.ph

Abstract: Laughter is often associated with happiness but recent studies show that there are actually five types of Filipino laughter and these are happiness, giddiness, excitement, embarrassment, and hurtful laughter. Facial expressions and vocalization related to Filipino laughter have been the focus in many studies, however body movements with regards to laughter are still left unexplored. The main focus of this research would be determining what body movements are associated to specific types of laughter. This research will make use of the Microsoft Kinect for capturing the different body points of the participants that will result to the low-level features. High-level feature extraction will be applied in order to translate the points gathered through the Microsoft Kinect into a more understandable data. This research aims to classify these types of laughter with the use of one modality, which is the body movement. The researchers will determine which specific body parts are important to track in order to determine the type of laughter through feature selection. With the use of different machine learning techniques, this research aims to build a model that would classify which type of laughter is being performed. The machine learning algorithms used are Bayesian Networks, k-Nearest Neighbor for k=3 and k=5, and J48 decision tree. Based on preliminary results, J48 is the recurring algorithm that holds the best accuracy results, although the kappa is still low for all experiments.

Key Words: Laughter; Gestures; Filipino Laughter; k-Nearest Neighbor; J48

1. INTRODUCTION

Laughter is one of the many ways to express emotion. It is most often associated with the happy emotion. Despite the many studies about laughter, laughter is considered to be unexplored particularly in gestures. Facial expressions, vocalizations, and gestures are some of the communicative acts which can be used to express laughter. These communicative acts are used for affective recognition, and are referred to as three modalities of laughter. For the case of gestures, whole-body movements, arm movements, and a variety of functional movements can be taken into consideration for detecting and recognizing affective expressions (Karg et al., 2013).

According to Romero (2010), Filipinos are considered to be one of the happiest people in the world as the Philippines places 14th in the most joyful places in the world. It is obvious that Filipinos love to laugh and have fun in any situation, even in difficult ones. Just like how people express happiness through tears, laughter can also be used to express other emotions aside from happiness. In fact, in the study of Suarez et al. (2012), five types of Filipino Affective Laughter were identified, and these are happiness, giddiness, excitement, embarrassment, and hurtful laughter. On the other hand, Filipino laughter have been studied in terms of vocalizations and facial expressions by Alonzo et al. (2010), Galvan et al. (2011), Lee et al. (2011), Suarez et al. (2012), Cu et al. (2012),

Sta. Maria et al. (2012), and Canillas et al. (2013). However Niewiadomski et al. (2013) and Griffin et al. (2013) mentioned that full body movements associated with these types of laughter in general were still unexplored.

Body movements and changes in posture are significant to laughter as it was stressed out by Niewiadomski et al. (2013). These body movements such as head backwards movements and forced exhalations that lead to shoulders vibrations are very important in laughter detection and synthesis. Realistic body movements are significant in determining between laughter and smile visual patterns in which analysis of laughter episodes' body movements may be advantageous to laughter synthesis.

Karg et al. (2013) stated in their research that the emotion of a person can be inferred from the gestures being done. For example, anger can be expressed through gestures either by modulation of movement such as fast movement or through specific body gestures such as clenching one's fist. This proves that other types of emotions can be expressed through different gestures as well.

There are laughter databases already available that can be used for recognizing and synthesizing different types of laughter (Mackeown et al., 2013). These databases are the AVLaughter Cycle database, MANHOB Laughter database, and ILHAIRE Laughter database. However, these databases only focus on the facial expressions of seated subjects. There is also the Multimodal Multiperson Corpus of Laughter in Interaction (MMLI) which contains multimodal data on laughter that focuses on full body movements which includes the shoulders, torso and respiration.

This research aims to determine which body movements are associated to specific types of laughter. Specifically, this research aims to determine the features that can be extracted from body movements to differentiate affective laughter. Also, this research aims to determine the mathematical model that can be used to differentiate affective laughter. Lastly, this research aims to implement the system in real time.

The results of this research can be used to help computers better understand human emotions. In particular, the results of this research can be used to aid Virtual Agents in affect recognition. Understanding human emotions will help them give better responses and human interaction will be smoother. Also, part of this research is building a laughter database which can be used in future researches and also in making a system that can teach people who lack social interaction skills.

