MIT Media Laboratory Technical Note 579, appears ICDL'04, San Diego, Oct. 20-22

Social Dynamics: Signals and Behavior

Alex Pentland
MIT Media Lab, E15-387, 20 Ames St, Cambridge MA 02139

Abstract

Nonlinguistic 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 measure of such social signalling, and shows that they can be used to form powerful predictosr of objective and subjective outcomes in several important situations. Finally, it is argued that such signals are important determinants of social position.

1. Introduction

Animals communicate their social structure in many ways, including dominance displays, relative positioning, access to resources, etc. Humans add to that a wide variety of cultural mechanisms such as clothing, seating arrangements, and name-dropping. Most of these culture-specific social communications are conscious and are often manipulated.

However in many situations non-linguistic social signals (body language, facial expression, tone of voice) are as important as linguistic content in predicting behavioral 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 counseling, 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 behavioral outcome (which corresponds to a 70% binary decision accuracy) [1]. It is astounding that observation of social signals within such a `thin slice' of behavior can predict important behavioral outcomes (divorce, student grade, criminal conviction, etc.) when the predicted outcome is sometimes months or years in the future.

Nonlinguistic vocal signaling is a particularly familiar part of human behavior. For instance, we speak of someone `taking change' of a conversation, and in such a case this person might be described as `driving the conversation' or `setting the tone' of the conversation. Such dominance of the conversational dynamics is popularly associated with higher social status or a leadership role. Similarly, some people seem skilled at establishing a `friendly' interaction. The ability to set conversational tone in this manner is popularly associated with good social skills, and is typical of skilled salespeople to social `connectors' [3].

The machine understanding has studied human communication at many time scales --- e.g., phonemes, words, phrases, dialogs --- and both semantic structure and prosodic structure has been analyzed. However the sort of longer-term, multi-utterance structure associated with social signaling has received relatively little attention [4]. In this paper I develop an automatic measurement method for quantifying some of these non-linguistic social signals, and describe how these measurements can be used to form powerful predictors of behavioral outcome in some very important types of social interaction: getting a date, getting a job, and getting a raise.

2. Measuring Social Signals

I have constructed measures for four types of vocal social signaling, which I have 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. To date they have been used to predict outcomes in salary negotiation, dating, friendship, and business preferences with accuracy comparable to that of human experts in analogous situations.

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 [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-talking. When two people are interacting, their individual turn-taking dynamics influences 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 (described in more detail in the next section) on a minute-by-minute basis. We observed that high numerical values of any one measure typically occur by themselves, e.g., during periods in which participants showed high engagement they did not use high stress, etc., so that each participant exhibits four `social display' states, plus a `neutral' relaxed state in which the participant is typically asking neutral questions or just listening. The fact that these display states were largely unmixed provides evidence that they are measuring separate social displays.

The signaling state of the two participants was strongly coupled, so that (ignoring symmetries and outliers) the joint state space has only six states rather than the expected fifteen. For instance, when one participant displayed engagement, the other participant almost always followed suit (90% of the time), resulting in a highly engaged, roughly equal conversation. When one participant displayed mirroring behavior, the other would usually join in (74% of the time). When one participant became active, the other became neutral (75% of the time). However when one participant used stress, the result differed according to status. If the high status participant used stress, then low-status participant would usually (66% of the time) signal activity and only 11% of the time would the low-status participant also show stress. When the low-status participant used stress, the high-status participant would usually become active (54% of the time) but 24% of the time would respond with matching stress.

2.2. Negotiation Experiment

In this experiment we investigated what might be thought to be a prototypically rational form of communication: negotiating a salary package with your boss. The intuition is that negotiation participants who `take change' of the dynamics of the conversation, what might be described as `driving the conversation' will do better than those who are more passive.

In Pentland, Curhan, et al [10] we collected audio from forty-six gender-matched dyads (either male/male or female/female, 28 male dyads and 18 female dyads) that were asked to conduct a face-to-face negotiation as part of their class work. The mock negotiation involved a Middle Manager (MM) applying for a transfer to a Vice President's (VP) division in a fictitious company. Many aspects of the job were subject to negotiation including salary, vacation, company car, division, and health care benefits; these aspects were summed into an overall objective score based on their market value. Participants were offered a real monetary incentive for maximizing their own individual outcome in the negotiation. Subjects were first year business students at MIT Sloan School of Management, almost all with previous work experience.

