J Multimodal User Interfaces
DOI 10.1007/s12193-012-0114-8

ORIGINAL PAPER

A multimodal emotion corpus for Filipino and its uses

Jocelynn Cu · Katrina Ysabel Solomon · Merlin Teodosia Suarez · Madelene Sta. Maria

Received: 9 March 2012 / Accepted: 3 October 2012

Abstract This paper describes the Filipino multimodal emotion database (FilMED). FilMED was built with the purpose of developing affective systems for TALA, which is an ambient intelligent empathic space. We collected a total of 11,430 audio–video clips showing acted and spontaneous expressions of emotion involving 25 subjects. We used Filipino emotion labels to annotate the emotion, which includes: *kasiyahan* (happiness), *kalungkutan* (sadness), *galit* (anger), *takot* (fear), *gulat* (surprise), and *pandidiri* (disgust). We also engaged 20 coders to annotate the clips with valence and arousal values using Feeltrace. To show the usefulness of the database, we presented three automatic affect recognition systems that used FilMED to build the affect models.

Keywords Multimodal emotion database · Filipino · Affect recognition

1 Introduction

Affect enhances human communication. It is often expressed through facial expression, voice, and body movement (e.g., head tilt, hand gestures, posture, and gait). The ability to interpret these expressions correctly can help a person maintain good interpersonal relationships, close a business deal, or provide timely support to their peers.

For the past decade, a number of researches in affective computing had enabled machines to automatically detect and identify these affective expressions. The successes of these affective systems lie heavily on carefully designed emotion databases, for example, the HUMAINE Database developed by [9]. This database is a collection of multimodal affective expressions, occurring in both naturalistic and induced forms, annotated with contextual information, and labeled with varying levels of descriptors.

Some of the useful databases are as follows (summarized in Table 1): Zervas et al. [21] built a Greek emotional speech database to study how prosodic features can be used to improve the naturalness of voice in a text-to-speech system. Texts that are semantically neutral are read by an actress. It included words, short and long sentences, and passages. The actress was instructed to read the texts in the following emotional states: sadness, anger, fear, joy and neutral. You et al. [20] built a Chinese emotion database with materials from broadcast programs and laboratory recordings. 42 Chinese-speaking participants are asked to express the texts in the seven basic emotions (i.e., happy, sad, fear, angry, surprise, and disgust). Their collection also includes a dataset of prosodic features extracted from the audio files. Abrilian et al. [1] built the French emotion database using materials from French television news interviews. The database is composed of 51 video clips covering 24 different topics spoken by 48 subjects. Emotions of the subjects were labeled categorically and dimensionally. Contextual information such as theme, degree of implication, to whom, what for, and causes of emotion are also annotated. Burkhardt et al. [4] built a German emotional speech database with recordings taken

J. Cu (✉)· K. Y. Solomon · M. T. Suarez
Center for Empathic Human Computer Interactions,
De La Salle University, 2401 Taft Avenue, Manila, Philippines
e-mail: jiji.cu@delasalle.ph

K. Y. Solomon
e-mail: kaye.solomon@delasalle.ph

M. T. Suarez
e-mail: merlin.suarez@delasalle.ph

M. Sta. Maria
Psychology Department, De La Salle University,
2401 Taft Avenue, Manila, Philippines
e-mail: madelene.stamaria@dlsu.edu.ph

Published online: 18 October 2012

Table 1 Some multimodal emotion databases recorded in languages other than English

References	Language	Source	Labels
Abrilian et. al. [1]	French	28 Participants	Basic emotions, neutral Dimensional
Barra-Chicote et. al. [3]	Spanish	2 Actors	Basic emotions
Burkhardt et. al. [4]	German	10 Actors	Neutral, anger, fear, joy, sadness, disgust, boredom
Chitu et. al. [5]	Dutch	25 Actors	Admiration, amusement, anger, boredom, contempt, etc.
Gajsek et. al. [13]	Slovenian	19 Participants	Dimensional
You et. al. [20]	Chinese	42 Participants	Basic emotions, neutral
Zervas et. al. [21]	Greek	1 Actor	Sadness, anger, fear, joy, neutral

from ten actors. Emotional labels used include neutral, anger, fear, joy, sadness, disgust, and boredom. The database also includes electro-glottogram recordings of the speech sound for study on speech synthesis. Barra-Chicote et al. [3] built a Spanish emotion database with recordings from a male and a female actor. Emotional content of the database is focused on the six basic emotions. The database was built for the studies on multimodal emotion recognition, emotional speech synthesis, prosodic modelling among others. Chitu et al. [5] built a Dutch emotion database with recordings from 25 actors. Actors were instructed to transpose themselves into the target emotional state as they read stories that were believed to carry strong emotional load. 21 Emotional labels were used to label the recorded clips. Gajsek et al. [13] built the Slovenian emotion database with recordings from 19 participants. Each recording was segmented into shorter utterances prior to labeling. Annotators freely assign emotion labels anywhere in the clip as they see fit.

