OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

Facial Expression of Emotion Assessment from Video in Adult Second Language Learners

*Zvetomira Svetleff [1], Suzan Anwar [2, 3]
*Department of Educational Psychology& Higher Education, University of Nevada Las Vegas
** Department of Computer Science, University of Arkansas at Little Rock
*** Department of Computer Science, Salahaddin University
Corresponding author: *Zvetomira Svetleff

ABSTRACT: *The study looks at how second language (L2) learners make use of their semiotic resources (e.g., native and target languages) while engaging in an unfamiliar task of storytelling and attempting to maintain self-regulation within two culturally and linguistically different contexts. A system for recognizing emotions through facial expressions from video is presented in this paper. This technique is applied for the automated identification of the psychological state that exhibits a very strong correlation with the detected features. Results indicate that that second language speakers expressed more emotions both positive and negative while narrating a story in the target language, and maintained a "neutral" facial expression while narrating the same story in the native language.*
Keywords: *Affect and Emotions; Embodiment; Second Language Acquisition; Bilingualism, Face Expression; Face Detection; Active shape Model Classification; SVM*

I. INTRODUCTION

During social communication, we simultaneously use speech and gesture. Nonverbal communication involves conveying messages through body movements, head nods, hand-arm gestures, facial expressions, and eye gaze, etc. These visual cues are helpful in communication [1] and facial expressions can play a significant role in social communication conveying emotions such as happiness, saddens, anger, surprise and/or disgust. By reading a person's face we may be able to observe a person's internal emotional state. The face we look at is a mix of both physical characteristics and emotive expressions. While through speech, we verbalize and we can hear the words, it is through gesture and facial expressions that we communicate important non-linguistic cues that help clarify the verbal message and enrich the conversation. According to Vygotsky's Sociocultural Theory (SCT) [2] human learning is largely a social process. SCT theory provides a framework for studying human cognition while emphasizing on the interaction between developing people and the culture in which they live. Because different communities share different values, understanding, history and practices, individuals from different cultures have not only different table manners, child-rearing practices, and schooling, but also they have different ways of emotional expression [2], often decoded in non-verbal cues such as facial expressions. Cross-cultural differences in how people express emotions are significant in L2 conversation analysis. People interact in face-to-face situations through nonverbal forms of communication such as facial expressions, gaze, gestures, and proxemics and thus, making meaning is not limited to verbal mediation alone [3]. Previous research demonstrates that the use of gesture helps clarify the message, as in Riseborough's story-retelling task [4] where a story was told to participants in four conditions. All four conditions included audio, however they each varied in visual cues: a) no visual cues; b) a speaker with no movement; c) a speaker with vague body movement; d) and a speaker with gestures. More information from the story was recalled by the group that saw the gestures of the speaker. L2 gesture research also has reported that learners use more gestures when producing L2 English than their L1s [4]. Based on our review of literature on real time extractions of emotion from video, no current research explains storytelling from a sociocultural perspective, while analyzing the emotion through the subject's facial expressions. In this paper, we propose our preliminary studies for designing a system that aims at identifying the psychological state, affect and emotion of individuals making use of their semiotic resources (e.g., native and target languages) while engaging in unfamiliar tasks (e.g. storytelling and answering a questionnaire in collaborative activity –RECOLA [5]) and attempting to maintain self-regulation within two culturally and linguistically different contexts. In particular, we focused on the identification of self-regulation and anxiety that may be of principal interest for second language acquisition assessment. The system is composed by custom acquisition software, recording both subject face and the video screen, an automatic and real-time module that considers few important facial regions and performs face detection, feature extraction, and facial expression recognition, and a decision module that correlates the extracted data with psychological features assessed by traditional means. In this study, the software "FaceEmo"[6] is used for the analysis of the facial expressions of the subjects represented as recorded in the videos. Details about the FaceEmo software are explained in section 3. In this paper, we apply the proposed technique for extraction of emotion from facial expression to answer the following questions: Does the development of meaning, affect and self-regulation associated with L2 words processing depend on the learner's level of language proficiency? When Suzan Anwar 2 Department of Computer Science University of Arkansas at Little Rock sxanwar@ualr.edu L2 learners start to process L2 words affectively and semantically? Several studies investigate the extent to which emotionality in L2 is reduced [7]. When L2 is learned post puberty or even after early childhood the two languages of an individual may differ in their emotional impact, with the native, first language (L1) being the language in which personal involvement is expressed, and the second language (L2) being the language of distance and detachment [8]. The learning process is essentially different under different interactive conditions, and the processing of L2 words and L1 words must be different in terms in affect and emotion. Understanding non-verbal communication behaviors is paramount in understanding emotions in social interactions with individuals who speak more than one language. The paper is organized as follows: Sections 2 describes the components of the proposed system. Section 3 presents the FaceEmo Emotion Classification Software. Section 4, includes results from two studies. Section 5 gives the conclusion and future possible research.