Data collected included individual voice recordings of both parties in a closed room plus ratings of subjective features. There was no time limit set and the negotiations length ranged from 10 to 80 minutes in length, with an average duration of approximately 35 minutes, for a total of 54 hours of data.

Subjective features analyzed were the answers to the questions `What kind of impression do you think you made on your counterpart?' `To what extent did your counterpart deliberately let you get a better deal than

he/she did?' and `To what extent did you steer clear of disagreements?

2.2.1. Results

Our hypothesis was that negotiation participants who showed the most engagement, stress and mirroring would do better than those who were more passive, i.e., that the time-averaged influence on each participant + amount of stress + amount of mirroring would predict the objective outcome of the negotiation. Following [1], we measured signaling in only the first five minutes of the negotiation and used that `thin slice' of behavior to predict the final negotiation outcome.

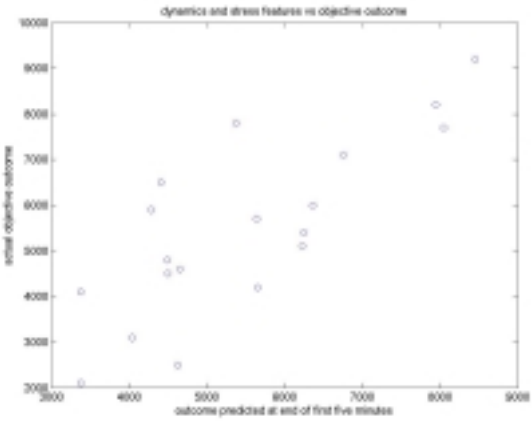

Figure 2: Outcome predicted from social signals at end of first five minutes of negotiation. (Female VPs shown).

This predictor had a strong (r= 0.57, p=0.001) correlation with the objective outcome of the negotiation. Thus the accuracy of this predictor is similar to that of human experts performing similar tasks [1].

Post-hoc analysis showed that the relationship differed for high- and low-status participants. For VPs, engagement + stress predicted almost half of their variation in outcome (r=0.75). For MMs, the mirroring measure alone predicted almost a third of the variation in their objective outcome (r=0.57).

The engagement measure had a significant positive correlation (r=0.63) with the subjective "impression I thought I made on my partner" rating and a with the "did your partner let you win" rating (r=0.65). The mirroring measure had a significant positive correlation with the extent to which participants said they were seeking to avoid disagreements (r=0.62).

2.3. 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 Madan, Caneel and Pentland [11] we analyzed 57 five-minute speed-dating sessions. In addition to the `romantically attracted' question (where a postive answer from both participants resulting in sharing of contact information), 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 in any case, but allowed us to explore whether vocal signals of romantic attraction could be differentiated from other types of attraction.

2.3.1. Results

Liniar regression was used to form predictors of the question responses using the values of the four social signaling measures. For each question the resulting predictor could account for more than 1/3rd of the variance, providing approximately 70% accuracy at predicting the questions response. This accuracy is comparable to that of human experts performing similar tasks [1].

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.

For the ‘attracted’ question the most predictive individual feature was the female activity measure. 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.

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.

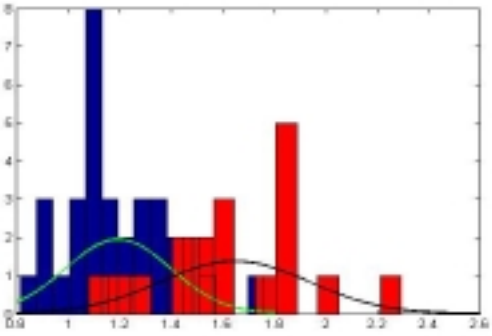

Figure 3: Frequency of female `attracted' responses (black=no) vs. predictor value. The cross-validated linear decision rule produces 71% accuracy.

Figure 3 illustrates a two-class linear classifier 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.

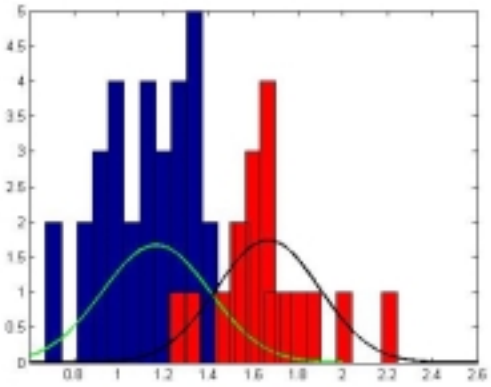

Figure 4. Frequency of female `business' responses (black=no) vs. predictor value. The cross-validated linear decision rule produces 74% accuracy..

