A Negotiation Analyzer

Pentland, A., Curhan, J., Khilnani, R., Martin, M., Eagle, N., Caneel, R., Madan A.
Massachusetts Institute of Technology
Room E15-387, 20 Ames St, Cambridge MA 02139,USA
{pentland,curhan,reshma_k,mcm,natecow,rcaneel,anmol }@mit.edu

ABSTRACT
Nonlinguistic social signals are often as important as linguistic content in predicting behavioral outcomes [1,2]. We show that an automated measure of non-linguistic vocal signaling (`tone of voice') predicts more than 1/3d of the variation in a negotiation's objective and subjective outcomes**.**

Categories and Subject Descriptors: H.5.2 [**User Interfaces**]: Graphical User Interfaces (GUI), Natural Language, I.2.7 [**Natural Language Processing**]: Speech Recognition and Synthesis

Additional Keywords and Phrases: Negotiation, Social Signals

INTRODUCTION

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 counciling, student performance assessment, jury decisions, etc.) an expert observer can reliably quantify these social signals and with only a few minutes of observation and use them to predict the behavioral outcome with an average accuracy of about r=0.6, which corresponds to a 70% binary decision accuracy [1].

Pentland [3] developed an automatic measurement method for quantifying these non-linguistic social signals, and in this paper we use these measurements to develop a powerful predictor of the outcome of one very important type of interpersonal behavior: negotiating a salary package.

MEASURING SOCIAL SIGNALS IN VOICE

In Pentland [3] measures were constructed for four types of social signaling. These were 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 accurately predict outcomes in negotiation, speed dating, and hiring preferences, with r=0.6 or greater and binary decision accuracies averaging around 70%.

Calculation of all four measures 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]. 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 influences each other and can be modeled as a Markov process [5]. 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 [6]. Our method is similar to the classic method of Jaffe et al. [5], 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 or worry).

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 [7]. In our experiments the distribution of utterance length is often bimodal. Sentences and sentence

UIST '04, October 24--27, 2004, Santa Fe, New Mexico, USA
ACM 1-58113-962-4/04/0010

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 our measure of mirroring. In our data these short utterance exchanges were also periods of tension release.

In our data `signals' (high values of one measure) typically occur by themselves, e.g., periods with high engagement do not show high stress, so that each participant exhibits four display states. The state of the two participants are strongly coupled, so that the joint state space has only seven states rather than sixteen, and the transitioning between states is similarly constrained.

Negotiation experiment

Forty-six gender-matched dyads (either male/male or female/female, 28 male dyads and 18 female dyads) 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. Participants 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, the instrumental outcomes, plus ratings of subjective features. There was no time limit set and the negotiations ranged from 10 to 80 minutes in length, with an average duration of approximately 35 minutes. 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?'

Experimental 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 the other 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.

This predictor had a strong ($r = 0.57$, $p = 0.001$) correlation with the objective outcome of the negotiation. 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 significant correlation 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 ($r = 0.62$).

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. Pentland, A., (2004) Social Dynamics: Signals and Outcomes, ICDL04, San Diego, CA. See publications at http://hd.media.mit.edu for this and related papers.
4. Basu, B., (2002) Conversational Scene Analysis, doctoral thesis, Dept. of Electrical Engineering and Computer Science, MIT. 2002.
5. 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.
6. Choudhury, T., and Pentland, A., (2004), NAASCOS, Pittsburgh, PA. See publications list at http://hd.media.mit.edu for this and related papers.
7. 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