II. PROPOSED METHOD

The system aims to detect a set of emotions (surprise, anger, happiness, sadness, fear, and disgust) in subjects that speak more than one language. In Study 1, subjects from different cultural backgrounds (Hispanic and Eastern European) are asked to construct a narrative based on a story in both their native and target language. The story is portrayed through a series of six sequential drawings. During this task, both the video screen and the subject face are recorded with a video camera. The user is not aware of the video recording before the experiment, to maintain a natural facial mimic. We focused on the identification of how self-regulation is of interest in non-verbal communication. To investigate the difference between how L2 learners process L2 words affectively and semantically, we compared the L2 words used to the L1 words used in both narratives of the short story. To study the affective state of the L2 learners, the emotion recognition software tracked face emotions as the L2 learner narrated the story in both English and their native language, documenting each facial expression of the speakers. In Study 2, subjects from different cultural backgrounds (French, German and Italian) were asked to participate in a collaborative activity. The experiments are done using 3 videos. The database consists of 9.5 hours of audio, visual, and physiological recordings of online dyadic interactions between 46 French speaking participants, who were solving a task in collaboration RECOLA [5] The system is composed by custom acquisition software, recording both subject face and the video screen, an automatic and real-time module that considers few important facial regions and performs face detection, feature extraction, and facial expression recognition, and a decision module that correlates the extracted data with psychological features assessed by traditional means.

III. EMOTION CLASSIFICATION SOFTWARE AND RESULTS

The proposed system pursues a different track and has three main steps as shown in Fig.1 which represents the block diagram for the proposed system, the first step is the detection of facial triangulation points using a tracker based on multiresolution active shape model. In the second step, local changes in specific regions of the face (forehead wrinkles, eye brow wrinkles, distance eyes to eyebrows, wrinkles in cheeks, vertical and horizontal measures of the mouth) are calculated with the help of point location tracking. Finally, a face expression is identified according to distance of the obtained attribute vector to that of in the learning set.

1. Face Detection and Feature Tracking

Active Shape Model (ASM) introduced by Cootes et al is one of the most prevalent techniques for detection and tracking of triangulation point [9]. To identify the triangulation points in an image, first the location of face is detected with an overall face detector (such as Viola-Hones). The average face shape which is aligned according to position of the face constitutes the starting point of the search. Then the steps described below are repeated until the shape converges. 1. For each point, best matching position with the template is identified by using the gradient of image texture in the proximity of that point. 2. The identified points are projected from their point locations in training set to the shape eigenvalues which is obtained by Principal Component Analysis (PCA). ASM tracking is developed by Wei [10]. It is stated that ASM gives better results when it is trained with model specific to the person [11]. Fig. 1. The Proposed Method's Algorithm

2. Post Processing

The ASM-based tracker tracks 116 triangulation points indicated on left side of Fig. 2. The tracker works sturdy on eyebrows, eyes, chin and nose points; however, since it cannot correctly track flexible lip points, it is not reasonable to directly use the location of this points as attributes. There are two reasons for this phenomenon, first reason is ASM's holistic modeling of all triangulation point's locations, and second reason is losing small changes in location of lip points to constraints made on shape with PCA. Further, the difference in intensity at lip edges is not as significant as other face components. Therefore, instead of directly using the locations of triangulation points being tracked, attributes depicted in the right side of Fig. 2 and in the list, below is obtained by (a) (b) Figure. 2. Face Landmarks (a), Areas Of Interest Used In Attribute Derivation(b) identification of specific length and regions in face using these points. 1- Distance of the mid-point of eye-gap to eyebrow mid points. 2- Mouth width 3- Mouth height 4- Domain of vertical edge in the foreface 5- Domain of horizontal edge in the mid-forehead 6- Sum of vertical and horizontal edge domains in right cheek 7- Sum of vertical and horizontal edge domains in left cheek First three attributes are obtained by Mahalanobis distance of the corresponding triangulation points to each other. For the other attributes, the image is first smoothened by filtering with Gauss core, then by filtering with Sorbel vertical and horizontal cores separately, edge domains are calculated. Next, absolute value corresponding to each region is calculated. Horizontal and vertical edge domains in facial expression are shown in Fig. 3 (a) (b) Figure. 3. Horizontal and Vertical Domain in Expression (a) and Corresponding Vertical Edge Domain (b). For instance, the average vertical edge domain value corresponding to blue box in the forehead constitutes the fourth attribute.

