Signals and Speech

Alex Pentland

Human Dynamics Group, The Media Laboratory, MIT
Pentland@mit.edu

Abstract

Our ancestors were a social species before they had language, and like other apes likely used a system of signals (dominance, interest, warning, etc) to coordinate their activities. Speech appears to have evolved as an elaboration of this earlier communication framework, with signaling remaining as a major functional structure in both dyadic and group interactions.

1. Introduction

All social species have developed a communication system, albeit based on various sorts of signals, in order to coordinate behavior between individuals. Typically the signals include gestures, expressions or calls. A special sort of signal is called `honest' not only because it provides usually trustworthy cues, but because it triggers changes in individuals receiving signals that are advantageous to the individual who sends them. Mood contagion through heightened activity levels and mimicry are perhaps the two best-known examples of such honest signals [1].

Our human ancestors used such signals to coordinate their actions long before sophisticated human language evolved. By better understanding the influence of social signals, we can shed light on the structure and function of modern social interactions. For instance, honest signals can increase the energy level within a hunting team or, for that matter a creative team, through contagious excitement. They create a more cohesive groups by increasing empathy and trust through mimicry signaling.

A relative newcomer in hominid evolution, language was likely layered upon older primate signaling mechanisms that used social strategies to find resources, make decisions, and coordinate group action. When we watch a conversation between two people and carefully measure the timing, energy and variability of the interaction, we find several examples of honest signals. My research group concentrates on four components of this human signaling. *Mimicry* is the reflexive copying of one person by another during a conversation, resulting in an unconscious back-and-forth trading of smiles, interjections and head nodding. *Activity* indicates interest and excitement, familiar to us from the connection between excitement and the activity level of children. *Influence* of one person over another can be measured by the extent to which one person causes the other person's pattern of speaking to match theirs. And *consistency*, or fluidity, of speech and movement is perceived as a marker of expertise.

It is important to note that these signals are not instantaneous quantities, but rather they are statistical averages over 30 seconds or longer. As a consequence they are not directly related to linguistic structure. They are similar to human `thin slice' social perception, the ability that we have to `read' compatibility, interest, and other social relationships from a few dozen seconds of observing another person engaging in a social interaction. Indeed, it seems likely that social signals are key mechanism by which thin slice social perception functions.

To measure these largely unconscious social signals, we have developed algorithms to parse an audio stream into utterances and backchannel activity, and then methods to characterize variations in pitch, pauses, and amplitude, and finally models for the statistical dependencies (`influence') between utterance timing [1]. The result is a robust system that can assess activity, influence, mimicry, and consistency, typically averaged over 30 second intervals.

Each of these honest signals has its roots in our brain structure. For instance, mimicry is believed to be related to cortical mirror neurons, a distributed brain structure that seems to be unique to primates and is especially prominent in humans. Mirror neurons react to other people's actions and provide a direct feedback channel between people. Newborns, for instance, mimic their parents' facial movements despite their general lack of coordination. Similarly, our activity level is related to the state of our autonomic nervous system, an extremely old neural structure. Whenever we need to react more vigorously—say in fight-or-flight situations or when sexually aroused—this system increases our animation levels. On the other hand, we tend to be listless and less reactive when our autonomic nervous system is blunted, as during clinical depression.

The relationship between mental illness and signaling behavior is tight enough that it is now being used on commercial basis to screen for depression. Signaling behavior is also being used to assess patient engagement and intent, and works well for both native and non-native speakers. For more detail see http://www.cogitohealth.com .

We typically collect data using custom-designed electronic `sociometric badges' which are able to record both audio features and coarse-grain body language. This sociometric badge system is able to accurately measure patterns of communication between up to several hundred people over periods of a month or more. The badges are available on a not-for-profit basis for scientific experiments, and are being used commercially to assess patterns of communication within organizations in order to improve productivity, reduce employee stress, and provide a scientific basis for space utilization and organizational re-engineering. For more detail see http://www.sociometricsolutions.com .

2. Influential communication

To illustrate the role of unconscious social signaling, consider our study on business-plan pitches. In that study, a group of rising-star business executives gathered at MIT for an important task. Each executive presented a business plan to the group and the group, then chose the best ideas. The executives wore our badges, which captured their styles of social signaling. By analyzing that signaling, we were able to predict with a high degree of accuracy which business plans the executives would choose. Our executives, it seems, were busy measuring the social content of the presentations, quite apart from the spoken, informational part.