2. METHODOLOGY

2.1 Corpus Building

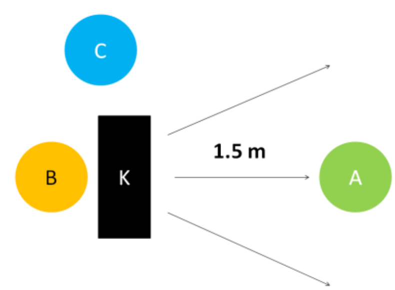

Figure 1: Illustration of the setup for the data collection (Legend: A – Person A, B – Person B, C – Sony HD video recorder, K – Microsoft Kinect)

In this research, the meaning of the word "corpus" is still the same that it is a large collection not of texts but of data that are video clips of laughter episodes. Similarly, it is more on being a database.

This study utilized two tools to perform the data collection namely, Microsoft Kinect (K) and Sony HD video recorder with a model number of HDR-XR260 that has 8.9 megapixels (C). For every collection made, there will be two people conversing, A and B with a duration of 20 minutes. Person A will stand 1.5 meters away facing the device, K while, person B will be at the back of K, facing person A. In addition, C will capture the whole body of person A while having a conversation with person B. Induction techniques that were used are in line with the ones used in the study of Ito et al. (2005) in which the subjects talked about topics of interest. The techniques used to induce laughter from person A is to talk about the following topics, best moments in college, inside jokes, likes, admiration, insults, fail moments, games, projects in school, professors, and first impression. The output of the data collection will be a Kinect data stream. Figure 1 illustrates the setup.

The Kinect data stream will be saved in a CSV file which will include the time of the recording and the points of the different body parts of person A. During the process of data collection, aside from the video camera that records the set-up of the environment where the recording would take place, there would be another video camera that captures the whole session of the conversation of the subjects to match the significant timeframes in the CSV files to the video.

After collecting the data from the Kinect Data Stream, three people will label the video clips that contains the laughter episodes. The coders will label all the video clips from all the subjects. The coders will need to pass the Baron-Cohen test (Baron Cohen and Wheelwright, 2004) before being asked to do performing the annotation.

The annotation in the previous step involves three coders but one unified result is needed. To achieve the unified result, the researchers will use an inter-coder agreement called the Fleiss' Kappa.

The data will then be segmented into parts in order to eliminate the timeframes where person A did not laugh. The only data that will be needed should only contain the points of the different body parts that contain the different laughter episodes done by person A. To conceal the identity of the subjects, their faces were pixelated through the use of Lightworks. Segmentation will be done using Windows Movie Maker. The contents of the segmented video clip will not include the conversation and silence during the data collection.

2.2. Feature Extraction

Feature extraction is divided into two parts: low-level feature extraction and high-level feature extraction. To aid in the latter phase of feature extraction, some formulas were used to detect gestures from the subjects' laughter episodes.

Low-level feature extraction refers to the analysis of determining the significant body parts or joints of the subject that will be useful for the latter methods of the model. It is done via a Kinect device and a method in C# that saves the required points in a comma separated values (CSV) file. Additionally, a timestamp is included in the CSV file, for the purpose of easier relation to the recordings. In the data collection, the points (X, Y, Z) of the different body parts of the subject that would be obtained includes head, left shoulder, right shoulder, left knee, right knee, left elbow, right elbow, spine, left hand, right hand, left foot and right foot.

In addition, the high-level feature extraction is to determine and identify what certain body parts of the subject have moved or were constantly being moved by the subject during his/her laughter episodes. The movements of the subject would be detected with the use of the low-level features through the use of other formulas also written in C#. To find out which formulas were applicable to the study, the segmented videos would have to be examined first for the most recurring gestures.

