Social Dynamics: The Voice of Power and Influence

Alex Pentland
Massachusetts Institute of Technology
Room E15-387, 20 Ames St, Cambridge MA 02139,USA
pentland@mit.edu

Abstract
People communicate and negotiate their social position through the *dynamics* of their conversational turn-taking. This channel of communication seems to be an important determinant of social control and information flow, and has a strong effect on social interactions ranging from starting salary to determining who is a 'connector' in a social network.

Keywords: Social dominance, gender differences, negotiation, information diffusion, human dynamics, prosody, conversational turn-taking, Coupled Hidden Markov Models (CHMM), influence model.

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 social communications are conscious and are often manipulated. However might there also be effective social communications mechanisms that are largely unconscious?

Human communication has been studied at many levels --- phonemes, words, phrases, dialogs --- and both semantic structure and prosodic structure has been analyzed. Long-term, multi-utterance prosodic structure, however, has had relatively less attention from both the machine understanding and linguistic communities [1]. In the popular culture, however, this longer-term structure is considered important and is related to social qualities such as `being empathetic' or `being in charge' [2].

When two people are interacting, the dynamics of their conversational turn-taking --- how often they talk and how long they talk --- must adapt to each other and the resulting turn-taking behavior will be a blend of the participants' typical individual behaviors. 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' dynamics with lots of quick back-and-forth turn taking. The ability to set conversational dynamics in this manner is popularly associated with good social skills, and attributed to skilled salespeople and those seeking favors from higher status individuals [2].

Another intuition is that in many conversations the interaction is `driven' by a question-and-answer interaction. This can be a `teacher' using the Socratic method, an administrator seeking a full report of some situation, or a storyteller interacting with her audience. In such cases there are distinct social roles, and we would expect a substantial and directional propagation of information. Thus we might be able to measure propagation of information by measuring who is `driving' the conversation.

In this paper I propose that the dynamic structure of conversational turn-taking is an important channel for communicating social information, and one that has been largely neglected by the machine understanding community. Information such as social dominance and the social importance of new information seems to be communicated in this implicit, largely unconscious manner.

Conversational dynamics has a strong effect on social interactions ranging from starting salary to determining who is a `connector' in a social network. In addition, people's social status appears to be at least partly negotiated via this same mechanism.

2. Measuring Conversational Dynamics
We begin by extracting features from the speech stream of each person. Basic speech features we currently extract include voiced vs non-voiced, the frequency of the fundamental formant, the spectral entropy of voiced segments. From these basic features we use a multi-level HMM structure to classify speaking vs not speaking [3], measure the variance of the formant frequency and spectral entropy, and estimate the speaking rate (as average number of voiced segments per second of speaking).

Once we have characterized a conversation in terms of speech features, the next challenge is to build a computational model that can be used to predict the longer-term dynamics of the individuals and their interactions. The learnability and interpretability of a model greatly depends on its parameterization. The requirement for a minimal parameterization has motivated our development of Coupled Hidden Markov Models (CHMMs) to describe interactions between two people, where the interaction parameters are limited to the inner products of the individual Markov chains [4]. This allows a simple parameterization in terms of the "influence" each person has on the other. The two-person CHMM model has more recently been generalized to handle interactions between many people by use of the so-called "Influence Model'', which describes the connections between many Markov chains as a network of convex combinations of the chains [5,6].

The graphical model for the influence model is identical to that of the generalized *N*-chain coupled HMM, but there is one very important simplification. Instead of keeping the entire $P(S_t^i \mid S_{t-1}^1, ..., S_{t-1}^N)$, we only keep $P(S_t^i \mid S_{t-1}^j)$ and approximate the former with $P(S_t^i \mid S_{t-1}^1, ..., S_{t-1}^N) = \sum_j \alpha_{ij} P(S_t^i \mid S_{t-1}^j)$. In other words, we form our probability for the next state by taking a convex combination of the pair wise conditional probabilities for our next state given our previous state and the neighbors' previous state. As a result, we only have N QxQ tables and N α parameters per chain, resulting in a total of $NQ^2 + N^2$ transition parameters - far fewer parameters than any of the above models.

This simplification seems reasonable for the domain of human interactions and potentially for many other domains. Furthermore, it gives us a small set of interpretable parameters, the α values, which summarize the interactions between the chains. By estimating these parameters, we can gain an understanding of how much the chains influence each other.