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%. The two fitted Gaussians are simply to aid visualization of the distributions' separability.

2.4. Social Network Experiment

In Choudhury and Pentland [7] we collected audio data from 23 subjects from 4 different research groups over a period of 11 days, resulting in an average of 66 hours of data per subject. The subjects were a representative sample of the community, including students, faculty and administrative staff. During data collection users had a small audio recording device on them for six hours a day (11AM -5PM) while they were on the MIT campus. The data was automatically analyzed to detect the pair-wise conversations, and this was used to analyze the distribution of conversational statistics within the sampled conversations, including the engagement and activity measures

2.4.1. Results

Our first finding was that the Markov statistics describing individuals' turn-taking styles are distinctive and stable across different conversational partners, and that these turn-taking patterns are not just a noisy variation of the same average style ($p < 0.001$). Since these Markov statistics effectively determine the average values of the activity and engagement measures, the implication is that people have characteristic patterns of activity and engagement signaling.

Male and female patterns were extremely different, with only slight overlap between the range of parameters observed for males and those observed for females. Surprisingly, the total speaking time for males and females was nearly equal.

Choudhury [12] in her PhD thesis investigated the hypothesis that the `engagement' measure (e.g., the inter-Markov process influence parameter) would be correlated with information flow within their social network structure. To investigate this hypothesis she compared the engagement measure to the individual subject's betweeness centrality, which is a standard social science measure of how important an individual is to information flow within a social network [13]. The correlation value between this centrality measure and the influence parameter was 0.90 (p-value < 0.0004, rank correlation 0.92). Thus the amount of time an individual displayed engagement was a nearly perfect predictor of how much of a `connector' they were.

3. Discussion

In this paper I have proposed a method of measuring social signaling using non-linguistic vocal features, and shown that the measured signaling can be used to create powerful predictors of both objective and subjective outcome in social situations. In addition, at least some aspects of people's position in a social network appears to be signaled and negotiated via this same mechanism.

In our negotiation experiment we showed that signaling during the first five minutes of a negotiation account for more than 1/3d of the variation in objective outcome, and that the `winning' strategy is different for high-status vs low-status participants. For the high-status participants engagement and use of stress was most

important. For low-status participants the use of mirroring was most important. The correlation between the engagement measure and the subjective questions concerning control, and between the mirroring measure and the subjective question concerning cooperation, support the validity of these features as measures of social signaling.

In our attraction experiment we showed that signaling during the first five minutes of conversation again accounted for more than 1/3d of the variation in outcome, and that the signaling and interaction is different for male and female participants. The high correlation between the female activity measure and the `attraction' responses aupports the validity of this feature as a measure of social signaling.

We found that in a research laboratory environment peoples' signaling mirrored the information flow within the social network. The more a person was a `connector' within the social network, the more they displayed engagement. This suggests that social signals are part of a continual, implicit `negotiation' between members of a social network that establishes each individual's appropriate position in the network.

Is this social signaling just a part of `normal' speaking prosody? Prosody is most commonly studied within the framework of speech understanding, where pitch, duration, and amplitude are used to modify, select, or emphasize the semantics conveyed by the words [2,4]. In contrast to this type of prosody the vocal features measured in these experiments occur at time scales that are far too long to be related to individual words or phrases.

The social signaling discussed in this paper instead 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 phenomina. 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 that people in these experiments were only vaguely aware of their own vocal characteristics, and they were unable to articulate the connection between these characteristics and the behavioral outcome. 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 behavior 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] Gladwell, M. (2000) The Tipping Point: How little things can make a big difference. New York: Little Brown

[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

[11] Madan, A., Caneel, R., and Pentland, A. (2004) GroupMedia: Distributed Multimodal Interfaces, ICMI, St. College, PA, Oct. 12-14, 2004. IEEE Press, PDF available at http://hd.media.mit.edu

[12] Choudhury, T., (2003) Sensing and Modeling Human Networks, Ph.D. Thesis, Dept. of Media Arts, and Sciences, MIT. Advisor: A. Pentland.

[13] Wasserman, S. and K. Faust, Social Network Analysis Methods and Applications. 1994: Cambridge University Press