Multimodal Perception of Attitudes: A Study on Video Blogs

Noor Alhusna Madzlan[1][3], Justine Reverdy[2], Francesca Bonin[2], Loredana Sundberg Cerrato[2] and Nick Campbell[2]

[1]CLCS, School of Linguistics, Speech and Communication Sciences, Trinity College Dublin
[2]SCSS, School of Computer Science and Statistics, Trinity College Dublin, Ireland
[3]ELLD, Faculty of Languages and Communication, UPSI, Malaysia

madzlann@tcd.ie, justine.reverdy@gmail.com, boninf@tcd.ie, cerratol@tcd.ie, nick@tcd.ie

Our paper reports a study on perception of multimodal attitude expressions of video bloggers (in short,vloggers) on YouTube. In communicative settings, expressions of attitude, affect or emotion are seen as integral for effective communication. Correct attitude expression is essential to avoid any misinterpretation between interlocutors. Speakers dynamically express their attitudes through multimodal signals such as verbal signals as well as visual signals, in particular, facial gestures [2], [3]. We developed an annotation scheme to annotate a corpus of 250 vlogs. This annotation scheme, named N5 is a derivation of the standard A10 attitude annotation by Henrichsen and Allwood [1]. The A10 attitudes are: Interested, Friendly, Casual, Bored, Thoughtful, Confident, Amused, Enthusiastic, Uninterested and Impatient. We aim at assessing whether our subset of the A10 attitude annotation scheme, the N5 Attitude Annotation Scheme (Amusement, Enthusiasm, Friendliness, Frustration and Impatience), is also a valid representation of the vlog genre. To achieve this, we conduct a perception test involving a group of 20 anonymous non-expert public participants.

This study aims at measuring respondents' attitude identification as well as the confidence rates of their attitude judgments. The respondents were asked to select attitude states that best represent the vloggers' expression. These speaker attitude expressions are presented through several modalities; audio only, video only and audio-visual modality. This is to determine which modality provides most information for respondents' to identify the relevant attitude. We specified the five attitudes and included the remaining 6 attitudes from the A10 attitude annotation scheme in the “Other” category for the public respondents to select. After

selecting an attitude, respondents were asked to decide, on a scale ranging from 1 to 7 (Unsure to Very Certain), how certain they were about their judgments on the attitude selection.

We analysed the results from three perspectives; inter-annotator agreement, certainty level for each modality and certainty level of attitude choice. For results of inter-annotator agreement, we found a "fair agreement" between all 20 raters with a k-value of 0.27 using weighted Fleiss Kappa [4]. The low value for agreement is expected due to the large number of raters involved in the test. It is in fact challenging to obtain higher scores for attitude perception due to several factors such as age, gender and cultural backgrounds of the participants themselves. We further analysed the relevance of multimodality for attitude perception and the results indicate that a fusion of audio and visual information is most helpful for participants to perceive attitude expressions of vlog speakers. Following that, we conducted analysis on the confidence level of participants with their attitude choice. Results show that the attitude ``Frustration" shows highest selection rate compared to other attitudes. This justifies our inclusion of the "Frustration" state as an attitude class that is salient in our vlog dataset. There is also not enough consistency in the "Other" category to justify inclusion of an extra attitude in our attitude scheme.

Results from this perception study lead us to report on a more valid justification for selection of attitudinal states in our attitude annotation scheme. Our findings suggest that our N5 attitude categories seems to be a sufficient scheme to annotate attitudes in vlogs.

References

[1] Henrichsen, Peter J and Allwood, J, *Predicting the attitude flow in dialogue based on multi-modal speech cues*, NEALT PROCEEDINGS SERIES, 2012.

[2] Madzlan, N, Han, J, Bonin, F and Campbell, N , *Towards Automatic Recognition of Attitudes: Prosodic Analysis of Video Blogs*, SPEECH PROSODY, 2014.

[3] Madzlan, N, Han, J, Bonin, F and Campbell, N , *Automatic Recognition of Attitudes in Video Blogs - Prosodic and Visual Feature Analysis*, INTERSPEECH, 2014.

[4] Fleiss, J. L. (1981) *Statistical methods for rates and proportions*. 2nd ed. (New York: John Wiley) pp. 38–46