To estimate the influence parameters we maximize the per-chain likelihood by gradient ascent using the derivative w.r.t. α_{ij}:

$$\frac{\partial}{\partial \alpha_{ij}}(.) = \sum_t \frac{P(S_t^i \mid S_{t-1}^j)}{\sum_k \alpha_{ik} P(S_t^i \mid S_{t-1}^k)} = \sum_t \frac{P(S_t^i \mid S_{t-1}^j)}{\langle \alpha_i, B_t^i \rangle}$$

where $\alpha = \begin{bmatrix} \alpha_{i0} \\ \vdots \\ \alpha_{iN} \end{bmatrix}$ and $B_t^i = \begin{bmatrix} P(S_t^i \mid S_{t-1}^0) \\ \vdots \\ P(S_t^i \mid S_{t-1}^N) \end{bmatrix}$.

Typically no more than 20 iterations are required to ensure convergence [6].

Using this model we will next analyze some of the dynamics of the interactions, focusing primarily on the turn-taking patterns of individuals and how they differ from each other as a function of social status, gender, social role, and similar factors [7,8].

We start by defining a "turn". Every one-tenth of a second we estimate how much time each of the participants speaks, and the participant who has the highest fraction of speaking time is considered to hold the "turn" for that time unit. For a given interaction, we can then quantify how participants' transition between turns by fitting a two-state HMM to their observed frequency of transitioning between speaking and not-speaking. Once we have estimated the individual turn-taking transition probabilities we can measure the coupling or influence parameters between the two participants. We use turn-taking behavior within one-minute segments of the conversations to estimate the interaction parameters.

The distribution of utterance length that we have observed is bimodal. Sentences and sentence fragments typically occur at several-second and longer time scales, and are modeled as described above. At time scales less than one second there are interjections and confirmations typically consisting of a single word. Within conversations there are frequently quick back-and-forth exchanges consisting solely of these short utterances. It is difficult to reliably estimate influence parameters or other statistics for these very short utterances, and so we characterize these interactions by simply measuring their frequency.

3. Experiment: Negotiation Dynamics

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, Eagle, and Martin [8] 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?

3.1 Results

In this experiment the hypothesis was that negotiation participants who establish a favorable conversational dynamics at the beginning of the negotiation would do better than those who are more passive.

To test this hypothesis the conversational dynamics were averaged over the first five minutes of the negotiation and compared to the objective outcome and to the subjective responses. A multilinear regression between the vocal dynamics and the overall objective outcome showed a significant predictive relationship across all participants ($r^2 = 0.31, p = 0.0089$). That is, we could do a good job of predicting your salary by measuring the conversational dynamics at the start of the negotiation.

The relationship between turn-taking dynamics and outcome differed for high- and low-status participants. For VPs, the extent to which they `set the tone' of the conversation by dominating the turn taking dynamics predicted almost half of their variation in outcome ( $r^2 = 0.47$ ). For MMs, the extent to which they established a `friendly' dynamics with lots of quick back-and-forth turn taking predicted almost a third of the variation in their objective outcome ($r^2 = 0.32$). Both relationships were similar for male and female dyads, but the effect was somewhat stronger in the female dyads.

The dynamics features were also significantly correlated with the subjective ``impression-on-partner'' rating and the ``did-your-partner-let-you-win'' rating. The amount of quick back-and-forth turn taking was significantly correlated with the extent to which participants said they were seeking to avoid disagreements.

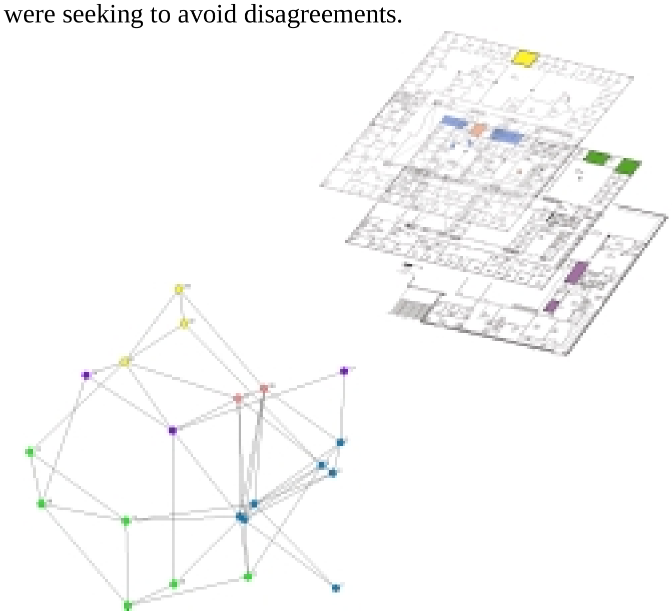

Figure 1: Physical locations of the participating research groups and a visualization of their interaction patterns.

