

Towards Socially-Intelligent Wearable Networks

Anmol Madan, Ron Caneel, Alex “Sandy” Pentland

MIT Media Laboratory

20 Ames Street, Cambridge 02139

{anmol,rcaneel,sandy} @ media.mit.edu

Abstract

We propose a wearable system that uses machine perception to quantify a user's social context and propagate this information to others in the user's social network. The social context is evaluated for the user's instantaneous, face-to-face interactions by evaluating proximity, collective speech features, head-movements, and galvanic skin responses. This information is then propagated to others within the user's social or work group who have pre-approved permission to 'patch in' to interesting conversations. We believe that propagation of social context will allow distant users to become better integrated into ongoing projects or discussions, and thus improve distance-separated social interaction, teamwork, and social networking

1. Idea & Motivation

People usually operate in groups, and their needs and goals are often defined by their group context. Most current wearable systems, however, only consider individual preferences and physical context variables such as location. We believe that it is essential to consider the people in the user's face-to-face proximity, as well as the user's social or work network, in order to meaningfully model their needs and expectations.

Quantification of social context may have its largest impact on distance interactions, where social context can be propagated over physical distance in order to better integrate distant participants. One can imagine harnessing the computing and sensing power of today's nearly ubiquitous wearable devices to provide a 'social intelligence' that improves the functioning of distributed work groups.

There has been past work to build mathematical understanding group behavior using speech feature analysis, physiology, activity classification and context mapping. Our feature space currently consists of speech features, galvanic skin responses (GSR) and head-movements, with which we attempt to capture quantitative elements of non-verbal communication like body language, prosody, empathy, and mirroring behavior. We add to this

information about the user's location and their proximity with others.

Choudhury [4,15] was able to show how turn-taking in face-to-face conversation (modeled using a generalized coupled Hidden Markov Model) could be a measure of influence, which was shown to have an extremely high correlation with one measure of the social novelty (and thus presumably the interestingness) of the information being presented. Pentland et al. [14, 16] have attempted to generalize speech features as measures of activity, engagement, stress and mirroring behavior. Our early work with the influence model also indicates how it could be representative of direction of information flow in face-to-face conversation.

GSR has been used as a measurable parameter of a person's internal "state" [9,17], associated with fear, anger, startle response, orienting response and sexual feelings. Studies in psychophysiology have used galvanic skin responses (or electro-dermal activity) as a physiological index for multiple individuals exposed to the same cognitive or emotional events [12]. The role of head nodding and head movement has been acknowledged in conversational feedback for the speaker [11], mirroring behavior and the *chameleon effect* (where people mimic body movements of their conversational partners and that this is reflected more in the behavior of a seemingly more empathetic person) and in *isopraxism* [3]. In [13] we have shown how a combination of speech features and similar GSR epochs and head-movements for a group, could be used to build a *group interest index*.

By capturing all these features, it is possible to incorporate *group social awareness* in wearable devices (based on sensing and real-time machine learning). The propagation of just group location and activity awareness within peers has been shown to enhance distance separated collaboration [10] and has also been used to infer shared interests [19]. We believe that by adding mathematical understanding of behavior based on our feature analysis, we can enhance distance-separated teamwork,

collaboration, decision-making and the sharing of social experiences. *Social intelligence*, captured using wearable devices and propagated over distance will give users some fashion of information they are accustomed to acquiring from body language, speaking style and sub-conscious understanding of behavior, if they were in the same physical-temporal space.

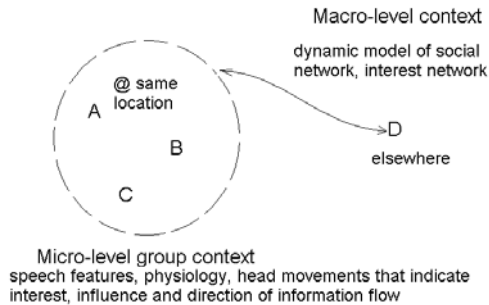


Fig 1: Propagating quantified social context over distance

Some scenarios to illustrate our broad vision of quantitative social intelligence in cell phones and PDAs are,

