

Why Tag Conversations?

Invited Paper

Dorée Duncan Seligmann Ajita John Shreeharsh Kelkar

Department of Collaborative Applications Research

Avaya Labs

New York, NY; Lincroft, NJ, USA

doree, ajita, skelkar@avaya.com

Abstract— Conversations conducted over media such as face-to-face, email, instant messaging, video and audio provide a rich means of knowledge creation and sharing using collaborative processes. They also provide a source of information that people can tap into for their work and can also be used to drive intelligent communications systems. In this paper, we discuss how a specific property of conversations - persistence - enables conversations to be stored, searched and retrieved in ways that provide contextualized histories of interactions. We have implemented a system to tag and visualize tagged conversations. In this paper we discuss the broader issues of enabling audio conversations to be persistent and show the potential benefits of applying collaborative tagging to conversations.

Keywords- tagging; folksonomy; conversation; enterprise; audio; social network analysis

I. INTRODUCTION

Our research goal is to build intelligent communications solutions that fold communications seamlessly into business processes, enhance a user's communications experience and bring communications to the core of social software. What makes these solutions intelligent is that selection of values for the "who, what, where, when, and how" of communications can be automated and based on communication context ranging from people's identity and roles, presence and availability, to social cohesion and common interests. To this end, we have developed a flexible set of algorithms and models to reason about communications; capabilities to model, capture, store, and manage users' communication context [11][12][25][26]; and a platform [23] on which a wide array of applications execute, ranging from personal communication management to communication enabled business processes.

We are building systems for the enterprise, to improve communications between workers, partners, and companies and their customers. Our focus is on communications and communications related data, a superset to actual content. This includes not only what people are conversing about but who is participating in the conversations, to what degree, when, for how long, on what devices, over which networks, each participant's quality of service (connection integrity, voice quality, SMS latency), the clients and software they are using, location, what they are doing at the same time, how successful or useful the conversation (or segment) is, the type of language used, associations with prior and future conversations, as well as conversational threads across media. Some of this

information is recorded and reported using traditional means, some through analytics and inferences, and some through direct intervention by the users.

We believe that future communications will be driven by communications reasoning and analytics. Our approach is based on what we call a communication feedback loop illustrated in Figure 1. where conversations and collaboration both drive and are derived by communications reasoning; and communications reasoning both drives and is derived by conversations and collaboration.



Figure 1: Communications Feedback Loop

At the core of our approach is users' communication context. This includes values of communication parameters for people (identity, roles, and groups), devices, communication capabilities (modalities, service), activity (past, present, future), environment (location, characterization), social circles, conversational threads and histories, and rules and policies. We use communications reasoning to infer communication context based on the information we can collect about users – ranging from simple presence to predictions of availability or communication response time. This is used in turn to select the people, time, location, devices, modalities, media, networks, interactions, scripts, flow-based applications, sets and meanings of user controls, to filter choices, and disambiguate references. Examples include an application that assembles decision makers to resolve an insurance claim to meet a deadline, a calendar reminder alerting system that selects the best time and mode to deliver a reminder based on how long it will take a person to get to their next appointment, an intelligent call back system that keeps presence information private and yet selects the time and way to connect two people, and a system that authenticates users by requesting they go to the physical location where a certain event occurred.

This paper derives from our work on social software aspects of communications [12][22]: how people connect and find each other, form and maintain communities, and share information and opinions. We wish to enhance these aspects through the use of technology and well-designed tools. We

have developed methods that leverage collaborative tagging to help solve communications problems such as identifying experts within a community [24][27]. We are now investigating the application of collaborative tagging to audio conversations in the enterprise. The system we have implemented is described in [22][36]. In this paper, we focus on the broader issues of applying collaborative tagging of audio conversations and how it achieves the notion of *persistence* for audio conversations for users to access, reflect, and rely upon their interactions in their daily tasks of knowledge creation and sharing.

A. *Motivating Scenario*

In the following scenario we show how, through tagging conversations we can solve abstract communications problems and render communications more efficient: between workers, an enterprise and its customers, between enterprises, and among social communities.

While she was in Korea, Ajita bought a neat new type of wireless peripheral device, CoolDevice, the latest craze there, made by a company called TinyCompany. She knew it was a risk and was unsupported, but she had the idea to use it with a Mainstream multimedia system that she had previously bought through RetailStoreUS. She just could not get the two devices to work together, despite some proficiency in electronics and such.

We begin with the intelligent search for the right person to help Ajita out. This problem is not new. Traditional call centers and skill-based routing systems that we encounter every day utilize a combination of tracking and a sophisticated set of algorithms to select the customer agent to whom to route an incoming customer call. The selection is made by factoring information about the customer and the pool of available agents (from their skill to current workload). On the other end of the spectrum are the popular “ask a question” services that enable users to post a question to an entire social community.