To understand why this makes sense, consider the situation in more detail. Imagine you are listening to a business plan pitch on an unfamiliar topic. Although you don't know much about the subject, the speaker's presentation is fluid and practiced. Also, the speaker is noticeably energetic and clearly excited. Your unconscious judgment seems to conclude that you may not know much about this topic, but the presenter is clearly expert and she is excited … so I guess it must be a good plan.

By examining the back and forth of signaling behavior in dyads and small groups—paying no attention to words or the identity of individuals—we can also accurately predict outcomes of speed-dating encounters, job interviews, even salary negotiation outcomes to within $1,000. In a wide variety of situations ranging from business management to first dates to the effects of political opinion, we find that roughly 40 percent of variation in outcomes can be attributed to signaling-based models of social information processing. This allows us to predict outcomes of dyadic interactions with about 80 percent accuracy without taking account of the words or the properties of the individuals. In other words, the estimated influence of these social signals is similar to the estimated influence of genetic makeup on individual behavior [1].

3. Roles, signals and speech

How do these signaling mechanisms interact with language? Evolution rarely discards successful working parts. It generally either builds additional structures while retaining the old capabilities or subsumes old structures as elements of the new. When our language capabilities began to evolve, our existing signaling mechanisms most likely were incorporated into the new design. The question, then, is how modern human interactions have been shaped by our ancient signaling mechanisms, and to what extent do these mechanisms still govern our lives?

Perhaps the simplest hypothesis is that simple elements of language --- deixis, signs --- first evolved in order to elaborate and expand our pre-existing signaling capabilities. In this case we might expect that the social and informational roles of speech will act in parallel with signaling. Following this line of reasoning we can even hypothesize that the information contained in signaling will be redundant for information about social context, such as who is the protagonist and who the supporter, but not for fine-grain content information.

A method of testing this hypothesis is to observe groups of subjects solving tasks like the Mission Survival Task [2], then have trained observers annotate the social and task roles for each individual using a content-classification framework such as the Bales Interaction Process Analysis method, and finally ask if the roles are equally well characterized by both their content and the accompanying signaling. If we find that analysis of content and context produces similar role classifications as analysis of signaling, then there is a strong argument that speech acts in parallel with the older signaling framework.

This experiment was conducted using the Mission Survival Corpus developed in [3], a multimodal annotated corpus based on the audio and the video recordings of eight meetings that took place in a lab setting appropriately equipped with cameras and microphones. Each meeting consisted of four people engaged in the solution of the mission survival task. This task is frequently used in experimental and social psychology to elicit decision making processes in small groups, and is considered a good probe of group decision making processes.

This corpus was annotated by trained human observers who marked every individual at each second of time in terms of the social roles of protagonist, supporter, attacker, and neutral, as well as the task roles of information giver, seeker, orienteer, and neutral [4]. Audio signaling was then computed by taking the pattern of speaking amplitude for all the participants, and broken into activity, influence, mimicry, and consistency measures. Video signaling was computed using the average amount of motion energy associated with each person's body and hands, as measured by automatically tracking skin region features [5]. The video features are thus quite analogous to the audio features. No measurements of facial motion or eye gaze were used.

From these base features we construct an influence model to classify the social and task roles of each speaker. The influence model builds Hidden Markov Models for each speaker, where the hidden states can be the social or task roles, and then the conditional dependencies between the states of all of the speakers are computed [6]. See http://vismod.media.mit.edu/vismod/demos/influence-model/index.html for code and additional examples.

Average accuracy for major roles	Audio Signals Only	Visual Signals Only	Both Audio and Visual Signals
Social Roles: Protagonist, Supporter, Neutral	0.77	0.72	0.78
Task Roles: Information Giver, Seeker, Orienteer, Neutral	0.71	0.68	0.71

Table 1: *Accuracy at classifying social and task roles using audio signaling, visual signaling, and combined audio-visual signaling. These accuracies are similar to the inter-rater reliability of the human rater..*

Table 1 shows the accuracies obtained when classifying social and task roles using audio signaling, visual signaling, and combined audio-visual signaling. It can be seen that the audio and visual channels are largely redundant. Most importantly, these accuracies are similar to human inter-rater accuracies obtained when classifying the social and task roles using speech content and context.