- Knowing the flow of information and influence for participants in a teleconference negotiation
- Sharing social experiences and common interests with separated friends and family – “Its Saturday night, who’s on my interest network?”
- Next-generation Instant Messaging software where users do not manually type emoticons, but use wireless physiology sensors and speech-analyzing cell phones
- Mobile social software (like Friendster or Orkut) that captures state information – interest, influence, engagement and direction of information flow

2. Proposed Implementation & Early Work

2.1 Hardware and Software Platform

Our hardware and software platform is derived from the Mithril system [6]. The system core is the Zaurus SL6000 Linux PDA, capable of real-time data analysis, peer-to-peer wireless networking, bluetooth wireless sensing, full-duplex audio, local data storage, graphical interaction, and a keyboard/touch-screen input. A sensor hub connects it to an accelerometer and GSR leads, and also supports measurement of heart rate, EKG and other physiological signals. We use the Enchantment whiteboard and signaling software as a lightweight means of routing information transparently across

processes or distributed devices, and propose bluetooth for proximity detection [8].

We have strongly felt the need for wireless sensing and an omnipresent form factor, and hence are integrating a wearable system with bluetooth accelerometers [5], bluetooth GSR sensors [18] and Motorola A760 Linux cell phones. We think that a robust, trendy system that sufficiently captures state information can be made using a Zaurus /cell phone and bluetooth accelerometers and GSR sensors.



Fig 2: The present system featuring a Zaurus PDA, Hoarder sensor board, bluetooth accelerometer (on hat), GSR leads and a microphone

2.2 Feature Analysis

Our proposed system calculates speech features (devoid of any linguistic content) of two different types - individual features and the group features. Individual features include fraction of speaking time, standard deviation of energy, standard deviation of formant frequency, standard deviation of spectral entropy and the voicing rate, and we use them as measures of speaker prosody and emphasis. We use a multi-level HMM structure [1] to classify the voiced/non-voiced sections and hence the speaking /not-speaking regions of the audio stream.

Group features represent the dynamics of interaction between individuals, and include the influence parameters and back-and-forth exchanges. The influence model is a Coupled Hidden Markov Models (CHMMs) designed to describe interactions between two people, where the interaction parameters are limited to the inner products of the individual Markov chains [2,4]. This model allows a simple parameterization in terms of the *influence* each person has on the other. We also label short interjections of a time scale less than 1 second as back-and-forth exchanges (typically single words like ‘okay’, ‘aha’, ‘yup’ or sounds of approval), and consider them as mirroring behavior [14], where one person unconsciously mimics the other’s prosodic pattern.

Head-nodding is detected using a Gaussian mixture model on the frequency domain representation of head movements[7].

Galvanic skin responses are detected using slope detection.

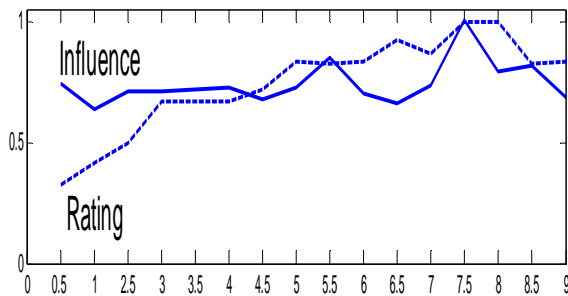


Fig 3: Graph of influence parameters and real-time group-interest feedback on a PDA

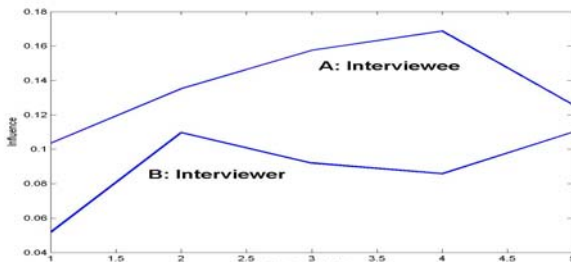


Fig 4: Graph of influence parameters indicating flow of information in an interview situation

2.3 Propagation of Social Context

Our goal is to develop lightweight, unobtrusive wearable systems that can identify face-to-face groups, capture collective social information, extract a meaningful group index, and transmit the group context to remote group members.