$$D_M(\underline{x}) = \sqrt{(\underline{x}-\underline{\mu})^{TS-1}(\underline{x}-\underline{\mu})}$$

For motivation on selecting attributes, movement descriptions corresponding to emotional expression given in Table 1 can be examined. These clues are based on leading research done by Ekman and Friesen [12].

3. Emotion Classification

In the first step of the video recording, subjects are asked to assume a neutral expression and to wait for 2 seconds. In each following frame, attribute vector for the face is calculated. The average distance value of each element in the attribute vector to each class is calculated which is di , i=1,…,N vectors. The distance values depict differences, whereas Si = e -d i values depict similarities and used as similarity metrics. Finally, Si vector is normalized as the sum of their elements will equal to 1. These values can be used as a probability for each class. Because of the superior classification performance and its ability to deal with high dimensional input data, Support Vector Machine (SVM) is the choice of the classifier in this study for facial expression recognition.

IV. Results

4.1. Study 1

The purpose of Study 1 was to investigate how L2 learners attempt to gain self-regulation in a communicative task in both their L1 and L2, while examining differences in the affective states of the speakers as tracked by the emotion recognition software. Subjects in this study were asked to construct a narrative based on a story in both their native and target language. Previous research suggests that that processes of self-regulation should differ cross-culturally with respect to L2 learning [13]. In the analysis of L1 and L2 differences in the verbal message we found that both the more proficient and less proficient L2 learners used object-regulation and framed the discourse using various extratextual elements, such as providing background for the setting or labeling characters. The less proficient L2 learners engaged more frequently in other-regulation. The more proficient L2 learners engaged more frequently in self-regulation. In the analysis of L1 and L2 differences in the non-verbal message (emotions through facial expressions), we analyzed the differences in the affective states of the speaker while narrating the story in both their own native language (e.g. Spanish) and in the foreign language (e.g. English) by tracking the emotions on the speaker's face. FaceEmo software was applied on 4 different videos:

Video label: C1, Sex: Female, Age: 41, Native Language: Spanish
Video label: C2, Sex: Female, Age: 41, Native Language: Spanish; Second Language: English
Video label: N1, Sex: Female, Age: 43, Native Language :Bulgarian
Video label: N2, Sex: Female, Age: 43, Native Language :Bulgarian; Second Language: English

After analyzing the results, we found that the highest three emotions probability for each video are as following:

C1:
1.neutral

2.happiness
3.disgust

C2:
1.neutral
2.happiness
3.disgust

N1:
1.neutral
2.happiness
3.disgust

N2:
1.neutral
2.happiness
3.disgust

Results from Study 1 indicate that subjects expressed more emotions (both positive and negative) while narrating the story in their L2, while the overall affective state of the speaker was accompanied by a “neutral” facial expression when narrating the same story in their L1.

4.2. Study 2

To compare results from Study 1, we used the Remote Collaborative and Affective Interactions (RECOLA) database, which is a multimodal corpus of spontaneous collaborative and affective interactions in French where participants are recorded in dyads during a video conference while completing a task requiring collaboration. In Study 2, we performed experiments on 3 videos selected from the database of recordings of online dyadic interactions between 46 French speaking participants solving a task in collaboration. To analyze differences in affect and emotion as expressed in L2 words processing, FaceEmo software was applied on 3 different videos selected from the RECOLA dataset:
Video label: P23, Sex: Female, Age: 30, Native Language:French
Video label: P41, Sex: Male, Age: 23, Native Language :German
Video label: P30: Sex: Female, Age: 22, Native Language :Italian

After analyzing the results we found that the highest three emotions probability for each video are as following:
P23:
1.neutral
2.happiness
3.disgust

P41:
1.sadness
2.surprise
3.fear

P30:
1.sadness
2.anger
3.fear

Results from Study 2 indicate that subjects express a “neutral” emotion when they collaborate in task solving in their native language (e.g. P23 –native French) and that they express a larger quantity of negative emotions (sadness, anger, etc.) when attempting to solve the task when French is not their native language (e.g. P41-native German and P30 –native Italian).