There is a Filipino speech database that was built for the purpose of Filipino speech recognition [14]. However, it has no audio–visual recordings of multimodal emotional expressions that can be used to build automatic affect recognition systems.

The need for an automatic affect recognition system stemmed from TALA [8]. TALA is an ambient intelligent space that is capable of automatically identifying its occupant, determine his/her affective states and activities, and provide appropriate empathic support by changing the ambient settings (e.g., temperature, lighting, music) in the space. Since its occupants are mainly Filipinos, we believe that the affect recognition systems should also be trained using Filipino data.

This paper describes the Filipino multimodal emotion corpus, or FilMED. Section 2 discusses the design and content of the database, which is basically composed of acted emotions (i.e., FilMED1) and spontaneous emotions (i.e, FilMED2). Section 3 presents the affective systems developed so far using FilMED. And finally, the concluding remarks are presented in the last section.

2 The Filipino multimodal emotion database

Filipino multimodal emotion database (FilMED) is composed of two parts. The first part of the database is a collection of acted emotions. We choose to collect acted emotions because we expect acted emotional expressions to be exaggerated. This type of data is useful because recording conditions can be controlled, we can have recordings focusing on specific emotions, and it would be easy to differentiate the various emotions visually and computationally. The second part of the database is a collection of spontaneous emotions. We expect spontaneous expressions to be more natural, more subtle, and varied. It will provide us with samples showing various expressions of the same emotional state in a natural setting.

2.1 FilMED1—the acted emotion database

The acted database or FilMED1 has a total of 10,500 audio–video (AV) clips of posed expressions. This is equivalent to 885 MB of data. Clips range from 0.32 to 5.64 s. Of the 10,500 clips, 7,000 of which were taken from five professional actors (3 males and 2 females) and 3,500 of which were taken from five volunteers (2 males and 3 females). All subjects are fluent in Filipino, aged 18 to 25 years old, residents in the urban area.

Recording of the AV clips were done in a quiet room whose walls are padded with acoustic boards to reduce the effects of echo and noise. The physical setup for the recording sessions is shown in Fig. 1. The subject was seated in front of a computer monitor that serves as a prompter. A webcam and a microphone were used for recording. The webcam was placed on top of the monitor to capture the full frontal view of the face. Control texts, chosen with the help of a linguist, are flashed on the prompter. There are four types of texts, which include 2-syllabic words, 5-syllabic words, phrases, and sentences (Table 2). There are 50 examples for each type of texts. These texts ensure that all possible combinations

Fig. 1 Physical setup of the recording session

Table 2 Examples of control texts used in FilMED1

Type	Examples
2-Syllabic words	*Lambat* (fishing net) *Tamad* (lazy)
5-Syllabic words	*Pamahalaan* (government) *Tagapagtago* (keeper)
Phrases	*Sa may talipapa* (in the market) *Itlog na maalat* (salted egg)
Short sentences	*Masustansya ang gulay*. (Vegetables are nutricious.) *Araw-araw akong nagdarasal*. (I pray every day.)

and variations in the Filipino language are captured. Examples are shown in Fig. 2.

Each actor was asked to transpose themselves into the six basic emotional states identified by Ekman. Prior to each recording session, the actors were briefed as to the type of emotion we want them to express or act out. They were given ample time to internalize the emotion and prepare for the recording. And when they are ready, the actors will read the texts in a particular emotional fashion.

Filipino labels were used to categorize the AV clips. These are *kasiyahan* (happiness), *kalungkutan* (sadness), *galit* (anger), *takot* (fear), *gulat* (surprise), and *pandidiri* (disgust). The neutral expression was included to serve as the baseline or reference emotion. The reason for using Filipino labels is to ensure that there are no misinterpretations on the specific type of emotions to be posed by the actors. A psychologist was asked to manually evaluate each clip and determine if the emotions are identifiable both from the face and voice of the subject. If the emotion was not evident on either the face or the voice, the clip was marked for re-recording.