Head tilting is one of the common gestures that the subjects make during laughter episodes. Therefore, in order to detect this as a gesture, the angle of the subject's head would have to be computed. The formula that will be used is from the research of Cheung (2012). The output, which is an angle, will determine whether the subject's head has moved towards the right or the left. Equation 1 is the formula for obtaining the angle of the subject's head. In the formula, CenterSH refers to the subject's center shoulder point, while Head is the head point, and RightSH refers to the right shoulder point.

$$arccos\left(\frac{P_{CenterSH-Head}^{\ \ 2} + P_{CenterSH-RightSH}^{\ \ 2} - P_{Head-RightSH}^{\ \ 2}}{2 * P_{CenterSH-Head} * P_{CenterSH-RightSH}}\right) \quad \text{(Eq.1)}$$

As what the segmented videos suggest, people tend to lean their bodies either forward or backward during laughter. In order to detect forward or backward leaning of the body, a function from the study of Cheung (2012) will be used in this research. As mentioned in a previous section, this function will just compare the subject's shoulder depth values. Equation 2 illustrates the function for detecting the state of the subject's leaning. The two variables that will be compared are the subject's center shoulder value from two consecutive frames.

$$f(n) = \begin{cases} \textit{Leaning forward}, \text{if } NewShoulderCenter_z \leq origShoulderCenter_z \\ \textit{Leaning backward}, \text{if } NewShoulderCenter_z > origShoulderCenter_z \end{cases} \quad \text{(Eq.2)}$$

Another recurring gesture in the recordings is the tendency of the subjects to move their hands about. Their hands may be somewhere below the head, on the head, or even covering the mouth. For the system to recognize these movements, another function from the research of Cheung (2012) will be utilized. Equation 3 shows the function for determining the position of the subject's hands. The important variables for detecting this kind of gesture are the y values of the subject's hand and head points, as well as θ, which is an upper and lower threshold that is set to a value that is higher and lower than the subject's head y value, respectively.

$$f(n) = \begin{cases} \textit{Hands up}, & \text{if } Hand_y < Head_y \text{ and } Hand_y < \theta_{upper} \\ \textit{Hands on head or face}, & \text{if } \theta_{lower} \geq Hand_y \geq \theta_{upper} \\ \textit{Hands down}, & \text{if } Hand_y > Head_y \text{ and } Hand_y > \theta_{lower} \end{cases} \quad \text{(Eq.3)}$$

In this research, forward or backward movement is made when the subject propels his or her body in a forward or backward motion. Moving a foot forward will not be considered as forward movement at all, therefore either of the feet are not part of the points needed to use the function. The subject's z value for the spine was chosen as the variable to be looked at. Comparison is similar to how Equation 2 works.

$$f(n) = \begin{cases} \textit{Moving forward}, & \text{if } NewSpine_z < OldSpine_z \text{ and } NewSpine_x = OldSpine_x \\ \textit{Moving backward}, & \text{if } NewSpine_z > OldSpine_z \text{ and } NewSpine_x = OldSpine_x \end{cases} \quad \text{(Eq.4)}$$

2.3. Modelling and Training

The data collected will be divided into training data and testing data. The training data will be used for training and creating the model while the testing data will be used for validation. Modelling will be done using Weka, data mining software in Java that has a collection of algorithms for machine learning. Machine Learning algorithms to be used are the following with its definition:

- Bayesian Network - A Bayesian Network is a classification technique that uses a Direct Acyclic Graph (DAG) to encode a joint probability distribution using a set of random variables (Friedman et al., 1997).
- K-Nearest Neighbor – It learns from analogy and Training samples are described by n-dimensional attributes that is represented by a point in an n-dimensional space.
- J48 Algorithm – The java implementation of the C.45 Algorithm. It is used to create univariate decision trees (Bhargava et al., 2013)

Modeling and training will be done in an iterative manner using the different algorithms. Modelling will also undergo feature selection to improve each algorithm's performance. It will be a cycle until the algorithm with the highest performance is identified.

3. RESULTS AND DISCUSSION

The initial data set contained 15084 instances of happiness laughter, 90 instances of excitement laughter, 540 instances of giddiness laughter, 3264 instances of embarrassment laughter and 868 instances of hurtful laughter. Each instance in the data set is one frame of a recording. This extremely unbalanced number of instances resulted to models that classified all types of laughter as happiness.

To improve the model's performance, the data set is balanced into multiple data sets by getting 1000 instances for one type of laughter and comparing it to 1000 mixed instances of the other types of laughter. For cases where the number of instances is insufficient from the initial data set, some instances are randomly duplicated to achieve the 1000 instances.

