Some Observations on the "Live" Collaborative Tagging of Audio Conferences in the Enterprise

Shreeharsh Kelkar   **Ajita John**   **Doree Duncan Seligmann**

Avaya Labs Research, 233 Mt. Airy Road, Basking Ridge NJ 07920

skelkar, ajita, doree @avaya.com

ABSTRACT

This paper describes preliminary findings related to a system for "live" collaborative tagging of enterprise meetings taking place on an audio bridge between distributed participants. Participants can apply tags to different points of the interaction as it is ongoing and can see, in near real-time, the "flow" of tags as they are being contributed. Two novel types of tags are proposed: "deep tags" that apply to a portion of the interaction and "instant tags" that apply to an instant of the interaction. Our system is being used by enterprise users and we analyze a corpus of 737 live-tags collected from 16 conversations that took place over several months. We found that the live-tags for audio have slightly different characteristics from Web 2.0 tags: they are longer and confer affordances on the audio like description and summarization. Some observations on the "cognitive cost" of live-tagging are offered.

ACM Classification Keywords

H5.3. Information interfaces and presentation (e.g., HCI): Group and Organization Interfaces.

General Terms

Design, Human Factors.

RATIONALE AND PRIOR WORK

Audio conferences are the lifeblood of modern geographically distributed enterprises. They allow employees from remote locations and home offices to interact and collaborate without having to travel physically. While the aids for these conferences such as presentations are often saved, the audio is generally not recorded. If recorded, then an audio repository of conversations could be tapped into by users who miss a meeting, users who want to inform others of something relevant that was spoken in a meeting, etc. This repository needs to be easily searchable, rich in context for navigation, and browsable.

In this paper, we present a system that aims to create such a repository by allowing the "live" collaborative tagging of distributed enterprise conversations i.e. by allowing participants to *tag conversations as they are ongoing.* A

web-based application, called the Live Conference Dashboard (henceforth LCD), allows geographically distributed meeting participants to tag an ongoing meeting. The tags can be applied to specific portions of the ongoing interaction (whole interaction, segments of it, and instants of it). The audio in the meeting gets recorded and participants are shown, in real-time, the tags being contributed during the meeting. The audio, indexed by the tags, can be accessed by the participants after the meeting is over using another web-based interface.

Distributed enterprise conferences are generally accompanied by a parallel IM channel where the participants exchange messages that are parenthetical to the discussion. Recently, Twitter has emerged as a medium which participants in a public interaction (a keynote talk, an exhibition) use to comment on it. IM and Twitter are principally communication media; tagging is for annotation and self-presentation. We believe that the way a system is built and named influences how people use it. E.g. Clay Shirky [7] points out that one of the reasons for Wikipedia's success is its name which meant that its "early users were guided by the rhetorical models of existing encyclopedias." We wanted our participants to be *tagging* the interaction rather than commenting on it or talking to each other (although the free-form nature of the tags allowed them to do this too). This would bring to the artifacts of such interactions the affordances that tagging brings to web-artifacts (like description), thus facilitating the searching and browsing of rich-media streams.

This work is a continuation of our work on Echoes [5, 6], a system that allowed users to archive, tag and replay distributed audio conversations; however, the tagging could only happen after a meeting ends. The overall goal of the LCD is similar to that of [4] i.e. to "salvage" the artifacts of an interaction. It is also somewhat similar to [2] which investigates real-time social annotation of (audio) lectures in the classroom by students; however, our focus is on a more collaborative environment (employees of a company discussing issues on an agenda, say) and where the participants are not face-to-face.

In [1, 3], we see that tags on the web are used for a variety of purposes: self-presentation, description and categorization of content, action items, comments, etc. The tagging activity in the LCD is different from web-tagging in the following ways: (1) it is primarily producers

of the content (i.e. the audio recording) who provide the tags rather than the consumers. (2) The content is the product of intense and real-time collaboration between participants. (3) The tags are added as the content is being created rather than afterwards.

This paper is organized as follows. We first describe the LCD and its affordances. We go on to describe our dataset. We then perform a detailed content analysis of the tags that were contributed. Finally, we offer some preliminary observations on the "cognitive cost" of live-tagging.