There are no efficient ways to help Ajita today. If she contacts RetailStoreUS, at best they will refer her to customer support for Mainstream or a Mainstream product support page. If she contacts Mainstream, officially they do not support this peripheral, TinyCompany or CoolDevice is not listed in the skills database associated with their technical help people. At best, a willing tech support might try to troubleshoot the problem, but let us assume that without direct knowledge of CoolDevice she will not be able to. Ajita’s problem will not appear on the FAQ page. Contacting chat groups, user forums, and the like will require some effort and few guarantees that helpful suggestions posted in return will be nothing more than a time sink. And as cosmopolitan as Ajita is, she just doesn’t speak or read Korean so getting information from TinyCompany’s website is not apparently feasible.

However, by capturing and mining conversations and collaborative activity, Ajita’s experience can be transformed – she can get help and may even broaden her social circle along the way. At the same time all three companies can benefit through the interaction with Ajita.

Through Ajita’s past interactions with Mainstream and RetailStoreUS, her purchase and technical support histories, her personal tagging activity, postings, comments, and participation in user forums, a profile can be developed that not only captures her various interests but can also be used to infer her level and type of expertise. We can also use that information to characterize her communication skills and to rank the various available ways to communicate with her depending upon task at hand. For example, in the past the only successful technical support sessions have been through phone use (not with email or IM) and with persons of a high level of technical expertise. There is a particular technician with whom she had a uniquely successful session for which there exists some social cohesion. However, for simple product inquiries, IM has been successful.

At the same time the workers at Mainstream can be profiled. Through a combination of collaborative tagging, conversational history, documents, meeting topics, a pool of people at Mainstream that have some knowledge of TinyCompany and CoolDevice, the Mainstream product in question and associated topics can be identified in near-real-time. This reveals that a member of the sales team in Korea has already experimented with the CoolDevice, a division in CTO is considering a device like CoolDevice as a potential addition to the product line, and a small development group has had meetings about potential add-ons of which CoolDevice is one. We also can classify each person’s level and type of expertise. By examining conversational histories we can detect different social groupings and relationships to determine, for example, who is the go-to person for a certain subject. By combining this information with each person’s role and availability, communication skills, language, and more, a selection can be made of not only who should interact with Ajita but when and how.

At the same time the social communities of both RestailStore (in US and abroad) and Mainstream can also be profiled and tracked to provide an able expert to assist Ajita: a fellow customer or someone with whom she has already interacted. As some people find great value in being helpful, providing expertise in various forums, it behooves enterprises to draw these individuals further into their community and leverage their expertise. Thus an enterprise’s pool of subject matter experts can be extended to these individuals. By mining social activities on RetailStore sites a number of people who have bought, retained, tagged, or positively rated the CoolDevice as well as the Mainstream product can be identified. A reward system could be used to motivate the power users, or patient helpers. These social communities are enhanced by bringing people together to help each other, and in turn, by hosting their interactions enterprises can better support and understand its customers.

Through the tagging of conversations, internal operations can be further improved. The use of tagging allows for instantaneous introduction and search of new topics. Field technicians can more easily find who, if anyone, knows how to troubleshoot a new problem; marketing can find who knows about a recent news item. Trends in expertise within the enterprise can be tracked and used to drive staffing and

training. Conversations can be analyzed to detect bottlenecks, or better understand intra-group relationships.

The paper is organized as follows. Section II discusses the role of conversations in the enterprise. Section III discusses a property of conversations – persistence – that enables conversations to be presented as contextualized histories of interactions. Section IV discusses the tagging of audio conversations and conclusions are presented in Section V.

II. CONVERSATIONS

While documents, code, webpages, wikis, blogs, etc. form repositories of knowledge in the enterprise, conversations embody knowledge creation and sharing through interactions in enterprise communities [32]. Conversations are never simple back-and-forth transmissions of information; instead people, with different backgrounds and skills, use all the resources at their disposal, at every moment, to engage in a continuous interpretation and thereby achieve mutual understanding. Moreover this interpretation is influenced by an unarticulated background of assumptions, practices and social relationships that the participants are enmeshed in.

People use several resources for interpretation during conversations. While on the phone, participants take notes, on paper or on their workstations. They use shared visual resources: for instance, if they are editing a document collaboratively, both have the document open, in its printed version or on the computer. They may use the internet to find information about certain things. They use their background – the unarticulated common knowledge that they share – to ground their interpretation and speed up the accomplishment of tasks. While having a conversation over email (or IM), they may use a combination of links or attachments to make an email (or an IM) more expressive. They may use media simultaneously, such as engaging in multiple IM chats simultaneously or using both the phone and IM together. They may use recollections of past related conversations or *histories* to bring someone up-to-date with an issue such as when people forward an entire thread to someone. To be able to use these resources and to be able to draw upon related conversations of oneself and others in a community during an active conversation or a task is a powerful aid to the user. While human beings are naturally ingenious, technology and tools, if done right, can make available even more resources to guide interpretation and help efficient accomplishment of tasks.