4. Experiment: Research Lab Dynamics

In Choudhury and Pentland [7] we collected data from 23 subjects from 4 different research groups over a period of 11 days (over two full work weeks and 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 the device on them for six hours a day (11AM –5PM) while they are on the MIT campus. The almost 1700 hours of data was automatically analyzed to detect the pair-wise conversations, and this was used to analyze the actual communication patterns that occur within the community.

An intuitive picture of this datasets' structure can be obtained by visualizing the network diagram via MDS (multi-dimensional scaling) on the geodesic distances. Geodesic distance is used as the distance metric because it corresponds directly with the number of `degrees of separation' within the social network. The link structure for the nodes is calculated by thresholding the number of interactions, and the distances between a pair of nodes is the length of the shortest path connecting the two nodes. Figure 1 shows both the physical layout of the participating research groups, and the network visualization obtained via MDS. The nodes are colored according to physical closeness of office location. People whose offices are in the same general space are close in the communication space as well, as expected from previous studies of group interaction patterns.

4.1 Results

In this experiment we wanted to investigate the role of conversational dynamics in everyday conversations. Our first hypothesis was that people had characteristic turn-taking dynamics. Figure 2 shows the output of multidimensional scaling of the subjects' turn-taking transition probabilities using a Euclidean distance metric. As this figure illustrates, we found that individuals have distinctive turn-taking styles and that these turn-taking patterns are not just a noisy variation of the same average style (p < 0.001).

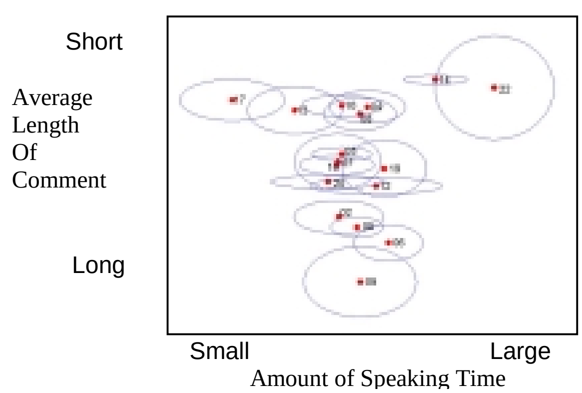

Figure 2: Subjects have chacteristic turn-taking dynamics.

In this figure the women subjects' data lies along the top of the diagram: on average they spoke as much as the men, but they did so in much shorter chunks. The exception to this rule was the single female professor, whose speaking style was near the median for the male group.

Our second hypothesis was that many conversations are `driven' by question-and-answer interactions, and therefore we might be able to measure propagation of information by measuring the influence parameters.

To test this hypothesis Choudhury [9] in her PhD thesis compared the measured influence parameters 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 [10]. The correlation value between this centrality measure and the influence parameter was 0.90 (p-value < 0.0004, rank correlation 0.92). This finding strongly supports the hypothesis that the influence parameters are a good measure of information propagation within organizations.

5. Discussion

In this paper I have argued that information such as social status and the social importance of new information is communicated by conversational dynamics. In addition, people's social status appears to be negotiated via this same mechanism. The relative importance of conversational dynamics versus other well-known channels of social communication has not been established, however the magnitude of the effects we have found are surprisingly large.

In our negotiation experiment we showed that interaction dynamics established during the first five minutes of a negotiation account for more than 30% of the variation in objective outcome, and that the `winning' strategy is different for high-status vs low-status participants. High-status participants do better by dominating the dynamics and forcing a slower-paced interaction, while low-status participants do better by fostering more rapid, interactive dynamics.

A similar variation in conversational dynamics was seen between genders in our research laboratory experiment. The women's dynamics were more rapid and interactive, while the males favored longer, less frequent utterances. The single exception was the one high-status women in the data set, whose dynamics were similar to the median male dynamics.