HOW LIVE-TAGGING WORKS

Distributed meetings in an enterprise generally take place on a "meet me" type of Conferencing Bridge. Participants in a meeting typically use their desk-phones to dial in. If they wished to live-tag the conference, we asked them to log in to the LCD. On doing so, the LCD would show them the tags contributed to this current meeting by other users and allow them to add tags to it as well. The LCD is built as a web-based Java Applet and consists of 3 parts: a tag cloud, a tagging dashboard and a timeline, arranged vertically in sequence. We constructed the LCD such that it occupied approximately 25% of the screen-width so that it would not interfere with the participants' other activities, such as viewing a PowerPoint presentation.

The tag cloud and the tagging dashboard are shown in Figure 1. The tag cloud shows the 25 most recent tags, arranged in chronological order with the most recently contributed tag at the top. The size of the tag is proportional to the number of times it has been used. The number of people who used that tag is appended to the tag. If only 1 user has contributed a tag, then the user's name is displayed next to it. Tags are color-coded according to the user who contributed a tag.

Tags can be words or phrases of any arbitrary length and can be of three types. *Whole tags* apply to the whole conversation. They are "one-click" tags i.e. the user types in the tags in the textbox provided, separates them with commas, and clicks "Apply Whole Tag". *Deep tags* apply to a part of the conversation (i.e. a segment of audio). They are "two-click" tags i.e. the user needs to click two buttons in order to apply a deep tag to a segment. First, the user clicks on "Start Deep Tag" when she hears something she wants to tag. A message to the effect of "Segment started at 11.40 am" (i.e. the time at which the segment is started) is shown. Because some time will elapse between users hearing something interesting and deciding to tag it, we allow the user to offset the beginning of a segment: the segment can start *n* minutes before the user pressed the "Start Deep Tag" button. When the user thinks the segment she wishes to tag is over, she presses "Apply Deep Tag" and the tags get applied to that segment of the audio. *Instant tags* apply to the current instant of the conversation. They can be applied with a single-click. Instant tags make possible the nesting of segments: allowing a user to apply an instant tag after starting a segment for deep tagging.

Figure 1: The left figure shows the default state of the LCD. The right figure shows the state of the LCD when a deep tag segment has been started.

The timeline appears below the tagging dashboard and is shown separately in Figure 2. The tags are color-coded, depending on the user who contributed the tag and shown as dots overlaid on a timeline. By moving the mouse-over the dot, the tag, the user who contributed it, along with the other tags attached to that segment is revealed. At this point, a user can rate a tag as well as add more tags to the audio segment a tag is attached to.
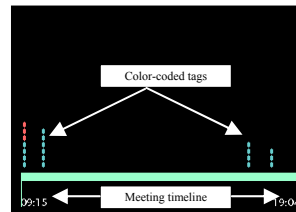

Figure 2: The timeline of the tags is shown below the tagging dashboard. The tags are shown overlaid on top of the timeline of the meeting.

The tag cloud shows participants the most recent "topics" discussed in the meeting. The size of a tag (which depends on the number of participants who contributed the tag) shows them the "importance" of a certain topic. The timeline allows participants to see the "flow" of the meeting, in terms of the topics (tags) discussed. The LCD automatically refreshes the tag cloud every 30 seconds. A "Refresh" button at the top provides a forced refresh.

DESCRIPTION OF OUR DATASET

16 distributed meetings in our company that took place between April 16 and September 4, 2009 were live-tagged by their participants (see Table 1) and the 737 contributed tags are analyzed in this paper. Live-tagging a meeting was optional for the participants of these meetings. The host of the group that used the LCD consistently put up the tagged audio on a wiki for others to catch up on missed meetings. The host himself consistently tagged the conversations throughout the last few months, indicating its usefulness.

Table 1: Detailed data on the live-tagged meetings

Meetings (arranged chronologically from first to last)	Estimated number of participants in the meeting	Number of participants who logged in to the LCD	Number of taggers
1	30	15	12
2	8	4	2
3	3	3	2
4	3	3	3
5	3	3	1
6	15	5	3
7	15	5	3
8	1000	10	5
9	4	2	2
10	11	7	4
11	1000	3	2
12	8	4	3
13	15	9	2
14	4	2	2
15	15	7	3
16	15	2	2

CONTENT ANALYSIS OF LIVE-TAGS

We performed a detailed content analysis on the contributed tags which revealed that live-tags for audio conversations have characteristics that are somewhat different than those used on the web. One difference is the sheer length of the tags in terms of the number of words. Tags like "defining the value of social media slide 5" are not uncommon. Analyzing the tags we found **78% of our tags were more than 1 word in length, and 25% more than 4 words, unlike web-tags.** Some tags were up to 12 words in length.