While communities in the enterprise have relied on a variety of media such as face-to-face, email, instant message (IM) and voice for engaging in conversations, we focus on audio conversations because:

1. Spoken conversations are among the most used forms of communication in an enterprise, and the richest in terms of collaborative power i.e. spoken language helps us convey context-rich information.
2. Spoken conversations are more interactive than email or IM and are more effective for discussing complex issues in meeting-like situations.

3. While search-and-retrieve techniques for textual communications such as email and IM are available (though tagging is not used much for these media), it is notoriously hard to be able to search and look for relevant pieces of audio and video.
4. A collaborative tagging design for audio conversations may be extended to email and IM.

One way of looking at technology-mediated audio conversations would be to look at their form, which depends on the degree of spontaneity and reciprocity they permit. A rough classification would be: (1) scheduled conversations: Examples of scheduled conversations are conference calls. Here a time between participants is agreed upon – perhaps through email or IM or in their previous conversation. At the selected time, the participants call each other or they log in to a conference bridge. Scheduled conversations such as meetings may be structured according to an agenda. Particularly interesting examples of scheduled “conversations” are broadcasted talks. Here a “talk” or a speech is broadcasted on a conference bridge. Listeners interested in hearing the talk dial in to the bridge and can then listen to it through their phone. They are provided with some visual aid to understand the talk, generally in the form of a presentation, which is placed on the corporate network. (2) Unscheduled conversations: Here the time for a call between the participants is not fixed but is spontaneous. The purpose could be work-related or not, as for example, when one calls up an acquaintance to ask him if he is ready for lunch. Participants use a variety of communications – email, IM, phone conversations – as a work task advances and through these communications try and hash out the numerous contingent issues that arise.

Some work has been done in looking at enhancing phone-based computer-supported cooperative work. [39] discusses using the phone as a computer terminal i.e. to be able to access one’s email, voice-mail, calendar events by dialing to a number (clearly this would be useful when one is traveling). [7] looks at the design of the PC-phone interface, i.e. the desirable properties of an interface for soft-phones like Skype; it serves as a good way to begin the discussion of a PC-based tagging interface for soft-phones. [37][38] discuss the construction of a structured audio interface, i.e. how to present structured information such as emails, phone-messages and calendars, when the only medium of communication is audio. [20][21] discuss “ubiquitous audio” – storing audio conversations unobtrusively and making them available for later reflection and thereby enhancing collaborative work.

III. PERSISTENT CONVERSATIONS

We are interested in a characteristic of conversations called persistence. Persistent conversations are stored and retrievable and enable the enhancement of future communications. We use the term “persistent conversation” in a certain sense. The term “persistent conversation” itself was taken from [14]. One sense, perhaps the primary one, is storage – meaning we would like to be able to store, access and retrieve conversations. This, by itself, is not entirely novel, since emails, IMs, and meeting notes are routinely stored, although voice conversations are generally not. (Tools such as Google Desktop [1] help search

through some of this content.) However the notion of persistence is deeper than this. In our day-to-day life, we are enmeshed in a socio-cultural web of relationships and our primary contact with the world around us is through language. Language describes and constructs the world around us [15]. It allows us to act and enables us to see possibilities for action. When we talk about making computer-mediated conversations persistent, we want to make available new avenues for action. That is by appropriate, storage, retrieval and presentation of traces of past conversations, we hope to enable new ways for acting and doing, and develop more resources which can aid human beings as they go about accomplishing their daily tasks of interpretation. To use a line from [14], persistence enables us to search, browse, replay, annotate, visualize, restructure and recontextualize conversations.

To understand what persistent audio conversations could help us in, we look at two forms of conversation – face-to-face conversations, and textual conversations (i.e. IM and email). In the discussion below, we try and highlight the affordances [17] that our tagging system should aim for.

A. *Face-to-face (“Unmediated”) Conversations*

Daft, Lengel et al. [9][10][41] consider face-to-face (FTF) to be the “richest communication medium” because it is characterized by instant feedback, multiple communication cues, the presence of spoken natural language and the possibility of personal focus. FTF encounters have another important characteristic: they can be scheduled or serendipitous. Examples of scheduled FTF conversations in an enterprise could be staff meetings, project meetings and the like. Examples of serendipitous conversations: one bumps into someone in the corridor, or on the way to the restroom or simply because one passes by someone’s office and catches her eye. Informal conversations have an important role in an organization [28] since they supplement the more formal communication channels and generate new connections among people and their related areas of expertise.

To summarize, FTF conversations and co-location, together have the following affordances: (1) they enable “richer” interactions between people, and (2) they afford spontaneous access to multiple communities of knowledge and practice.