IV. FIGURES AND TABLES

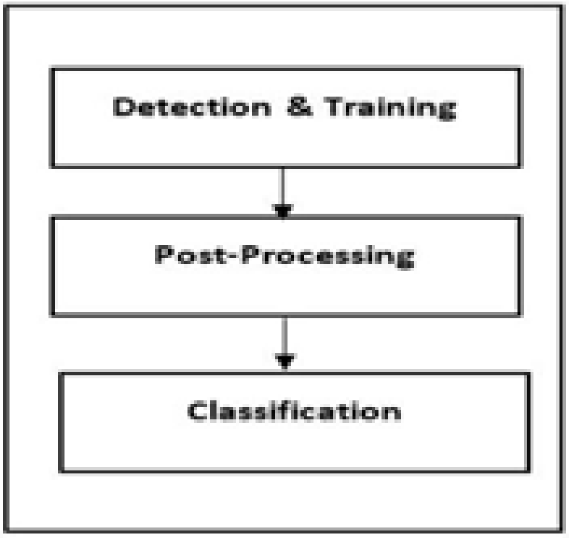

Fig. 1 the proposed method's algorithm

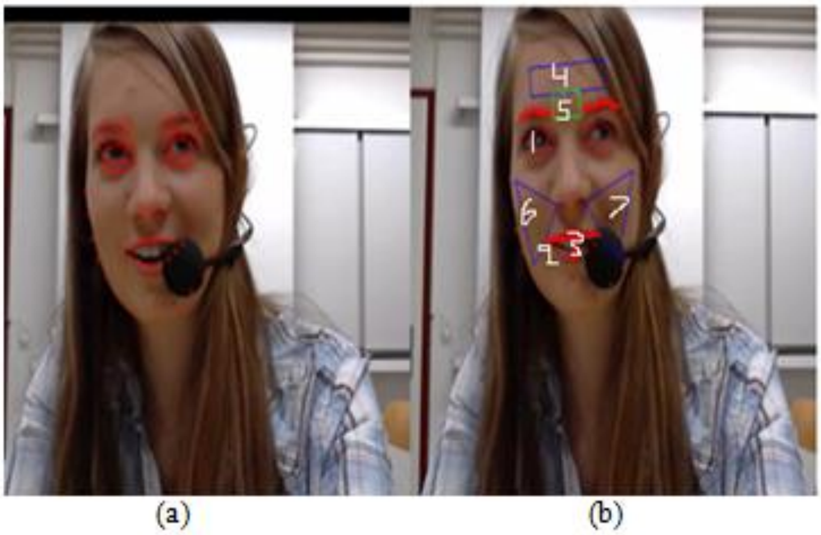

Fig. 2 face landmarks (a), areas of interest used in attribute derivation(b)

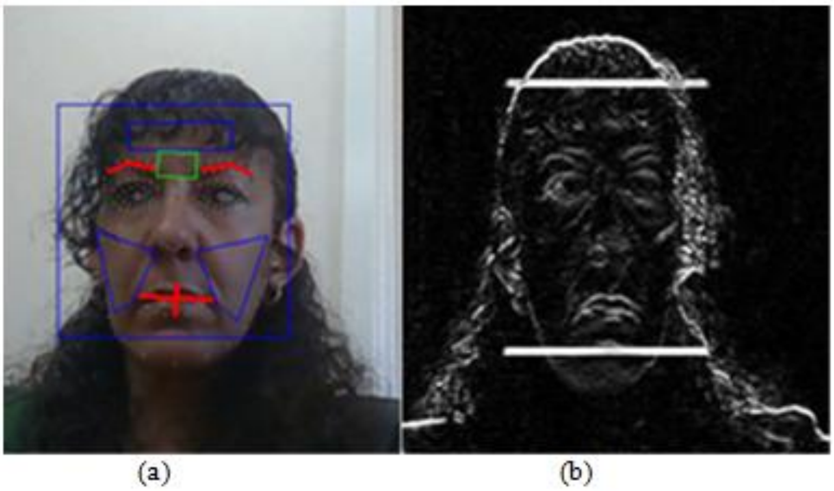

Fig. 3 horizontal and vertical domain in expression (a) and corresponding vertical edge domain (b)