Another difference relates to the *function* the tags perform. To discover this, we coded all the tags by assigning two kinds of *values* to every tag, similar to what [1, 3] do for web tags. **Function values** brought out the functions that the tag seemed to be performing e.g. providing a description of a specific audio segment. Assigning function values can also be thought of as discerning the *intent* of the tagger. Based on these function values, we could identify 6 functions that the tags perform. (a) *Provide topic of an audio segment*: E.g. "what is social media marketing?", "opportunities through twitter" etc. (b) *Provide Description of an audio segment*: E.g. "collaborate with partners", "cross collaboration system search", etc. (c) *Marking/Indexing parts of audio*: E.g. "silence", "slide 7", "break", etc. (d) *Indicating turn-taking in a conversation*: E.g. "louise key points". (e) *Providing commentary on an audio segment*: E.g. "not boring". (f) *Play*: Used by participants to communicate with each other in real-time or just to play around. E.g. "jxxx gxxxxxx is tagging a lot". The distribution of tags with respect to their function values is shown in Table 2. We see that the *description* and *topic* functions dominate (by default, all instant tags were

assigned the "marker" function value). Unlike Web 2.0 tags, the "categorization" and "self-presentation" functions were missing for live-tags.

Table 2: Distribution of the function values of the live-tags

Function values	Number of tags	Function values	Number of tags
Marker	507	Commentary	35
Topic	205	Play	11
Description	227	Summary	1
Turntaking	64		

Reference values listed the *entities* (e.g. people, technologies, events) that the tag referred to; these may be thought of as discerning the *content* of the tag. The main reference values are: *People, Groups, Organizations, Activities, Projects, Technology, Resources, Companies*. A tag can have multiple function and reference values.

The difference between the nature of live-tags and Web 2.0 tags may be attributed to the following factors: (1) the incentive for live-tagging interactions may be different from that of web-tagging; in our explanatory emails to users about how to use the LCD, we indicated that the tags would be used for searching and browsing audio, especially so that users who missed a meeting could get to the interesting segments. Several users indicated that they found the ability to point out interesting segments of their meetings to people who had missed them very appealing. (2) On the web, taggers are generally consumers of content, while our taggers are the creators of the content and in fact, enmeshed in its creation as they tag. Thus, they may tag more with the intent of describing what they are doing rather than categorizing it. This may also be one reason for the high word-length of the tags (since describing a portion of a meeting takes more than 1 word).

The content analysis suggests that from the point of view of *function*, tags may help in *audio navigation*, while from the point of view of *reference*, they may help *audio search*.

MUTUTAL AWARENESS AND COGNITIVE COST

In the first meeting in which the LCD was used, we asked the taggers in that meeting to answer a user survey. Out of the 12 taggers, 8 responded. To discern whether collaborative live-tagging promoted mutual awareness between participants in a meeting, we asked our users whether the flow of tags inspired them to tag as well. 6 out of 8 said "yes." When asked if the live-tags gave them a "different perspective" on the ongoing conversation, the response varied evenly between "rarely" and "sometimes."

Users also reported difficulties with live-tagging. One user remarked (minor spelling errors have been fixed): *"Had difficulty relating the context of the tag given the conversation. Once I saw a tag I was thinking about what was being said before and relating it to the context of the tag that I lost the conversation in process. Ended up as a bit*

confusing." In response to another question we were told that "*it [i.e. tagging] is very intrusive on your time while trying to pay attention*" and that "*adding them manually is a lot of work*". This brings out the fact that since live-tagging is done as the conversation is going on, it will therefore tend to divert a user's cognitive resources from the conversation itself. We call this the "cognitive cost" of live-tagging. 3 observations about this are listed.