In modeling embarrassment laughter vs. other types of laughter, J48 outperformed the other algorithms with the accuracy of 78.35 and a kappa statistic of 0.567. The kappa statistic of 0.8 is the acceptance criteria for most affect recognition models. This makes the kappa statistic of the embarrassment laughter vs. other types of laughter model considerably low. Feature selection is also applied to the model which identified that Head Nodding is irrelevant for identifying embarrassment laughter. By removing the Head Nodding feature, the accuracy and kappa statistic increased to 79.85% and 0.597 respectively.

The same was done for the model of excitement laughter vs. other types of laughter. J48 still outperformed the other algorithms with an accuracy of 95.90% and a kappa statistic of 0.918. The large difference between the performance of the excitement laughter and the embarrassment laughter model might be caused by the duplicated instances to cover up the shortage of instances. The possibility of over fitting is considered for this model. After applying feature selection, hand movements, nodding, and shrugging were considered to be irrelevant and were removed to increase the accuracy to 96.10% and the kappa statistic to 0.922.

Likewise, the same procedure of machine learning experiments was done on the dataset about giddy laughter. As with the first two experiments, J48 algorithm ranks first among the selection of algorithms. When compared with the previous experiments, the results for giddiness are closer to the results for embarrassment than the results for excitement. While the accuracy is higher for the results on the embarrassment dataset for the J48 algorithm, the giddiness dataset has a higher kappa in this area. Feature selection generally had no effect on this dataset.

The happiness laughter dataset underwent the same treatment as the datasets before it. Once again, J48 stood as the best algorithm for classification in this dataset. The results are also close with those of the embarrassment and giddiness dataset results. While this was the case, the J48 algorithm's kappa is closer with embarrassment than with giddiness. Feature selection only improved the results of this algorithm by a marginal amount of 0.2.

The hurtful dataset is the final dataset on which binary classification has been performed upon. Its results are close to the other results, except that of the results for excitement. Based on these results, it can be inferred that the problem for the excitement dataset may really be due to the duplication of instances, as mentioned before. Again, J48 garnered the highest accuracy on this dataset. The kappa is also greater than 0.6.

A final experiment was conducted on a dataset containing a balanced number of instances for all five types of laughter. It contained 1000 instances for each laughter type. This means that all there are more choices for classification for the algorithms. While the dataset was different, the same algorithms were chosen for this final experiment. The results were generally lower than the five previous experiments. However, J48 still had the best performance among all other algorithms. Its kappa is also above 0.6 before and after the application of feature selection. It is interesting to note, though, that feature selection diminished the results for J48.

4. CONCLUSIONS

J48 outperformed all the other algorithms in all the data sets used. However, most datasets still had a kappa statistic that is lower than 0.8 which is generally lower than most affect recognition models. The feature selection process raised the models' performance by a small amount and was able to identify that generally, head tilt angle, body movement and leaning were the most useful features for training the data. Also, the models that had binary comparison performed better than the model that had all types of laughter compared to each other.

5. ACKNOWLEDGMENTS

We would like to thank our adviser, Ms. Jocelynn Cu, for the unending support and guidance that she has given to us all throughout this project; the panelists, Dr. Merlin Suarez and Dr. Rafael Cabredo, for the feedbacks and suggestions that have helped in the improvement of this work; the data subjects for volunteering and giving their willingness to help us to complete the dataset; the whole Center for Empathic Human-Computer Interactions (CEHCI) community and the faculty and staff of College of Computer Studies, De La Salle University – Manila.

6. REFERENCES

Alonzo, J., Campita, J., Lucila, S., Miranda, M., and Suarez, M. T. (2010) Discovering emotions in Filipino Laughter. In 15th OU-DLSU Joint Academic Research Symposium, 29-30 Sept 2010, Manila, Philippines.

Baron-Cohen, S., & Wheelwright, S. (2004). The Empathy Quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. Journal of Autism and Developmental Disorders, 34, 163–175.

Canillas, R. M., Lachica, J. D., Sy, P. M., and Cu, J. (2013) Building an audio-based model for emotion recognition in Filipino laughter. In 13th Philippine Computing Science Congress, 7-8 Mar 2013, Manila, Philippines.