B. *Persistence of emails and IM*

Email and Instant Messaging (IM) are comparatively recent media for communication. Since both are text-based i.e. people communicate by writing (or, in this case, typing) messages to each other, one might assume that they are not rich media in the sense outlined above, lacking instant feedback, communication cues, spoken natural language, and personal focus. However, field studies [13][31] indicate that email and IM are used widely in enterprises, despite, what would seem their inherent deficiencies. In this case, what turns out to be a deficiency i.e. the fact that the messages are textual is also a resource: email and IMs can be stored and retrieved almost effortlessly – in other words email and IM are almost persistent media. In addition, both these media provide for a less interruptive form of communication than face-to-face. We now look at the kinds of actions that these media afford.

Email offers [13]:

- The ability to send visual artifacts (pictures, documents, files), which function as resources for interpretation between participants in a conversation, thereby improving the richness of interaction.
- The persistence of email messages, which means that previous exchanges related to the current topic can be easily retrieved and presented. . Email can hence be used for record-keeping purposes to enhance the execution of future tasks.
- The ability to turn emails themselves to become artifacts for further conversation (“threads” based on subject content). This becomes possible because written text can be easily copy-pasted and commented on. An example is an email that goes: “You said: <text from previous email>. I disagree. Here’s why...” Moreover email clients reproduce entire emails when one clicks on “reply”, thereby providing a history of the conversation up to that point. New topics are easily introduced to other people by “forwarding” an email. Further resources for interpretation are part of the email itself. Google’s email service has developed a very efficient way of grouping emails together as “conversations” by looking at the subject line.

IM services offer [31]:

- A “media space”[18]. This means that IM clients can function as a way to become aware of other people’s activities, both explicitly, by observing, for instance, their status: “busy”, “away”, “available” etc. or implicitly if they take longer time periods to respond.
- The ability to “stick”. IMs are prominently displayed on the computer monitor so an IM can function as a reminder of a task.
- Contextualized histories. When a person is sent a message by an acquaintance, services like Google Talk [2] make available a history of previous conversations with that acquaintance. These histories can be searched and retrieved and serve as resources for interpretation. The integration of IM and email in Google’s services further improves the user’s ability to rely on past conversations to enhance future ones.

Additionally, IMs have many of the affordances of email, including persistence and the ability to support shared visual objects.

These affordances are a reminder of the kind of actions that persistent conversations can provide. They act as a guideline in our investigation of designing tools to enrich the collaborative experience of audio conversations. Our tools for persistent audio conversations should aim (1) for no loss to the actual richness of audio interaction and its ability to convey nuance, (2) to provide access to people and communities (the kind that happens, when one goes to hear a public talk, or bumps into colleague at the water-cooler, or gets introduced to someone by an acquaintance), and which is missing when one talks to a person over a telephone, and which is particularly missing in modern-day distributed enterprises, where people work

together as collaborators but may rarely see their colleagues in person (3) to enrich audio conversations by letting people store, access, and retrieve different resources (visual or otherwise) that can help in better interpretation and faster accomplishment of tasks, (4) to make the participants in a conversation more aware of each other's activities, all of which serve as resources for interpretation. Keeping these in mind, we first discuss some issues with achieving persistence with audio and then how tagging may be used to address them.

C. Issues with Achieving Persistence with Audio

One of the ways to achieve persistence is to extract meta-data from the audio files for search, retrieval and analysis. Audio files (or video files) are signals, as opposed to text, which is symbol (discrete, easily identifiable, etc). Pattern recognition algorithms try to extract a clean "label" from a messy signal, such as extracting the label "cat" from the sound of a cat meowing in an audio file[16]. Sometimes this problem becomes: given a finite set of audio/video signals with their corresponding labels, can we propagate these labels to other un-annotated audio/video files (see, for example,[8][40])? While audio analysis has had some success, for instance in capturing the highlights of sports videos by looking at the features of audio [33] (by detecting cheering, commentator pitch etc), there is no consensus about how this is to be done or what features and classifiers to use in different circumstances. Moreover, pattern recognition algorithms require copious amount of training data and are not real-time.

Another approach may be to transcribe the audio. However, even though, Automatic Speech Recognition is fairly well-established[34][35], recognizing spoken natural language, in real-time, is still difficult. Even a dictation software tool like Dragon [3] needs to be trained for a long time to provide robust transcription for a particular speaker.

Other problems remain. How does one "skim" audio[20][21]? A video, for example, may be represented by a few of its key-frames; a representative key-frame can quickly give a user an idea of the contents of the video. What is a key-frame for audio? Short of playing back the whole (or part) of the audio, how can audio be presented back to the user? Can we represent audio with a visual? Will this visual give us an idea of audio's semantic content?

IV. TAGGING AUDIO CONVERSATIONS

Tagging is hugely popular on the World Wide Web and is already an important part of practices that have grown around it [19]. It is becoming an integral part of the new forms by which people express themselves collaboratively on the web: blogs, video-blogs, wikis, podcasts, etc. Moreover tagging seems to be adaptable to a wide variety of media (audio, video, text, documents, web-pages, etc) and forms (blogs, conversations, wikis, podcasts, etc)[29]. Additionally, tagging has great potential for the enterprise [24][30].