2.2 FilMED2—the spontaneous emotion database

The spontaneous database or FilMED2 is composed of clips taken from television series *Pinoy Big Brother* (PBB) *Season One* and the *Philippine's Scariest*

Fig. 2 Samples of FilMED1 images showing posed emotions happy, sad, angry, fear, surprise, and disgust (from *top* to *bottom*, *left* to *right*)

Challenge (PSC). The PBB series follows the Philippine version of the reality television show Big Brother. It follows the same premise as its foreign counterparts, where 12 residents are forced to live with each other inside a house for 100 days. The house was built with 26 surveillance cameras discreetly positioned in various parts of the house such as the living room, kitchen, bedroom, and bathroom to monitor the occupants' activities. The choice of using the first season of this type of program are mainly due to its "reality" flavour and its "non-actor" participants, in which we believe authentic emotions are more likely displayed. To get high quality of the recording of the show, we decided to manually extract emotion episodes from the DVD version of the series. The segmented emotion clips show the participants conversing and interacting with another participant, and sometimes being interviewed by the show's host. The PSC follows three contestants as they complete a task in a scary place. Each contestant enters the house or room individually. As they entered the room, they were to read aloud the texts narrating a scary event that was supposed to have happened in the room. This is to set the mind of the contestant as they complete a task within the room and to heighten the scary atmosphere of the place. Each contestant carries a camera mounted on their helmet capturing their facial expressions and a two-way radio capturing their conversations with the host of the show. We choose to collect clips from this program because we believe that this program could provide us with clips showing fear and surprise. Only those clips showing the full frontal face of the subject (i.e., left/right eyebrow, left/right eyes, nose and mouth are visible) and the subject speaking in Filipino were extracted.

The AV clips collected ranged from 1 to 32 s. A total of 930 clips, a combination of various multimedia formats like audio, video, and images totalling 1.07 GB, were collected and annotated by coders. For the labeling task, we invited 20 volunteers to participate. To dispel cultural differences in interpreting the expressed emotion, we ensure that the

Table 3 Content summary of FilMED2 totalling 930 clips in various multimedia format

Label/media	Audio	Video	Images
Happy	11	82	40
Sad	12	32	40
Angry	11	17	11
Fear	10	10	12
Neutral	10	48	12
Mixed or blended	158	283	131
Total	212	472	246

Fig. 3 Samples of FilMED2 images showing spontaneous emotions happy, sad, angry, fear, and neutral (from *left* to *right*)

volunteers who will do the annotation have approximately the same background as the participants in the television series, i.e., they are fluent in Filipino, aged 18–25 years old, and urban residents.

For the categorical labeling task, we expected that several labels might be assigned to the same clip because some coders have watched PBB and PSC before and they have additional contextual background for the particular expression. We used Fleiss' Kappa [12] measurement of $K = 0.6$ to consolidate the labels from all 20 coders. This value for K was selected because according to [11], the psychiatric community considers K values that are greater than 0.5 as acceptable. Table 3 shows the breakdown of the files categorized according to affect label and media type. Based on this table, 61.51 % of the clips were classified as containing mixed or blended affect, 14.30 % were happy, 9.03 % were sad, 4.19 % were angry, 3.44 % were fear, and 7.53 % were neutral. By mixed or blended affect, we mean that the clips were labeled with at least two emotion labels that coders think are relevant. Examples of FilMED2 clips are shown in Fig. 3.

For the dimensional labeling task, Feeltrace [7] was used to annotate the clips. With Feeltrace, affect is described in two-dimensions, i.e., valence (how positive or negative the affect is) and arousal (how excited or apathetic the affect is). To arrive at a common label for each clip, we computed the sign-agreement and correlation scores of the coders [16]. The results of dimensional labeling are shown in Fig. 4. Clips that were categorically labeled as happy are mainly found in the first quadrant (positive valence, high arousal); clips that were labeled angry and fear are mostly found in the second quadrant (negative valence, high arousal); and clips that were sad are found in the third quadrant (negative valence, low arousal). Neutral is scattered across the third and fourth quadrants (positive valence, low arousal). The centroids of each emotion clusters are also computed and plotted on the valence–arousal coordinates, shown in Fig. 5.

3 Observations and evaluation of FilMED

FilMED1 and FilMED2 were evaluated separately given the nature of the collected data.