Table 1 emotional expressions and their descriptions

EMOTIONAL	MOVEMENT CLUES
Surprise	Rise of eyebrows, sight opening of mouth, slight fall of chin
Anger	Frowning of eyebrows, tightening of lips and standing out of eyes.
Happiness	Rise and fall of mouth edges
Sadness	Fall of mouth edges and frowning of inner eyebrows
Fear	Rise of eye brows, standing out of eyes and slight opening of mouth
Disgust	Rise of upper lip, wrinkle of nose, fall of cheeks

V. CONCLUSION

The purpose of this analysis was to identify what differences exist between both the verbal and non-verbal expression of L2 learners when engaging in a simple narrative. Evidence suggests that there are significant differences in both the verbal (semantic) and non-verbal (affective, emotional facial expression) delivery of the message communicated when processing L2 words. The level of language proficiency also played a significant role in terms of affect, emotion and self-regulation. This research has future applications not only in language teaching and learning, but also in fields that already utilize linguistic deception detection, such as law enforcement, intelligence agencies, and forensic context.

REFERENCES

[1]. Sueyoshi, Ayano and Debra M. Hardison. 2005. "The Role of Gestures and Facial Cues in Second Language Listening Comprehension." Language Learning 55 (4): 661-699.

[2]. van Oers, Bert, Wim Wardekker, Ed Elbers, and René Van der Veer, eds. The transformation of learning: Advances in cultural-historical activity theory. Cambridge University Press, 2008.

[3]. McCafferty, Steven G. 2008. "MIMESIS AND SECOND LANGUAGE ACQUISITION: A Sociocultural Perspective." Studies in Second Language Acquisition 30 (2). Cambridge University Press: 147–67. doi:10.1017/S0272263108080297.

[4]. Gullberg, Marianne. Gesture as a communication strategy in second language discourse: A study of learners of French and Swedish. Vol. 35. Linguistics and Phonetics, 1998.

[5]. Ringeval, Fabien, Andreas Sonderegger, Juergen Sauer, and Denis Lalanne. "Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions." In Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on, pp. 1-8. IEEE, 2013.

[6]. Bebawy, Monica, Suzan Anwar, and Mariofanna Milanova. "Active Shape Model vs. Deep Learning for Facial Emotion Recognition in Security." In IAPR Workshop on Multimodal Pattern Recognition of Social Signals in Human-Computer Interaction, pp. 1-11. Springer, Cham, 2016.

[7]. Sianipar, Agnes, Renée Middelburg, and Ton Dijkstra. "When feelings arise with meanings: how emotion and meaning of a native language affect second language processing in adult learners." PloS one 10, no. 12 (2015): e0144576.

[8]. Dewaele, Jean-Marc, and Aneta Pavlenko. "Emotion vocabulary in interlanguage." Language learning 52, no. 2 (2002): 263-322.

[9]. T. F. Cootes, C. J. Taylor, D. H. Cooper ve J. Graham,others, "Active shape models-their training and application,"Computer vision and image understanding, c. 61, sf. 38–59,1995.

[10]. Y. Wei, "Research on Facial Expression Recognition and Synthesis," Y.L. Tezi, Department of Computer Science and Technology, Nanjing University, 2009. http://code.google.com/p/asmlibrary/

[11]. İ. Arı ve L. Akarun, "Yüz Özniteliklerinin Takibi ve İşaret Dili için İfade Tanıma," IEEE Sinyal İşleme ve İletişim Uygulamaları Konferansı, 2009

[12]. P. Ekman, W. Friesen, "Facial Action Coding System: A Technique for the Measurement of Facial Movement", Palo Alto, California: Consulting Psychologists Press. (1978)

[13]. McCafferty, Steven G. "The Use of Private Speech by Adult Second Language Learners: A Cross-Cultural Study." The Modern Language Journal76, no. 2 (1992): 179-189.
[14]. Note that thesis title is set in italics and the university that granted the degree is listed along with location information

*Zvetomira Svetleff . "Facial Expression of Emotion Assessment from Video in Adult Second Language Learners." International Journal Of Modern Engineering Research (IJMER), vol. 07, no. 08, 2017, pp. 27–33.