We also found that in a research laboratory environment peoples' conversational dynamics mirrored the information flow within the social network. The more a person was a `connector' within the social network, the more they dominated the conversational dynamics.

These findings support the view that peoples' conversational dynamics are a learned behavior, and are different depending on social context. Female subjects, for instance, adopted one style of dynamics in the context of the research laboratory, a similar dynamics in the low-status negotiation role, but a very different dynamics in the high-status negotiation and high-connector roles. Male subjects' dynamics were similar in the laboratory and the high-status negotiation role, but very different in the low-status or low-connector roles.

A persons' conversational dynamics is also a function of their social network, requiring consensus among other community members. You cannot have `high status' dynamics within the community without everyone else cooperating to let you behave that way. This suggests that there is a continual, implicit `negotiation' between members of a social network that establishes each individual's appropriate conversational dynamics behavior.

Finally, it is interesting that people in these experiments were only vaguely aware that there were differences in conversational dynamics. They were completely unaware of either the relationship between `connectedness' and `high status' dynamics, or the relationship between dynamics style and negotiation outcome. The conversational dynamics that people display seems to be learned unconsciously.

Are conversational dynamics 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 [1]. In contrast to this type of prosody the conversational dynamics measured in these experiments occur at time scales that are far too long to be related to individual words or phrases.

Prosody is also studied as a method of signaling emotional state [1]. In our experiments, however, affect was not a central variable nor was it uniform across time or subjects. It is therefore seems unlikely that conversational dynamics are directly related to subjects' affect.

Conversational dynamics instead seem to communicate and be involved in mediating social variables such as status or group interest, and arise from the interaction of two or more people rather than being a property of a single speaker. Semantics, utterance prosody, and signals of affect are important because they can be used to alter turn-taking behavior and utterance length, and thus alter the dynamics.

This raises the important question is whether conversational dynamics are a separate, independent communication channel…a sort of vocal body language…or arise only as a consequence of the semantic content of the conversation. I believe that the evidence is that this is a separate communication channel. The males and females in the research laboratory experiment perform the same sort of research work, have the same job responsibilities, and know mostly the same people, yet have very different conversational dynamics. The `connectors' also perform the same sort of research and know mostly the same people, and the `novel' information they transmit is presumably repeated and discussed by everyone in the experiment, yet their dynamics vary systematically with the centrality of their connectivity. The factual material discussed in the negotiation experiment is very limited and known to both parties, yet conversational style varied by status. In each example the claim that dynamics is a consequence of content seems weak.

Finally, It is interesting to speculate about what might happen if people were made more aware of their dynamics, through the use of a small wearable meter that could provide them 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. Handel, Stephen, (1989) Listening: an introduction to the perception of auditory events, Stephen Handel , Cambridge: MIT Press

2. Gladwell, M., (2000) *The Tipping Point: How little things make can make a big difference*. New York: Little Brown.

3. Basu, S., (2001) *Conversation Scene Analysis*, Ph.D Thesis in Dept. of Electrical Engineering and Computer scienc, MIT. Advisor: A. Pentland.

4. Oliver, N., Rosario, B., and Pentland, A., (2000*) A Bayesian Computer Vision System for Modeling Human Interaction,* IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8), pp. 831-843

5. Asavathiratham, C., (2002) *The Influence Model: A Tractable Representation for the Dynamics of Networked Markov Chains*, Ph.D. Thesis in Dept. of EECS., MIT. Advisor: G. Verghese.

6. Choudhury, T., Basu, S., Clarkson, B., and Pentland, A., (2003) "*Learning Communities: Connectivity and Dynamics of Interactive Agents*,'' Proc. Int'l Joint Conf. Neural Networks, Special Session on Autonomous Mental Development, IEEE Press Oct. 2003, pp: 2797-2802.

7. Choudhury, T., and Pentland, A., (2003) *"Sensing and Modeling Human Networks Using the Sociometer,"* 7th IEEE In'tl Symposium on Wearable Computing, IEEE Press, Oct. 2003, pp. 216-222.

8. Pentland, A., Curhan, J., Eagle, N., Martin, M., (2004) *``Money and Influence,''* MIT Media Lab Technical Report 577

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

10. Wasserman, S. and K. Faust, *Social Network Analysis Methods and Applications*. 1994: Cambridge University Press