3.1 FilMED1

Filipinos have almost the same expression for neutrality and sadness, which is characterized by the same drop in pitch and volume, and facial muscles relaxed. When asked to express a neutral emotion, they tend to sound sad and look sad. Most acted expressions of anger are consistently characterized by short burst of high energy at the onset of speech, that is, the actors usually shout to express anger. This is usually combined with eyes wide open and pointedly looking at the object of anger. There are very few expressions of anger where the voice is low, controlled and hard. This may suggest that when acted, the actors need to exaggerate the emotion to ensure that it is being recognized and labeled as anger. Disgust is difficult for Filipinos to express especially if the English word label is used. For Filipinos, disgust has a range of emotions usually associated with negative emotions. It may be interpreted as *suklam*, *yamot*, *suya*, *sama ng loob*, *rimarim*, and *pandidiri* among others. Not all these words have direct translation in English and all are scattered across the entire evaluation-activation space [18]. This indicates that a distinction needs to be made between lexical and linguistic equivalence [2]. However, translation equivalents do not guarantee equivalence in meanings. Happiness, surprise and fear are all characterized by a wide pitch range and higher pitch, but they differ in terms of duration of utterances. It should be noted that for the non-professional actors, they tend to mute their expressions, both on the face and the voice, especially when they are conscious of being watched. These may indicate that Filipino emotional expressions are highly influenced by cultural display rules, that positive emotions are encouraged and negative emotions are suppressed.

3.2 FilMED2

90 % of the clips in FilMED2 came from PBB. PBB creates a social environment where a Filipino is forced to live with a group of strangers and is forced to interact with each member. We believe that the clips we captured reflect authentic and spontaneous emotional expression. In FilMED2, 61.5 % of the clips were labeled with mixed or blended emotion. This is related to the observations we made in FilMED1, where