Tagging can offer a respite from the problems outlined for audio in Section III.C. This is because (1) we let users tag audio (or parts of audio), and thereby eliminate the need for pattern recognition algorithms that try to find the "meaning" of audio signals. Moreover, users may tag audio (mostly) for self-

interest: the tags are their way of remembering and note-taking. Signal analysis of the audio may be limited to simpler problems such as identifying speaker turns (for easier skimming of the audio) through silence-detection in individual audio streams (a much simpler problem) and looking for certain keywords in the audio for recommended tags (through Automatic Speech Recognition), both of which have been implemented robustly on commercial products [4] (2) Tagged audio means that search-and-retrieve operations are now performed on text rather than signals, a much simpler and robust process. Moreover, these tags are human-contributed and they are a direct product of the social and institutional contexts in which people tag. (3) Collaborative tagging means that tags contributed by all users can be collated and analyzed – or in other words, the task of "annotating" audio is distributed over a large number of people. (4) Tags offer a mediating layer between actual audio conversations and the people who participate in them (and tag them). Thus, a user may choose to reveal his tags for a conversation but not the actual conversation itself – another user who finds himself interested in the tags for that conversation can then contact the participants with his queries. Tags can thus help in negotiating a trade-off between access and privacy. (5) A variety of social network analysis techniques can be used to analyze tagged conversations [27] which can inform an individual of the different networks and communities of practice that she is embedded in and thereby help her access other people and other communities of shared/relevant interests. The ability to see conversations in the broader contexts of interactions between individuals, groups, and communities will enable an employee to figure out the broader context of his work and thus increase efficiency.

Specifically, tagged audio conversations can be used in several ways:

1. To serve as a memory aid to refer back and then easily retrieve a pertinent segment of a past conversation or to, in realtime, be presented with conversations with the same tags being issued.
2. To allow off-line search of both ongoing and previous conversations
3. To seamlessly share previous conversations with participants of an ongoing conversation – by playing it during the course of a conference call.
4. To associate parts of conversations with notes, reminders, and to-do items; or conversations with each other, for example one conversation may be the subject of another conversation.
5. To find people who have: had conversations with the same groups of people, or on the same topics, used the same tags to describe the same conversations, who have participated in the same (large) conference call.
6. To discover that there are multiple weekly meetings by different groups of people on a particular topic, that one particular person is engaged in most conversations about a certain topic, or that there are sub-networks of people

7. To search and find conversations based on any number of criteria such as all the conversations in which: someone raised their voice, a particular person spoke, that exceeded an hour, that resulted in email, or the language spoken was French.

In the rest of this section, we discuss how users may tag conversations and briefly describe our implemented visualization interface for tagged conversations.

A. *How users tag conversations*

Let us consider some specific scenarios that arise in an enterprise. In the most common one, an employee sits at his work-desk, with a telephone handset and a work-station, both easily accessible. In other scenarios, an employee may have access to only one of these. For instance, an employee who is traveling may not be able to access his computer. Sometimes the computer may be engaged in another task. Moreover, there is already a set of practices associated with the employees in an enterprise. The act of tagging and the use of tagged resources, must, as far as possible, be integrated into these practices and must serve as an easily-available resource that aids the completion of tasks. Thus, the act of tagging must not present too much of a cognitive load, and at the same time, must not disrupt the actual conversation at hand.

Let us consider the ways in which individuals can tag an audio conversation. We have three core requirements: (1) a user must be able to tag any conversation, including one that is currently going on, by logging into a visual web-based interface through a computer. (2) She must be able to tag the current (or previous conversations) through a simplified visual interface on her phone. And finally, (3) when neither of these options is available, tagging must be allowed through a voice interface via the telephone itself.

1) Tagging via a computer monitor.

Tagging via a visual interface on a traditional computer monitor offers the most in terms of action-possibilities. This is because (a) the computer monitor, as a display device, can be used for extraordinarily detailed visualizations of (relevant) conversations and social networks. Moreover, we are fairly well-versed in creating visual interfaces and users are also fairly experienced in using them. There is the additional possibility of using sound and audio (since most computers today are equipped with speakers) to convey more information. (b) The presence of rich input devices such as the mouse and the keyboard are further assets. Keyboards offer a fairly quick and simple way of typing out tags and descriptions. The affordances of using a mouse are an additional advantage. We can use the mouse to select particular parts of conversation for tagging. We can scroll up and down and quickly scan a conversation or social network visualization. If the interfaces are expressive and well-designed, this can be an extremely effective way of tagging parts of conversations. (c) Workstations with network access are themselves rich resources, since the internet, and the enterprise intranet are available if one needs further information and to add related hyperlinked references to a tagged conversation. At the same time, a work-station provides multiple ways of accessing

people – if one comes across someone one needs to contact, one can use email, IM or Skype to contact him.

2) Tagging via the phone with a visual interface.