This experiment provides evidence that social and task roles as defined by semantic content and context are very similar to the roles as defined by unconscious social signaling. This supports the view that the speech and semantic structures that characterize these roles operate in parallel with social signaling mechanisms.

From this result it might expected that this same process could be used to predict turn-taking order within the group. In a separate experiment we found that by combining activity level data with influence modeling of the speaking/not speaking state for each participant we could indeed predict turn-taking order with an accuracy similar to our ability to classify social and task role [7]. This level of prediction accuracy is again similar to human levels of performance at this task.

4. Signals and Performance

A second important question to ask is how social signaling mechanisms are connected to group problem solving. Our recent *Science* paper [8] we examined group performance across a suite of tasks and found a `collective intelligence factor,' analogous to the individual intelligence factor *g*, that is correlated with the participants `social intelligence' and their pattern of interaction, but uncorrelated with the individual members' intelligence scores. This suggests that the social signaling may play a significant role in group performance.

How might social signals be used in problem solving? Early human groups could likely pool information and solve problems, much as ape groups are observed to do today. It is easy to accomplish these tasks using social signaling. Participants introduce a suggested course of action or a fact (perhaps using deixis or signs), and then other participants respond by signaling their level of interest or approval. Finally, group members "add up" the signaling to pick the option with the most positive signaling.

This method of decision-making doesn't require language. In order to pick the winning course of action, each participant must only signal to the rest of the group how interested they are in each alternative and then be able to read the group's combined signaling. Animal behavior research supports the idea that this is what both bees and ape troops do when deciding about group movements. It also is similar to the initial reaction signaling seen in business meetings. The back-channel "ums" and "oks" that greet new ideas in today's conference rooms suggest that our modern decision-making processes may preserve and leverage these ancient mechanisms.

To investigate the hypothesis that modern group decision making retains this basic signaling structure we can revisit our *Science* paper data, where the performance of a large number of groups was carefully measured, and ask if the signaling behavior alone – without considering content, group membership, or other semantic features – accurately predicts the performance of the group. If it does, then this is evidence that the signaling behavior of the group is parallel to the functional semantics of the problem solving process, e.g., the structure of signaling parallels the structure of the problem solving process.

Audio data was recorded for 51 of the groups (from a total of 190 in the original dataset) and four of the tasks reported in the *Science* paper using sociometric badges [9]. The tasks that were monitored with the badges included a brainstorming activity, an IQ test taken as a group, a group judgment task, and a shopping planning task. The tasks were conducted in two blocks, which we name MCI1 (n=84 group scores) and MCI2 (n=120 group scores). Within the MCI1 block all groups had three participants, while within the MCI2 block group size ranged from two to five. Both performance scores and social signaling measures were z-scored within each task to allow comparison across tasks, and signaling features were computed from the audio. In this experiment the average activity level for the group was estimated by the overall frequency of utterances, the frequency of backchannel reactions was taken as a proxy for the average group mimicry level, and within-group variation in influence between participants was estimated by computing the variation in turn-taking frequency across participants [10]. Consistency was not a significant factor.

Figure 1 shows the relationship between group performance and the pattern (not content) of speech (adj r=0.72 for MCI1 and r=0.67 for MCI2, both p<0.001). From these data we can define three characteristics that are typical of the highest performing groups: (1) many very short contributions, (2) frequent `back channel' statements of validation (comments such as `good,' `no!,' etc.); and (3) similar levels of turn-taking among participants.

These data show that objective problem solving performance across a reasonable range of tasks and group sizes can be surprisingly accurately measured without reference to semantic content or context. This supports the argument that the functional structure of the signaling behavior parallels the functional structure of the semantic problem solving processes.

5. Conclusion

I have argued that speech evolved `on top' of older signaling mechanisms, and that as a consequence the semantic structure of speech as observed in group interaction parallels the signaling structure. The evidence is that for group social roles, informational roles, turn-taking patterns, and objective problem solving performance the structure observed in the signaling behavior closely parallels the structure observed by analysis of the speech semantics including context.