(i) *No participant in any meeting tagged and spoke at the same time*: Active participants (e.g. the "host" of the meeting, whose job would be to coordinate the meeting) would tag *after* their speaking turn ended.
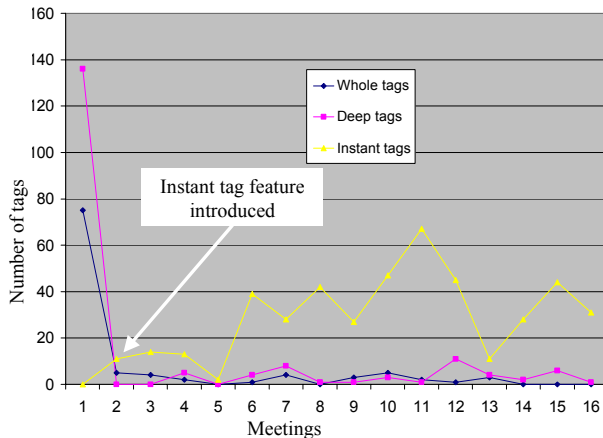

Figure 3: The distribution of whole, deep and instant tags in the 16 meetings.

(ii) *Taggers preferred to use the less 'effortful" tags i.e. instant tags over deep tags*: Deep tags took more effort; they needed two clicks (to indicate the starting and ending of a segment) which in turn required a high level of alertness. Instant tags, on the other hand, were "one-click" tags. By and large, users preferred to use instant tags rather than deep tags. In Figure 3, we see that in meeting 1, where the instant tag feature was not available, 72 whole tags and 141 deep tags were contributed. The number of whole tags is surprising although our surmise is that some taggers preferred the simpler-to-apply whole tags even when they knew that these tags could only point to the whole audio file (and consequently not help in audio navigation). Once the instant tag feature was introduced in the next meeting, the number of deep tags (and whole tags) dropped steeply. From meeting 2 onwards, the number of whole and deep tags is consistently lower than the number of instant tags.

Recently, we asked 5 of our users, who have used the LCD consistently over these 16 meetings, and are therefore very familiar with it, about their preference for instant tags. One user said, "*Most of the time I am tagging and talking at the same time (which is a real challenge) ... that's what causes me to fall back on instant tag ....*" Another said, *".... in meetings I'm in, a deep tagging segment may be overlapping with a segment that relates to a different point ... I just do instant tagging. It's easier for me.*"

Another factor for our taggers' preference for instant tags over deep tags could be the current deep-tagging interface feature itself and our users' unfamiliarity with it. At least 2 users have indicated to us how this could be improved. Suggestions include having a red "record" icon blink when a deep tag segment is started, selecting a portion of the timeline using the mouse, among others.

(iii) *The average length of a deep tagged segment was 3 minutes, while the standard deviation was slightly less than 3 minutes*. The length of the deep-tagged segment is the time between the two clicks required to deep-tag. Our hypothesis is that participants find it hard to deep-tag a larger segment, especially since the conversation keeps flowing, bringing out other things they may want to tag.

CONCLUSIONS AND FUTURE WORK

In this paper, we reported our analysis of 737 live-tags collected from 16 enterprise meetings. Our preliminary findings indicate that live-tagging has a cognitive cost. In the future, we plan to improve the LCD to reduce the cognitive cost of live-tagging and carry out user studies to: (a) understand empirically the cognitive cost of live-tagging (b) estimate the feasibility of live-tagging for different types of conversations, (c) observe how the live-tags allow the searching and browsing of audio for enterprise work.

REFERENCES

1. Golder, S. and Huberman, B. Usage Patterns of Collaborative Tagging Systems. *Journal of Information Science*, *32* (2). 198-208.

2. Kalnikaite, V. and Whittaker, S., Social Summarization: Does Social Feedback Improve Access to Speech Data? in *ACM Conference on Computer supported Cooperative Work*, (San Diego, CA, 2008).

3. Marlow, C., Naaman, M., et al., HT06, tagging paper, taxonomy, Flickr, academic article, to read. in *Conference on Hypertext and Hypermedia*, (Odense, Denmark, 2006), ACM, New York, NY, USA.

4. Moran, T.P., Palen, L., Harrison, S., et al., "I'll Get That Off the Audio": A Case Study of Salvaging Multimedia Meeting Records. in *CHI*, (Atlanta, GA, 1997).

5. Renduchintala, A., Kelkar, S., John, A. and Seligmann, D.D., Designing for Persistent Audio Conversations in the Enterprise. in *DUX '07*, (Chicago, IL, USA, 2007).

6. Seligmann, D.D., John, A. and Kelkar, S., Why Tag Conversations? in *International Conference on Collaborative Computing: Networking, Applications and Worksharing*, (White Plains, NY, 2007).

7. Shirky, C. *Here Comes Everybody: The Power of Organizing without Organizations*. Penguin Press, 2008.