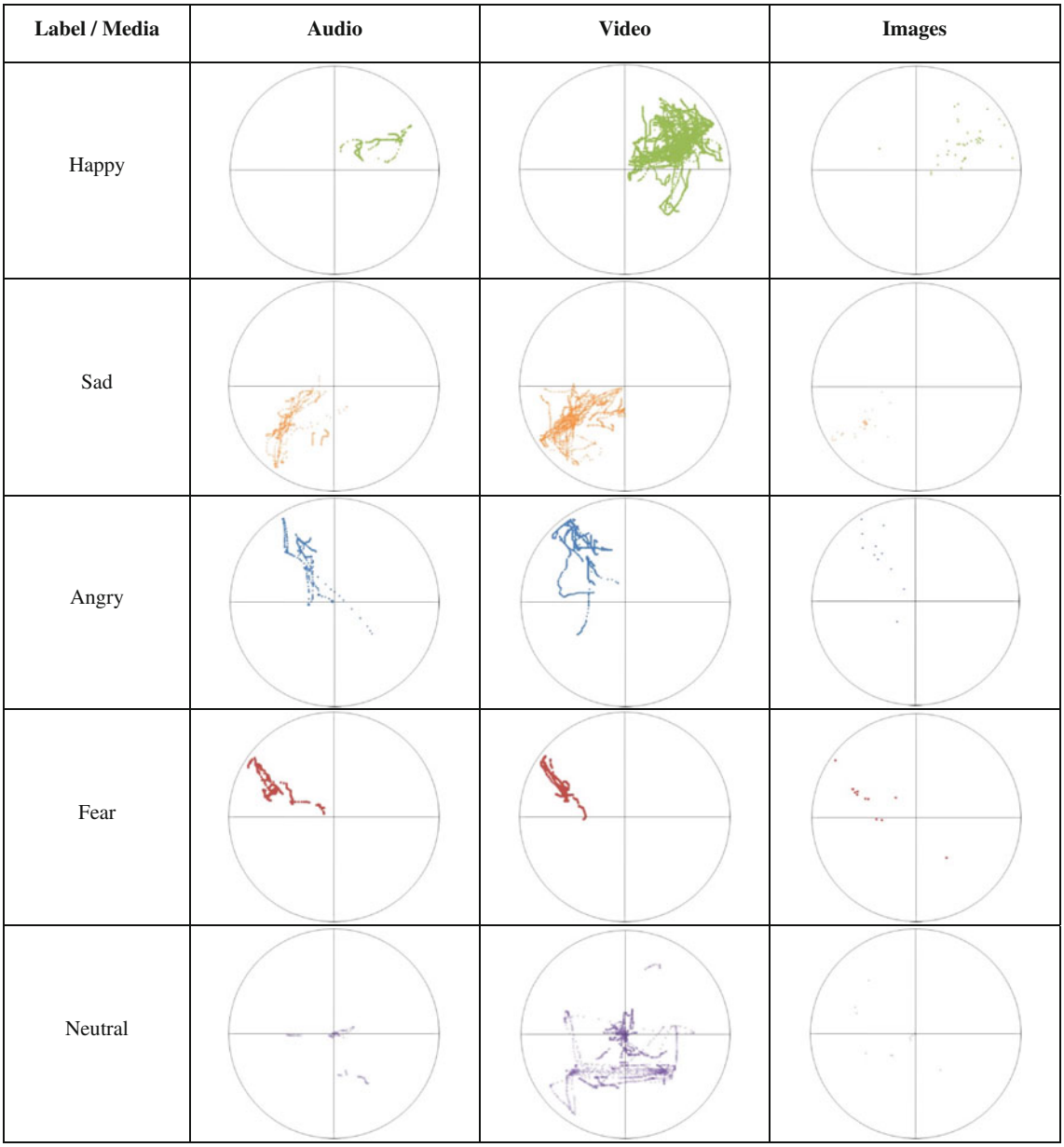

Fig. 4 FilMED2 labeled dimensionally using FEELTRACE. The *x*-axis indicates valence, the *y*-axis indicates arousal

actors tend to mute their emotional expression. These are evidence that Filipino emotional display are highly influenced by culture. For example, the Filipino trait of *pakikisama* (i.e., getting along well with others) and *pakikiramdam* (i.e., sensitivity to feelings) are just two of the important traits deeply ingrained in the Filipino psyche, to discourage criticism from others in the group and to maintain good relationship with the group. These may be why a typical Filipino would automatically mute his or her emotional expression to ensure conformity with the group. The next larger set of clips is that of happiness. It is the most amplified emotion across all cultures, and most especially for members of collectivistic culture like the Filipino. In a collectivistic culture, there is more expression of positive emotions among in-group members than are members of individualistic cultures [15]. This may be the case among participants of PBB who experienced prolonged interactions with the same group within a household. The occurrence of smiles and laughter in the clips, especially those labeled as mixed or blended emotion, further suggests that a positive emotional expression is commonly used as a tool to diffuse a highly charged emotional situation. We also observed that the kappa coefficient among annotators is highest if the clips presented to the coders contain more contextual information, e.g., they know the topic being discussed

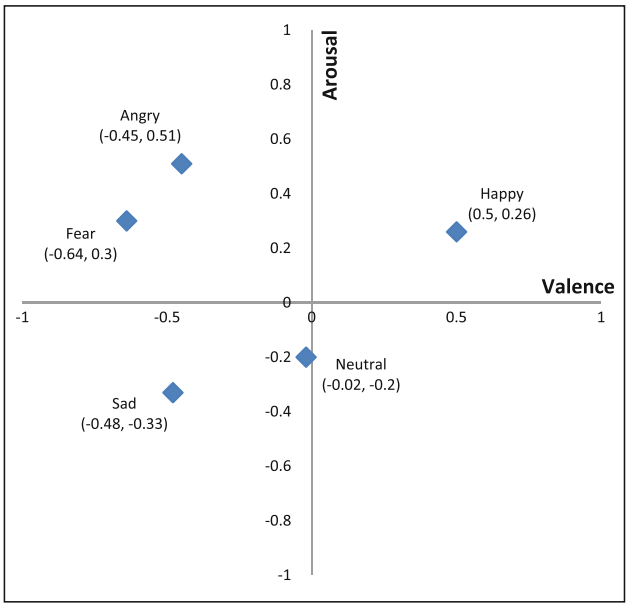

Fig. 5 Cluster centroid of the emotions happy (0.5, 0.26), sad (−0.48, −0.33), angry (−0.45, 0.51), fear (−0.64, 0.3), and neutral (−0.02, −0.2)

by the subject. Interviews with the annotators revealed that they rely more on the combined information given by the face, voice and body movement, rather than on one modality only.

4 Applications of FilMED

FilMED was built to have useful data on which we can build affect recognition systems for TALA. These affect recognition systems were Emoticon [6], SMERFS [10] and SAM-D2 [17]. Table 4 shows a summary of these studies.