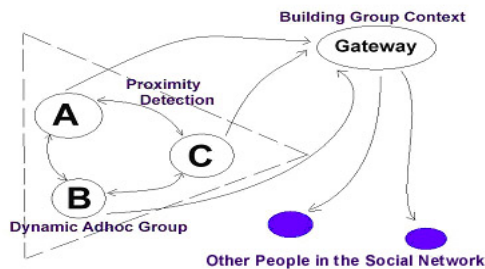


Fig 5: Group formation and information propagation

The gateway in Fig. 5 uses real-time machine learning methods to identify relevant group context changes, such as changes in influence parameters (indicating a change in information flow), or instances of group head-nodding (indicating changes in group interest) [13] or instances of group GSR concordance (indicating interesting group events).

Depending on the identities of the people taking face-to-face, the directionality of the information flow, and the interest levels, the gateway can then notify pre-approved distant users that they might want to 'patch in' to the conversation. Upon receiving such a notification

distance-separated group members can either subscribe to this information over the enchantment whiteboard and begin to receive the raw audio signal plus annotations of the social context, or they can choose to be notified by the system only in case of interesting events, like simultaneous GSR spikes in a number of people, or they can store the audio signal with social annotations for later review.

References

- [1] S. Basu. "Conversation Scene Analysis", in Dept. of Electrical Engineering and Computer science. Doctoral, MIT, 2002
- [2] M. Brand, N. Oliver & A. Pentland, "Coupled hidden Markov models for complex action," IEEE Conference on Computer Vision and Pattern Recognition (CVPR97), 1996.
- [3] T.L. Chartrand, & J.A. Bargh, "The chameleon effect: The perception-behavior link and social interaction." *Journal of Personality and Social Psychology*, 76, 893-910, 1999.
- [4] T. Choudhury, "Sensing and Modeling Human Networks." PhD thesis, Dept of MAS, MIT, 2003.
- [5] Contextual Computing Group, Georgia Tech, GA Bluetooth accelerometer design (<http://www.cc.gatech.edu/ccg/resources/btacc>)
- [6] R.W. DeVaul, M. Sung, J. Gips, S. Pentland, "MIThril 2003: Applications and Architecture", Proc. ISWC '03, White Plains, N.Y., October 2003
- [7] R. W. DeVaul and S. Pentland, "The MIThril Real-Time Context Engine and Activity Classification", Technical Report, MIT Media Lab, 2003
- [8] Eagle, Pentland, "Social Serendipity: Proximity Sensing and Cueing", Media Lab Technical Report 580, May 2004
- [9] G. D. Fuller, "GSR History, & Physiology" taken from Biofeedbacks Methods and Procedures in clinical practice, 1977.
- [10] Holmquist, Falk, "Supporting Group Collaboratin with InterPersonal devices"
- [11] A. Kapoor & R. Picard, "A real-time head and nod shake detector". In IEEE PUI, Orlando, FL, 2001.
- [12] C. Marci, "Psychotherapy Correlates of Empathy, An investigation of the patient-clinical dyad", 2002.
- [13] A. Madan, R. Caneel, S. Pentland, "GroupMedia: Distributed Multimodal Interfaces", ICMI 2004
- [14] Pentland A, Curhan J, Eagle N, Martin M., "Money and Influence", Media Lab Technical Note 577, May 2004 (<http://hd.media.mit.edu>)
- [15] Pentland A, "Social Dynamics: The Voice of Power and Influence", Media Lab Technical Note 579, May 2004 (<http://hd.media.mit.edu>)
- [16] Pentland, A. (2004) *Social Dynamics: Signals and Behavior*, ICDL, San Diego, CA Oct 20-23
- [17] Picard, R.W., "Affective Computing", MIT Press, Cambridge
- [18] Reynolds C. Strauss M. , Picard R.W., Bluetooth GSR sensor design, (<http://arsenal.media.mit.edu/notebook/>)
- [19] Terry, Mynatt, Ryall, Leigh, " Social Net: Using Patterns of Physical Proximity over Time to Infer Shared Interests", 2002