Corporate IP phones (and now cellular phones) are equipped with internet/intranet access and a visual interface. However, the visual interface of a phone does not afford the kind of visual expressiveness that a computer monitor allows. Thus the visualizations must be smaller in scale and adapted to the phone display. Also, the phone does not allow the kind of expressive input devices like the mouse or the keyboard. Instead, the input would be through audio, buttons and the keypad. However the fact that the telephone keypad is less expressive and cumbersome to use means that tagging recommendations have a greater role to play in this case. Tagging recommendations suggest to the user what the tags to use by analyzing the community's activity and/or by searching for keywords in the audio.

The visual interface on the phone may be used as a channel for mutual awareness between parties in a conversation. This could function as a critical resource that contributes to mutual understanding. For instance, when one talks to another person over the phone, our only contact with them is through the voice-channel. What the visual interface could do is make us more aware of the other person's actions that he may not always communicate to us through his voice. For example, if a user Ed tags a part of an active (current) conversation "CoolDevice", it could immediately appear on the visual interface of Ajita's phones; they could choose to ignore it, (or not notice it) or make it the point of discussion. This notion of real-time mutual awareness is somewhat akin to the way IM chat clients help us understand someone's presence and availability through a variety of cues, both explicit (like online status) or implicit (like someone's lack of response).

We discuss how the phone visual interface may be used for tagging the two types of conversations:

1. An active conversation: As a conversation is going on, the phone visual interface can constantly update itself – providing recommendations for tags and more awareness of the actions of other participants. Users may be able to tag specific parts of the conversation, depending on how expressive the visualization that is presented to them is. The visualizations on the phone-screen can persist for a while after the conversation is over so that the user can tag even after the conversation is finished.
2. Past conversation(s): While the problem of tagging a conversation remains the same as (1), the key question here is: how does the user find the relevant conversations? How will the phone-visual interface support search-and-retrieve mechanisms? How will search results be presented? How can groups of conversations be tagged? These are open questions and we have to resort to specific implementations on specific phone models to answer them.

3) Tagging via the phone through an audio interface

We discuss how the phone without a visual interface may be used for tagging the two types of conversations

1. An active conversation: To tag, a user could press a key on her keypad (say *) and then say explicitly into her phone, something such as “TAG interesting audio conference END-TAG”. By pressing the key before she says so, the system ensures that other users do not hear her words and that the over-all conversation not be interrupted (however they may be made aware of it in other ways). The system will interpret her words as a command to tag the portions of the conversation, i.e. the few turns before and after she says this, with the tags “interesting”, “audio” and “conference” by doing real-time speech recognition. It can also provide feedback to the user as, say, a “whisper” to her phone, with the system’s account of the tags that it received (this will be a good way of accounting for error in the speech recognition). In the case of the other users, it can display the tags on the visual interface of their phones. Discussion between participants about what tags to use is also possible.
2. Past conversation(s): A user may be able to tag a past conversation by dialing into a special number for the tagging service which identifies her (from her caller ID) and through a voice interface asks her if she wants to tag a recent conversation. It can then ask her to say each tag clearly and repeat it back to her. Note that in this case, it is hard to tag parts of the conversation. Tagging conversations older than the recent ones would be harder since the conversation to be tagged has to be identified. This would raise other issues: how would the user query for the conversation she wants to tag? How would the results of the search be presented to her via audio? How would she tag groups of conversations?

B. System for Tagging Audio Conversations

We have developed a web-based visual system for tagging audio conversations. It is described in [22][36] and enables the following:

1. Users dial into a conference bridge and the audio is recorded along with meta-information such as names of users, time of recording etc.
2. Users (participants and non-participants) can tag conversations.
3. Users can share their conversations and/or tags with other users.
4. Users can retrieve conversations by searching on keywords that have been used to tag conversations.
5. The retrieved conversations can be rendered as lists (that are time-ordered or sorted based on queries) or as a social network visualization that presents the inter-relationships between users and tags.

Figure 2 shows a screenshot of the visualization. It shows conversations that have been retrieved through a search box at the top left. The conversations are depicted in the left frame as particles where users (or tags, which are not shown) act as queries to position the conversations depending on

whether the users have participated in/tagged the conversations (or the tags have been associated with the conversations). Links are drawn to between users/tags and conversations to emphasize these associations. The conversations are also depicted in the bottom right frame as a list. The details of a selected conversation are shown in the top right view. This view allows the conversation to be tagged. More details are described in [36].



Figure 2: Visualization of Tagged Conversations

C. Putting collaboratively tagged data to use

We now discuss how collaboratively tagged data can be utilized from an individual’s and the enterprise’s perspectives.

From an individual’s perspective:

(1) At the most basic level, collaborative tagging allows a user better access to his and other users’ conversations. Since knowledge manifests and spreads through conversations, this may greatly boost a user’s productivity.