4.1 Emoticon [6]

Emoticon is an automatic affect detection system that detects the emotional state of a person by analysing his/her voice. Emoticon used 7,000 audio clips from FilMED1 (i.e., 1,000 clips each for the basic emotions of happiness, anger, fear, surprise, sadness, disgust and neutrality). This study focused mainly on processing the audio clips. The approach involves pre-processing the audio clips to reduce noise, using signal processing techniques to extract speech features, and using machine learning techniques to build models for classification of emotion. Features such as minimum pitch, maximum pitch, mean pitch, minimum energy, maximum energy, mean energy, duration, and formants F1, F2, and F3 were extracted using PRAAT[1] (example shown in Fig. 6). These features were used to form a feature vector. Each feature vector in the dataset represents one audio clip. The affect model is built using the K-Nearest Neighbour, where k = 3, machine learning approach.

In the first attempt, all 7,000 audio clips were used to build the affect database. Based on the tenfold cross validation results, Emoticon achieved an average emotion recognition rate of 40.12 %, this means that it can correctly identify the emotion at only less than 50 % of the time. For example, out of the 1,000 audio clips labeled as happiness, only 441 are correctly identified by the model as happiness. This result indicates that the model was not useful for automatic affect recognition. In the second attempt, 100 clips showing consistent expression of each type of emotion were carefully selected and used to build the model. Again using tenfold cross validation test, the model achieved an average emotion recognition accuracy of 77.29 %. Among the emotions tested, happiness is the most correctly recognized emotion at 95 %, i.e., 95 out of 100 clips are correctly recognized. The least or the most difficult to recognize emotion is fear at 50 % and disgust at 64 %. Comparing gender-based expression of emotion, the study found that the emotions sad, fear and disgust are better recognized from female speakers while anger is better recognized from male speakers. In short, male speakers are better at expressing strong emotions that require lots of energy, while female speakers are better at expressing emotions that require less energy. This result can be related to the Filipino culture where men are expected to be tough (thus better at expressing anger) and women are more timid (thus better at expressing sadness). But this is not conclusive and may require more meticulous study. The results also indicate that 60 % of the time, the emotions happy and sad are correctly recognized in short utterances. In long utterances (i.e., phrases and sentences), emotions anger, sadness and fear are have very low recognition.

4.2 SMERFS [10]

SMERFS performs automatic affect detection by analysing face features and voice features. It used 488 spontaneous audio–video clips from FilMED2. Facial features are computed from 16 facial points around the eyebrows, eyes, and mouth regions. Prosodic features are pitch, pitch contour, and energy. The emotions considered in this study are happy, sad, angry, fear and neutral.

A total of 488 clips were used to train a multi-class support vector machine (SVM) to recognize the affect. It achieved an accuracy of 80 %. This figure was achieved because 75 % of the weight in decision making was given to facial features.

The study also tried to identify a small feature set that can be used to determine affect. For the voice, the useful features include pitch, pitch contour and energy. For the face, the useful features are the 12 distance features derived from

[1] PRAAT can be downloaded from http://www.fon.hum.uva.nl/praat/.

Table 4 Affect recognition systems that used FilMED to generate their affect models

	Emoticon	SMERFS	SAM-D2
Database	FilMED1	FilMED2	FilMED2
Size of dataset	7,000 audio clips	488 AV clips	900 AV clips
Label	Categorical	Categorical	Dimensional
Modality	Voice	Face and voice	Face and voice
Voice features	Pitch, energy, duration, formants	Pitch, pitch contour, energy	Pitch, energy, formants f1–f3, MFCCs
Face features	None	16 facial points on eyebrows, eyes, mouth	68 facial points on the entire face
Classifiers	K-nearest neighbour (K = 3)	Multi-class SVM	SVM with regression
Evaluation	Recognition accuracy: 77.29 %	Recognition accuracy: 80 %	Valence RMSE is 0.4638; arousal RMSE is 0.2925