Corpus. (2014). In Oxford Dictionaries web. Oxford University Press.

Cu, J., Suarez, M. T., and Sta. Maria, M. (2012) Towards building a context-based laughter model. In 3rd Interdisciplinary Workshop on Laughter and other Non-Verbal Vocalisations in Speech, 26-27 Oct 2012, Dublin, Ireland.

Friedman, N., Geiger, D., and Goldszmidt, M. (1997) Bayesian network classifiers. Machine Learning, 29: 131–163, 1997.

Galvan, C., Manangan, D., Sanchez, M., Wong, J., and Cu, J. (2011) Audiovisual affect recognition in spontaneous Filipino Laughter. In 2nd Int'l Workshop on Empathic Computing in assoc. with 3rd Int'l Conf on Knowledge and Systems Eng'g, 14-17 Oct 2011, Hanoi, Vietnam.

Griffin, J. J., Aung, M. S. H., Romera-Paredes, B., McLoughlin, C., McKeown, G., Curran, W., and Bianchi-Berthouze, N. (2013) Laughter type recognition from whole body motion. In Proc of Humaine Association Conference on Affective Computing and Intelligent Interaction, IEEE Explore, 2-5 Sept 2013, Geneva, Switzerland, pp. 349-355.

Ito, A., Wang, X., Suzuki, M., and Makino, S. (2005) Smile and laughter recognition using speech processing and face recognition from conversation video. In Proceedings of the International Conference on Cyberworlds, pages 437–444, 2005.

Karg, M., Samadani, A. A., Gorbet, R., Kuhnlenz, K., Hoey, J., and Kulic, D. (2013) Body movements for affective expression: a survey of automatic recognition, and generation. IEEE Transactions on Affective Computing, preprint.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th international joint conference on artificial intelligence - volume 2 (pp. 1137-1143). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Lee, J. C., Miranda, M. C., Cu, J., and Suarez, M. T. (2011) Automatic laughter segmentation in meetings. In 11th Philippine Computing Science Congress, 4-5 Mar 2011, Naga, Philippines.

McKeown, G., Curran, W., McLoughlin, C., Griffin, H. J., and Bianchi-Berthouze, N. (2013) Laughter induction techniques suitable for generating motion capture data of laughter associated body movements. In 2nd Int'l Workshop on Emotion Representation, Analysis and Synthesis in Continuous Time and Space in conjunction with IEEE Conf on Automatic Face and Gesture Recognition.

Niewiadomski, R., Hofmann, J., Urbain, J., Platt, T., Wagner, J., Piot, B. (2013) Laugh-aware virtual agent and its impact on user amusement. In Proc. int. conf. on Autonomous agents and multi-agent systems, AAMAS.

Niewiadomski, R., Mancini, M., Baur, T., Varni, G., Griffin, H., and Aung, M. S. H. (2013) MMLI: Multimodal multiperson corpus of laughter in interaction. In Proc 4th Int'l Workshop on Human Behaviour Understanding, in conj with ACM Multimedia 2013, LNCS 8212, Springer Int'l Pub, pp 184-195, Barcelona, Spain.

Romero, G. (2010) Why the Filipino Laughs. Retrieved from http://www.ovcrd.upd.edu.ph/researchlines/2010/10/22/why-the-filipino-laughs/

Sharma, N., Bhargava, R., & Mathuria, M. (2013, June). Decision Tree Analysis on J48 Algorithm for Data Mining. International Journal of Advanced Research inComputer Science and Software Engineering, 3(6), 1114-1119. Retrieved from http://www.academia.edu/4375403/Decision_Tree_Analysis_on_J48_Algorithm_for_Data_Mining

Sta. Maria, M., Cu, J., and Suarez, M. T. (2012) Pagtawang pinoy: pagbubuo ng isang multimodal database para sa pagtawa bilang ekspresyon ng emosyon. In Ika-37 Pambansang Kumperensya sa Sikolohiyang Pilipino, 23 Nob 2012, Batangas, Philippines.

Suarez, M. T., Cu, J., and Sta. Maria, M. (2012) Building a multimodal laughter database for emotion recognition. In 8th Language Resources and Evaluation Conference, 21-27 May 2012, Istanbul, Turkey.