It appears that this largely unconscious social signaling behavior is a `para-semantic' communication channel, analogous to `para-lingustic' markers of syntax. This connection between ancient signaling mechanisms and our more recent speech capabilities make it likely that speech developed as an elaboration of signaling behaviors, and then that signaling and speech continued to co-evolve.

As a practical matter one can use this parallelism to improve the performance of groups. We have developed methods of providing real-time feedback in order to promote the signaling behaviors seen in the higher performance groups. The tool we have developed – called the meeting mediator – has two components: a sociometric badge to capture signaling behaviors and a mobile telephone to visualize the group's interactions [11]. The phone visualization provides real-time feedback to encourage balanced participation and high interactivity in the group, and has been found to be to be particularly effective for geographically distributed groups. In group brainstorming and problem solving experiments, this feedback increases average activity levels and the amount of backchannel signaling, and reduces variation in turn-taking. As would be expected from the results of the last experiment, this results in an overall increase in group performance [12].

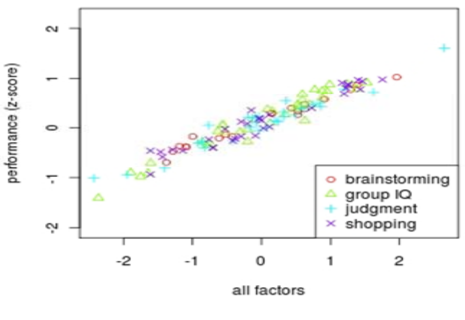

Figure 1: *Group performance vs signaling features for MCI 1 data.*

6. References

[1] Pentland, *Honest Signals: How They Shape Our World,* MIT Press, Cambridge, MA, 2008.

[2] J. Hall and W. H. Watson, `The effects of a normative intervention on group decision-making performance," Human Relations, vol. 23, no. 4, pp. 299–317, 1970

[3] R. F. Bales, Interaction Process Analysis: a Method for the Study of Small Groups. Addison-Wesley Press, 1951

[4] F. Pianesi, M. Zancanaro, B. Lepri, and A. Cappelletti, "A multimodal annotated corpus of concensus decision making meetings," Language Resources and Evaluation, vol. 41, no. 3-4, pp. 409–429, 2008

[5] P. Chippendale, "Towards automatic body language annotation," in Proc 7th IEEE In't Conf. Automatic Face and Gesture Recognition, Washington DC, 2006, ISBN:0-7695-2503-2

[6] W. Dong, B. Lepri, A. Cappelletti, A. Pentland, F. Pianesi, and M. Zancanaro, "Using the influence model to recognize functional roles in meetings," in ICMI, 2007, pp. 271–278

[7] W. Pan, M. Cebrian, W. Dong, T. Kim, and A. Pentland. Modeling Dynamical Influence in Human Interaction Patterns. *Arxiv preprint arXiv:1009.0240*, 2010.

[8] A.W. Woolley, C.F. Chabris, A. Pentland, N. Hashmi, and T.W. Malone, *Science* 330(6004) 686-688. (2010).

[9] Daniel Olguín Olguín, Benjamin Waber, Taemie Kim, Akshay Mohan, Koji Ara, and Alex (Sandy) Pentland. Sensible Organizations: Technology and Methodology for Automatically Measuring Organizational Behavior. IEEE Transactions on Systems, Man, and Cybernetics-B: Cybernetics. Vol. 39, No. 1, pp. 43-55. February, 2009.

[10] Dong, W., and Pentland, A. Quantifying group problem solving using social signal analysis. In Proceeding of International Conference on Multimodal Interaction (ICMI),2010

[11] T. Kim, A. Chang, L. Holland, and A. Pentland, Meeting mediator: enhancing group collaboration using sociometric feedback, In *Proceedings of ACM Conference on Supported Cooperative Work*, San Diego, CA, November 8-12, 2008, pp. 457-466. ISBN: 978-1-60558-007-4

[12]. W. Dong, T. Kim, and A. Pentland, A quantitative analysis of collective creativity in playing 20-questions games. In *Proceedings of the ACM Conference on Creativity and Cognition*, Berkeley, CA, October 27-29, 2009. pp. 365-366