MIT Media Laboratory Technical Note No. 584, September 2004

Voices of Attraction

Anmol Madan, Ron Caneel, Alex "Sandy" Pentland
MIT Media Laboratory
20 Ames Street, Cambridge 02139
{anmol,rcaneel,sandy} @ media.mit.edu

Abstract

Non-linguistic social signals (e.g., `tone of voice') are often as important as linguistic content in predicting behavioural outcomes [1,2]. This paper describes four automated measures of such social signalling within the experimental context of speech in speed dating.

1. Introduction

In many situations non-linguistic social signals (body language, facial expression, tone of voice) are as important as linguistic content in predicting behavioural outcome [1,2]. Tone of voice and prosodic style are among the most powerful of these social signals even though (and perhaps because) people are usually unaware of them [2]. In a wide range of situations (marriage counselling, student performance assessment, jury decisions, etc.) an expert observer can reliably quantify these social signals and with only a few minutes of observation predict about 1/3d of the variance in behavioural outcome (which corresponds to a 70% binary decision accuracy) [1]. It is astounding that observation of social signals within such a `thin slice' of behaviour can predict important behavioural outcomes (divorce, student grade, criminal conviction, etc.) when the predicted outcome is sometimes months or years in the future.

2. Measuring Social Signals

Pentland [3] constructed measures for four types of vocal social signaling, designated activity level, engagement, stress, and mirroring. These four measures were extrapolated from a broad reading of the voice analysis and social science literature, and we are now working to establish their general validity

Calculation of the activity measure begins by using a two-level HMM to segment the speech stream of each person into voiced and non-voiced segments, and then group the voiced segments into speaking vs. non-speaking [4,5]. Conversational activity level is measured by the z-scored percentage of speaking time plus the frequency of voiced segments.

Engagement is measured by the z-scored influence each person has on the other's turn taking. When two people are interacting, their individual turn-taking dynamics influence each other and can be modeled as a Markov process [6]. By quantifying the influence each participant has on the other we obtain a measure of their engagement...popularly speaking, were they driving the conversation? To measure these influences we model their individual turn-taking by an Hidden Markov Model (HMM) and measure the coupling of these two dynamic systems to estimate the influence each has on the others' turn-taking dynamics [7]. Our method is similar to the classic method of Jaffe et al. [6], but with a simpler parameterization that permits the direction of influence to be calculated and permits analysis of conversations involving many participants.

Stress is measured by the variation in prosodic emphasis. For each voiced segment we extract the mean energy, frequency of the fundamental format, and the spectral entropy. Averaging over longer time periods provides estimates of the mean-scaled standard deviation of the energy, formant frequency and spectral entropy. The z-scored sum of these standard deviations is taken as a measure speaker stress; such stress can be either purposeful (e.g., prosodic emphasis) or unintentional (e.g., physiological stress caused by discomfort).

Mirroring behavior, in which the prosody of one participant is `mirrored' by the other, is considered to signal empathy, and has been shown to positively influence the outcome of a negotiation [8]. In our experiments the distribution of utterance length is often bimodal. Sentences and sentence fragments typically occurred at several-second and longer time scales. At time scales less than one second there are short interjections (e.g., `uh-huh'), but also back-and-forth exchanges typically consisting of single words (e.g., `OK?', `OK!', `done?', `yup.'). The z-scored frequency of these short utterance exchanges is taken as a measure of mirroring. In our data these short utterance exchanges were also periods of tension release.

2.1. Signaling Dynamics

These measures of social signaling can be computed on a conventional PDA in real-time, using a one-minute lagging window during which the statistics are accumulated. It is therefore straightforward to investigate how these `social signals' are distributed in conversation. In [9] we analyzed social signaling in 54 hours of two-person negotiations on a minute-by-minute basis. We observed that high numerical values of any one measure typically occur by themselves, e.g., periods with high engagement do not show high stress, etc., so that each participant exhibits four `social display' states, plus a `neutral' relaxed state in which the participant is typically asking emotionally neutral questions or just listening. The fact that these display states were largely unmixed provides evidence that they are measuring separate social displays.

2.2. Attraction Experiment

Speed dating is relatively new way of meeting many potential matches during an evening. Participants interact for five minutes with their `date', at the end of which they decide if they would like to provide contact information to him/her, and then they move onto the next person. A 'match' is found when both singles answer yes, and they are later provided with mutual contact information.

In this experiment we analyzed 57 five-minute speed-dating sessions. In addition to the `romantically attracted' (provide contact information) question, participants were also asked two other yes/no questions: would they like to stay in touch just as friends, and would they like to stay in touch for a business relationship. These `stay in touch' questions were hypothetical, since contact information would not be exchanged, but allowed us to explore whether romantic attraction could be differentiated from other factors.

2.2.1. Results

The four social signaling measures for both male and female were compared by linear regression to the question responses, and in each case the resulting predictor could account for more than 1/3rd of the variance. For the females responses, for instance, the correlation with the 'attracted' responses were r=0.66, p=0.01, for the 'friendship' responses r=0.63, p=0.01, and for the 'business' reponses r=0.7, p=0.01. Corresponding values for the male responses were r=0.59, r=0.62, and r=0.57, each with p=0.01.