(2) Tagging is like note-taking, which functions both as a memory-aid and a summary of actual conversations. But because tags (and thus conversations) can be searched and retrieved, our system, in a way, resembles Vannevar Bush’s visionary device [6], the “memex” or “memory extender”, which functions as a hypermedia machine, allowing us free-associative access to the information and memories that saturate our life.

(3) The other important thing is that tagging is collaborative which means that multiple users tag their shared resources. By analyzing this data and looking at the relationship between users and tags, and then presenting this information back to the user, we can allow a user to see his social context, and the communities he is embedded in. This is important since it provides him access to different communities and other users, with whom he may have had no contact, in normal circumstances.

(4) Since the act of tagging can be performed during an active conversation, by making users aware of each other’s activities – mutual awareness – we can help in making their

conversation richer and provide one more resource for better interpretation i.e. the act of tagging and the tags themselves.

From an Enterprise perspective:

Analysis of collaboratively tagged data can reveal interesting patterns for an enterprise:

(1) Important conversations: If conversations are the way knowledge spreads and is manifested in the enterprise, analysis can reveal important things about how knowledge is flowing. For instance, it can help determine whether knowledge is “sticky” or “leaky”[5], and help an enterprise coordinate its knowledge flows.

(2) Expertise communities: If one thinks of the tags as “topics” of conversation, the analysis of people and tags gives an enterprise an idea of expertise communities, i.e. the networks of people that grow around a tag. This, in turn, can help it figure out how to connect different enterprise communities to each other, thereby increasing the efficiency of the enterprise.

(3) Enterprises contain wide networks of people, both formal and informal. By formal networks, we mean the formal hierarchies (manager, employee, assistant manager, etc.) and separations (sales, marketing, research, etc.) within an enterprise. The analysis of communities and networks within the enterprise can determine whether the hierarchies and separations are functioning well. For example, there is a problem in the inter-department channels within an enterprise if the marketing and sales communities do not seem to be communicating.

V. CONCLUSIONS

We have discussed the use of collaborative tagging to make audio conversations persistent which enables users to annotate, search, and retrieve active and past conversations to use as resources for individual and collaborative work. By sharing tags, a community of users can draw upon the collaborative work of others resulting in a greater collective power than what is enabled today. We envision that the popularity of collaborative tagging for entities such as webpages, photographs, blogs, and video can be successfully extended to highly interactive content such as audio conversations to provide contextualized histories that enable increased productivity in collaborative work.

REFERENCES

[1] [Google Desktop] <http://desktop.google.com/>
[2] [Google Talk] <http://www.google.com/talk>.
[3] [Nuance] <http://www.nuance.com/naturallyspeaking/>.
[4] [Avaya Unified Communication Center (UCC)] http://www.avaya.com/gcm/master-usa/en-us/products/offers/ucc_r2.htm.
[5] J. S. Brown and P. Duguid (2000). *The Social Life of Information*. Cambridge, MA, Harvard Business School Press.
[6] V. Bush (1945). *As We May Think*. *The Atlantic Monthly*. 176: 101-108.
[7] J. Cadiz, A. Narin, G. Jancke, et al. (2004). *Exploring PC-Telephone Convergence with the Enhanced Telephony Prototype*, CHI 2004, Vienna, Austria, April 24-29.