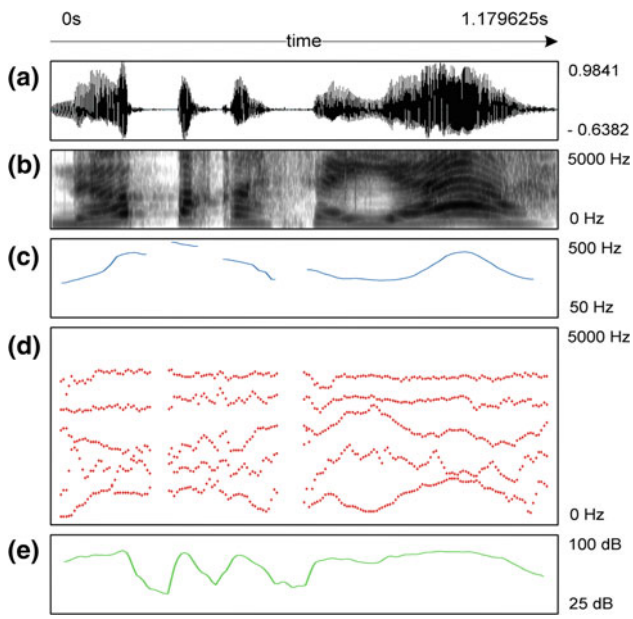

Fig. 6 Features extracted from the audio signal using PRAAT. **a** Recording of the word *napakasaya*, **b** spectrogram, **c** pitch contour, **d** formants f1–f5, and **e** intensity. The duration of the word is 1.179625 s

the facial feature points are useful. Fusing the modalities increases the recognition accuracy as compared to relying on only one modality.

4.3 SAM-D2 [17]

SAM-D2 also performs multimodal affect recognition using face and voice features. However, unlike SMERFS, SAM-D2 used the valence–arousal labels of the clips. 900 Audio–video clips showing spontaneous emotions were used in this study. Facial features were composed of distances computed from 68 facial points covering the entire face. These facial points were extracted using the Active Appearance Model (AAM). Voice features include pitch, intensity, formants F1, F2 and F3, energy, and 12 Mel-Frequency Cepstral Coefficients (MFCC). Two regression techniques, the Multilayer Perceptron (MLP) and the Support Vector Machine with Regression (SVR), were used to build the affect model.

The performance of the system was compared based on the value of Root Mean Square Error (RMSE). Based on results, the SVR outperforms MLP by achieving prediction values of 0.4638 for valence and 0.2925 for arousal.

4.4 Other studies using FilMED

Some of the on-going studies that are currently using FilMED include:

- A real-time automatic affect recognition system. To achieve this, the automatic affect recognition system should be able to handle streaming data, extract an optimal feature set, and classify the emotion with minimal delay. This study explores the use of Motion History Imaging (MHI) to keep track of changes in the facial expression.
- Analysis of mixed or blended emotions. More than 50 % of the clips in FilMED2 contain mixed or blended emotion. This study aims to answer the following questions: which emotions usually co-occur; which emotions are usually masked; which modality or feature set is more reliable source of information on the masked emotion; how emotions transition from one to another; and, in what contexts do mixed emotions occur.

5 Concluding remarks

This paper presented the FilMED and how it was used to build automatic affect recognizers. Currently, applications that we can develop using FilMED are still limited.

First, for the face analysis, we are limited by the frontal shots of the subject. As long as the subject does not turn his/her face to the left or to the right by more than 30°, we

can still extract the required facial points to correctly identify the affect. The database does not include clips where the face of the subject was occluded, either wholly or partially, by objects or shadows.

Second, aside from the face, voice and shoulder movements, we cannot do other forms of body movement analysis like hand gestures, body postures, or gait. These modalities, when combined with face and voice, are useful sources of a person's affective state and helpful in the study of mixed or blended emotions.

Third, the clips in the database do not contain enough samples where facial and body movements are used to convey more than just emotions. According to the studies of [2] and [15], cultural norms and social restraints influence how a person experience and express emotions. We believe this to be true because in the Philippines, for example, facial expressions that include pursed lips may not indicate any emotion at all. Pursed lips may be used to point to a location or to ask someone to perform a task, depending on the context [19]. Combining these lip movements with eye brow movements may indicate absolute agreement, encouragement, weak disapproval, and many others. It is important for us to be able to capture these nuances in our models to reduce the chances of wrong interpretations.

Annotation of FilMED is not yet complete. There is still a need to identify the types of contextual information that are available in the video clips and determine the best way to annotate these information. As the collection of FilMED grows, so does the number of studies and applications we can do with it.

Acknowledgments This project is funded by the Department of Science and Technology—Philippine Council for Industry, Energy and Emerging Technology Research and Development (DOST-PCIEERD) under the program titled "*Towards the Development of a Self-Improving and Ambient Intelligent Empathic Space: Data-centric, Multimodal Empathic Modelling from a Pluridisciplinary Perspective*" and the De La Salle University – University Research Coordination Office (URCO) under Project No. 24 F U/S 2 08.