The engagement measure was the most important individual feature for predicting the `friendship' and `business' responses. The mirroring measure was also significantly correlated with female `friendship' and `business' ratings, but not with with male ratings. The stress measure showed correlation with both participants saying 'yes' or both saying 'no' for the 'attraction' (r=0.6,p=0.01) and 'friendship' (r=0.58,p=0.01) questions.

An interesting observation was that for the `attracted' question female features alone showed far more correlation with both male (r=0.5,p=0.02) and female (r=0.48, p=0.03) responses than male features (no significant correlation). In other words, female social signaling is more important in determining a couples 'attracted' response than male signaling. The most predictive individual feature was the female activity measure.

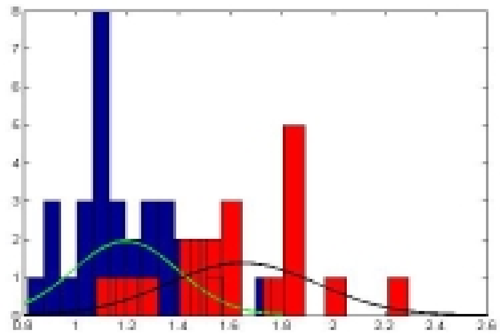

Figure 3: Frequency distribution of female `attracted' responses (red=yes) vs. predictor value. The cross-validated binary linear decision rule has 72% accuracy

Figure 3 shows a two-class linear classifier for the `attraction' responses, based on the social signaling measures; this classifier has a cross-validated accuracy of 71% for predicting the `attracted' response. The two fitted Gaussians are simply to aid visualization of the distributions' separability.

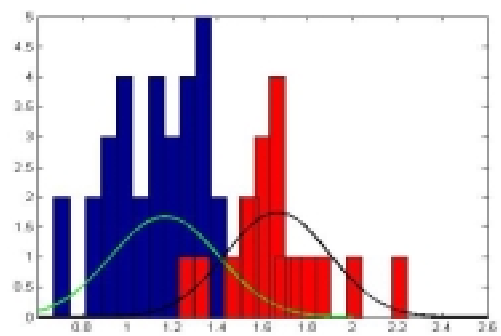

Figure 4. Frequency distribution of female `business' responses (red=yes) vs. predictor value. The cross-validated three-class linear decision rule produces 83 % accuracy.

Figure 4 illustrates a two-class linear classifier for the `business' responses, based on the social signaling measures; this classifier has a cross-validated accuracy of 74% for predicting the `attracted' response. By considering the overlapping region as a third class, we can increase the cross-validation accuracy to 83% for the yes and no response regions. The two fitted Gaussians are simply to aid visualization of the distributions' separability.

3. Discussion

The social signaling discussed in this paper seems to communicate and be involved in mediating social variables such as status, interest, determination, or cooperation, and arise from the interaction of two or more people rather than being a property of a single speaker. Semantics and affect are important in determining what signaling an individual will engage in, but they seem to be fundamentally different types of phenomena. The social signaling measured here seems to be a sort of `vocal body language' that operates relatively independently of linguistic or affective communication channels.

Finally, it is interesting to speculate about what might happen if people were made more aware of their social signaling. One idea is to construct a small wearable `social signaling' meter that could provide users with real-time feedback. We are now beginning tests with such a meter and expect to be able to report the results by the time of the conference.

References

[1] Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behaviour as predictors of interpersonal consequences: A meta-analysis. Psychological Bulletin, 111(2), 256-274.

[2] Nass, C., and Brave, S. (2004) Voice Activated: How People Are Wired for Speech and How Computers Will Speak with Us, MIT Press

[3] Pentland, A. (2004) Social Dynamics: Signals and Behavior, ICDL, San Diego, CA Oct 20-23, IEEE Press

[4] Handel, Stephen, (1989) Listening: an introduction to the perception of auditory events, Stephen Handel , Cambridge: MIT Press

[5] Basu, B., (2002) Conversational Scene Analysis, doctoral thesis, Dept. of Electrical Engineering and Computer Science, MIT. 2002. Advisor: A. Pentland

[6] Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., & Jasnow, M. (2001). Rhythms of dialogue in early infancy. Monographs of the Society for Research in Child Development, 66(2), No. 264.

[7] Choudhury, T., and Pentland, A., (2004), NAASCOS, June 27-29, Pittsburgh, PA. PDF available at http://hd.media.mit.edu

[8] Chartrand, T., and Bargh, J., (1999) The Chameleon Effect: The Perception-Behavior Link and Social Interaction, J. Personality and Social Psychology, Vo. 76, No. 6, 893-910

[9] Khilnani, R. (2004) Temporal Analysis of Stages in Negotiation, MEng Project, Advisor: A. Pentland.

[10] Pentland, A., Curhan, J., Khilnani, R., Martin, M., Eagle, N., Caneel, R., Madan A (2004) ``Toward a Negotiation Advisor,'' UIST 04, Oct 24-27, ACM. PDF available at http://hd.media.mit.edu