[8] E. Chang, K. Goh, G. Sychay, et al. (2003). CBSA: content-based soft annotation for multimodal image retrieval using Bayes point machines. *IEEE Transactions on Circuits and Systems for Video Technology* 13(1): 26-38.
[9] R. Daft, R. H. Lengel and L. K. Trevino (1987). Message Equivocality, Media Selection, and Manager Performance: Implications for Information Systems. *MIS Quarterly* 11(3): 355-366.
[10] R. L. Daft and R. H. Lengel (1986). Organizational Information Requirements, Media Richness and Structural Design. *Management Science* 32(5): 554-571.
[11] M. De Choudhury, H. Sundaram, A. John, and D. Seligmann (2007). Context-Aware Routing of Enterprise User Communications. *percomw*, pp. 39-44, Fifth Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PerComW'07), 2007.
[12] M. De Choudhury, H. Sundaram, A. John, and D. Seligmann (2007). Contextual Prediction of Communication Flow in Social Networks. *Sixth Annual IEEE International Conference on Web Intelligence*, 2007.
[13] N. Ducheneaut and V. Bellotti (2003). Ceci n'est pas un Objet? Talking about Objects in Email. *Human-Computer Interaction* 18: 85-110.
[14] T. Erickson (1999). Persistent Conversation: An Introduction. *Journal of Computer-Mediated Communication: Special Issue on Persistent Conversation* 4(4).
[15] F. Flores and T. Winograd (1986). *Understanding Computers and Cognition: A New Foundation for Design*, Ablex Publishing Corporation, Norwood, New Jersey.
[16] J. Foote (1999). An overview of audio information retrieval. *Multimedia Systems* 7(1): 2-10.
[17] W. W. Gaver (1991). Technology affordances. *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press: 79--84.
[18] W. W. Gaver (1992). The affordances of media spaces for collaboration. *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, ACM Press: 17--24.
[19] S. Golder and B. Huberman (2006). The Structure of Collaborative Tagging Systems. *Journal of Information Science* 32(2): 198-208.
[20] D. Hindus and C. Schmandt (1992). *Ubiquitous Audio: Capturing Spontaneous Collaboration*, CSCW 92,
[21] D. Hindus, C. Schmandt and C. Horner (1993). Capturing, Structuring and Representing Ubiquitous Audio. *ACM Transactions on Information Systems* 11(4): 376-400.
[22] A. John, S. Kelkar., E. Peebles, A. Renduchintala, and D. Seligmann (2007). Collaborative Tagging and Persistent Audio Conversations, *The European Conference on Computer Supported Cooperative Work (ECSCW 2007): Workshop on CSCW and Web 2.0*, Limerick, Ireland, 25th September 2007.
[23] A. John, R. Klemm, A. Mani, D. Seligmann (2006). "Hermes: a platform for context-aware enterprise communication," *Pervasive Computing and Communications Workshops, 2006. PerCom Workshops 2006. Fourth Annual IEEE International Conference on*, vol., no., pp. 6 pp.-, 13-17 March 2006.
[24] A. John and D. Seligmann (2006). *Collaborative Tagging and Expertise in the Enterprise*, WWW 2006: Collaborative Tagging Workshop, Edinburgh, Scotland.
[25] A. Kavimandan, R. Klemm, A. Gokhale, D. Seligmann (2007). Enhancing Enterprise User Productivity with Embedded Context-Aware Voice Applications, *International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM 2007)*, Papeete, French Polynesia, November 2007.
[26] A. Kavimandan, R. Klemm, A. John, D. Seligmann, A. Gokhale (2006). A Client-Side Architecture for Supporting Pervasive Enterprise Communications. In *Proceedings of ICPS'06 : IEEE International Conference on Pervasive Services 2006*, June 26-29, 2006.
[27] S. Kelkar, A. John and D. Seligmann (2007). An Activity-based Perspective of Collaborative Tagging, *International Conference on Weblogs and Social Media*, Boulder, Colorado.
[28] R. E. Kraut, R. S. Fish, R. W. Root, et al. (1990). *Informal Communication in Organizations: Form, Function and Technology*. *Human Reactions to Technology: The Claremont Symposium on*

Applied Social Psychology. I. S. Oskamp and S. Spacapan. Beverly Hills, CA, Sage Publications.

- [29] C. Marlow, M. Naaman, D. Boyd, et al. (2006). Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead., WWW 2006: Collaborative Tagging Workshop, Edinburgh, Scotland.
- [30] D. R. Millen, J. Feinberg and B. Kerr (2006). Dogear: Social bookmarking in the enterprise, SIGCHI conference on Human Factors in computing systems, Montreal, Québec, Canada.
- [31] B. A. Nardi, S. Whittaker and E. Bradner (2000). Interaction and Outeraction: Instant Messaging in Action, Computer-supported Cooperative Work, Philadelphia, PA, December 2-6.
- [32] J. E. Orr (1986). Narratives at Work: Story Telling as Cooperative Diagnostic Activity, Computer-Supported Cooperative Work, Austin, Texas.
- [33] I. Otsuka, R. Radharkishnan, M. Siracusa, et al. (2006). An Enhanced Video Summarization System Using Audio Features for a Personal Video Recorder. *IEEE Transactions on Consumer Electronics* 52,(1): 168-172.
- [34] L. R. Rabiner (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77(2): 257-286.
- [35] L. R. Rabiner and B. H. Juang (1993). *Fundamentals of speech recognition*. Englewood Cliffs, N.J., PTR Prentice Hall.
- [36] A. Renduchintala, S. Kelkar, A. John, and D. Seligmann. (2007). Designing for Persistent Audio Conversations in the Enterprise, The Third Designing for User Experience Conference, Chicago, USA, November 5-7, 2007.
- [37] P. Resnick (1992). HyperVoice: A Phone-based CSCW platform, CSCW 92.
- [38] P. Resnick (1993). Phone-Based CSCW: Tools and Trials. *ACM Transactions on Information Systems* 11(4): 401-424.
- [39] C. Schmandt (1993). Phoneshell: the telephone as computer terminal, ACM international conference on Multimedia, Anaheim, California, United States, 373-382,
- [40] G. C. Sychay, E.; Goh, K. (2002). Effective image annotation via active learning, *Proc. IEEE International Conference on Multimedia and Expo*, 2002, 209-212, Aug. 2002.
- [41] L. K. Trevino, R. H. Lengel and R. L. Daft (1987). Media Symbolism, Media Richness, and Media Choice in Organizations: A Symbolic Interactionist Perspective. *Communications Research* 14(5): 553-574.