Measuring the Perceived Difficulty of a Lecture Using Automatic Facial Expression Recognition

Jacob Whitehill, Marian Bartlett, and Javier Movellan

Machine Perception Laboratory
University of California, San Diego
{jake,movellan}@mplab.ucsd.edu, marni@salk.edu

Abstract. This project explores the idea of facial expression for automated feedback in teaching. We show how automatic real-time facial expression recognition can be effectively used to estimate the difficulty level, as perceived by an individual student, of a delivered lecture. We also show that facial expression is predictive of an individual student's preferred rate of curriculum presentation at each moment in time. On a video lecture viewing task, training on less than two minutes of recorded facial expression data and testing on a separate validation set, our system predicted the subjects' self-reported difficulty scores with mean accuracy of 0.42 (Pearson r) and their preferred viewing speeds with mean accuracy of 0.29. Our techniques are fully automatic and have potential applications for both intelligent tutoring systems (ITS) and standard classroom environments. [1]

1 Introduction

This paper explores the utility of *facial expression* as a feedback signal from student to teacher. Previous work [2,3] has investigated using facial expression for predicting the student's affective state. This paper investigates the usefulness of automatic expression recognition for two tasks: (1) measuring the student's perception of difficulty of a delivered lecture at each moment in time, and (2) determining a student's preferred viewing speed of lesson material.

2 Experiment

We conducted a pilot experiment in which subjects viewed a video lecture at an adjustable speed while their facial expressions were recognized automatically and recorded. Using the "difficulty" scores that the subjects report, the correlations between facial expression and difficulty, and between expression and preferred viewing speed, were assessed. Each of 8 human subjects watched a recorded video lecture 200 seconds (3 min 20 sec) in length. The video consisted of 7 short video clips concatenated together about a disparate range of topics, including physics, philosophy, math, and teenage gossip. The experiment proceeded as follows:

[1] The full version of this paper is available at [1].

B. Woolf et al. (Eds.): ITS 2008, LNCS 5091, pp. 668–670, 2008.

1. **Watch the video lecture.** The playback speed could be adjusted by the subject using the keyboard. Facial expression data of each human subject were recorded while they watched the video.
2. **Take the quiz.** The quiz consisted of 6 specific questions about the lecture.
3. **Self-report on the difficulty.** The lecture was re-played at a speed of 1.0. Subjects adjusted their rating of the video's difficulty on a scale of 1 to 10 on a frame-by-frame basis using arrow keys.

Facial expressions were analyzed and recorded automatically using the CERT system [4]. CERT outputs estimated intensities of facial "action units" (AUs) of face videos in real time. AUs are the component muscle movements that comprise facial expressions, analagous to phonemes for words, and are defined in the Facial Action Coding System [5], which is the standard framework for objectively coding human facial expression. CERT recognizes AUs 1, 2, 4, 5, 9, 10, 12, 14, 15, 17, 20, and 45 as well as Smile.

None of the subjects was aware of the purpose of the experiment or that expression data were measured. Subjects were encouraged to view the video efficiently (by adjusting its speed) by a conspicuous automatic "shut-off" timer and a quiz. In actuality, the timer had no effect, and the quiz was not graded.

3 Analysis

Our experiment yielded three data series which we time-aligned:

1. **Difficulty:** The subjects' self-reported scores of how difficult they perceived the lecture at each moment in time.
2. **Speed**: The lecture speed at each moment as controlled by the user.
3. **Expression:** A set of real-valued expression channels output by CERT which estimate the intensity of 13 expression channels (12 AUs + Smile). The channels were smoothed using local quadratic regression.

We employed linear regression over the Expression channels to predict both Difficulty and Speed (see Figure 1 for an example). Separate regression models were trained for each subject. The average correlation (over all 8 subjects) of the Difficulty scores predicted by the regression model (training over all data) with the self-reported Difficulty scores was 0.75, suggesting that facial expression contains a large amount of useful signal.

Subjects exhibited a wide variability of which facial muscle movements were correlated with Difficulty and Speed. The facial movement most consistently correlated was AU 45 ("blink"). For six out of eight subjects, the correlation was negative (subjects blinked less during more difficult lecture segments), which is consistent with evidence from experimental psychology that blink rate decreases when interest or mental load is high [6].

Generalization Performance: To assess the effectiveness of using facial expression to predict Difficulty and Speed values in a real intelligent tutoring system, we split the data series into disjoint "bands" ($\sim$ 12 sec long) of training

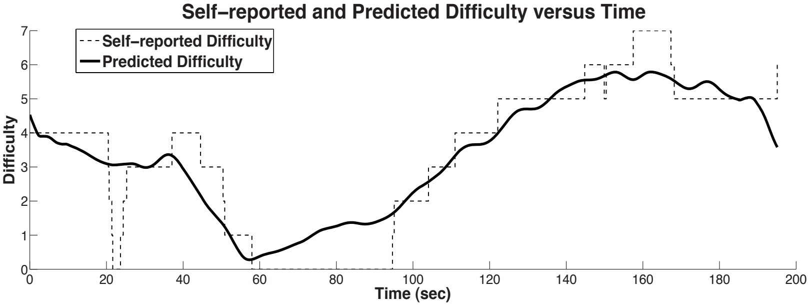

Fig. 1. Comparison of the self-reported Difficulty scores with the predicted Difficulty scores using linear regression over the facial expression channels for Subject #6. For this figure, all data were used for training the regression model.

and validation sets (see [1] for details). Linear regression models over the Expression data series were trained to predict Difficulty and Speed for each human subject. Generalization performance was assessed by correlating predicted to self-reported Difficulty and Speed values over the validation data.

The average correlation across all subjects of predicted to self-reported Difficulty was 0.42; the average correlation for Speed was 0.29. Individual correlations were statistically significant for all but one subject. While these results show room for improvement, they already demonstrate that automatic facial expression recognition can extract relevant information for ITS.

4 Conclusions

Our results show that automatic facial expression recognition is already a useful feedback signal for intelligent tutoring systems for two concrete tasks: perceived difficulty estimation, and preferred speed prediction. As expression recognition technology improves, its usefulness in ITS will continue to grow.

References

1. Whitehill, J., Bartlett, M., Movellan, J.: Measuring the perceived difficulty of a lecture using automatic facial expression recognition. Technical Report 2008.01, Machine Perception Lab (2008), `http://mplab.ucsd.edu/~jake/its08.pdf`
2. Rio, H., Soli, A.L., Aguirr, E., Guerrer, L., Santa, J.P.A.: Facial expression recognition and modeling for virtual intelligent tutoring systems. In: Proceedings of the Mexican International Conference on Artificial Intelligence (2000)
3. D'Mello, S., Picard, R., Graesser, A.: Towards an affect-sensitive autotutor. IEEE Intelligent Systems, Special issue on Intelligent Educational Systems 22(4) (2007)
4. Bartlett, M., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., Movellan, J.: Auto. recog. of facial actions in spontaneous expressions. Jour. of Multimedia (2006)
5. Ekman, P., Friesen, W.: The Facial Action Coding System: A Technique For The Measurement of Facial Movement. Consulting Psychologists Press (1978)
6. Holland, M., Tarlow, G.: Blinking and mental load. Psych. Reports 31(1) (1972)