References

1. Abrilian S, Devillers L, Buisine S, Martin JC (2005) EmoTV1: annotation of real-life emotions for the specification of multimodal affective interfaces. In: Proceedings of the 11th international conference on HCI, Las Vegas, USA
2. Altarriba J, Basnight DM, Canary TM (2003) Online readings in psychology and culture. In: Lonner WJ, Dinnel DL, Hayes SA, Sattler DN (eds) Emotion representation and perception across cultures. Center for Cross-Cultural Research, Western Washington University, Washington DC
3. Barra-Chicote R, Montero JM, Macias-Guarasa J, Lufti SL, Lucas JM, Fernandez-Martinez F, Dharo LF, San-Segundo R, Ferreiros J, Cordoba R, Pardo M (2008) Spanish expressive voices: corpus for emotion research in Spanish. In: LREC, 2008
4. Burkhardt F, Paeschke A, Rolfes M, Sendlmeier W, Weiss B (2005) A database of German emotional speech. In: Interspeech 2005, Lisbon, Portugal
5. Chitu A, Van Vulpen M, Takapoui P, Rothkrantz JM (2008) Building a Dutch multimodal corpus for emotion recognition. Workshop on corpora for research on emotion and affect, vol 6, pp 53–56
6. Chua J, De Guia O, Li C, Rojas JF (2009) Emoticon: emotion modelling using vocal features. In: 1st DLSU-OU Workshop in empathic computing, DLSU, Philippines
7. Cowie R, Douglas-Cowie E, Savvidou S, McMahon E, Sawey S, Schroder M (2000) Feeltrace: an instrument for recording perceived emotion in real time. ISCA workshop on speech and emotion, pp 19–24
8. Cu J, Cabredo R, Cu G, Legaspi R, Inventado PS, Trogo R, Suarez MT (2010) The TALA empathic space: integrating affect and activity recognition into a smart space. In: Proceedings of the 3rd International Conference on HumanCom, Cebu, Philippines
9. Douglas-Cowie E, Cowie R, Sneddon I, Cox C, Lowry O, McRoriw M, Martin JC, Devillers L, Abrilian S, Batliner A, Amir N, Karpouzos K (2007) The HUMAINE database: addressing the collection and annotation of naturalistic and induced emotional data. In: Paiva A, Prada R, Picard RW (eds) ACII 2007. LNCS 4738:488–500
10. Dy MLI, Espinosa IV, Go PP, Mendez CM, Cu J (2010) Multimodal emotion recognition using a spontaneous Filipino emotion database. In: 1st International workshop on empathic computing, Cebu, Philippines
11. Eugenio BD (2000) On the usage of kappa to evaluate agreement on coding tasks. In: Proceedings of 2nd International conference on language resource and evaluation, Chicago
12. Fleiss JL (1971) Measuring nominal scale agreement among many raters. Psychol Bull 76(5):378–382
13. Gajsek R, Struc V, Mihelic F (2009) Multimodal emotional database: AvID. Informatica 33:101–106
14. Guevara R, Co M, Espina E, Garcia I, Tan E, Ensomo R, Sagum R (2002) Development of a Filipino speech corpus. In: 3rd National ECE conference, UP Diliman, Philippines
15. Matsumoto D, Yoo SH, Hirayam S, Petrova G (2005) Development and validation of a measure of display rule knowledge: the display rule assessment inventory. Emotion 5(1):23–40
16. Nicolaou MA, Gunes H, Pantic M (2010) Automatic segmentation of spontaneous data using dimensional labels from multiple coders. In: Proceedings of the international workshop on multimodal corpora: advances in capturing, coding, and analyzing multimodality, in assoc with 7th LREC, Malta, pp 43–48
17. Solomon KS, Latorre AA, Tensuan JP, Cu J, Suarez MT (2011) Modeling spontaneous effect using continuous data. In Phil. Comput J 6(2):14–20
18. Sta. Maria M, Magno C (2007) Dimensions of Filipino negative social emotions. In: Proceedings of the 7th conference of Asian association of social psychology
19. Washington BD (2007) Understanding nonverbal communication of Filipinos: a traditional form of literacy. California State University East Bay Online Journal, CFSJ-CSUEB-2007
20. You M, Chen C, Bu J (2005) CHAD: a Chinese affective database. In: LNCS, vol 3784. Springer, Berlin, pp 542–549
21. Zervas P, Fakotakis N, Geourgo I, Kokkinakis G (2003) Greek emotional database: construction and linguistic analysis. In: Proceedings of the 6th international conference of Greek linguistics, Rethymno