Social Summarization: Does Social Feedback Improve Access to Speech Data?

Vaiva Kalnikaitė
The University of Sheffield
Regent Court, 211 Portobello Street,
Sheffield, S1 4DP, UK
v.kalnikaite@sheffield.ac.uk
+44(0)114 222 2666

Steve Whittaker
The University of Sheffield
Regent Court, 211 Portobello Street,
Sheffield, S1 4DP, UK
s.whittaker@sheffield.ac.uk
+44(0)114 222 2630

ABSTRACT
We extend the notion of social tagging to construct *social summaries* of complex multimedia materials. Our system allows students to apply time-indexed multimedia tags such as handwritten annotations or photos to different parts of lecture recordings. These tags can be used to straightforwardly access different parts of the lecture. The social component of the interface presents information about which tags are most frequently accessed by others: allowing students to infer those parts of the lecture of most interest to others. We demonstrate the utility of the approach in a 6 week fieldwork study. Social summaries are used much more than corresponding systems that do not provide social information. In addition, social tool use was correlated with high course marks.

Author Keywords
Tagging, Collaborative Systems, Notes, Photos, Popularity, Speech Browsing, Speech Retrieval, Web2.0.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
There has recently been huge interest in social computing, based on insights that the activities of many people can be exploited to help solve individual and small group tasks. For example, search engines rely on user linking behaviour and link labelling to rerank results generated using textual IR methods [1]. And social tagging systems such as dogear, dig, deli.cio.us, flickr, citeulike and reddit support the creation of labels and ratings allowing users to access content previously categorized by others.

Such systems tend to operate with large user bases on the web, with user tags and content generally (though not always) being textual. Here, in contrast, we built and evaluated a novel system that explores social tagging in a smaller user population, where tags and content are multimedia.

We investigate the utility of tagging to construct *social summaries* of complex multimedia materials. Our system allows students to apply time-indexed tags, such as handwritten annotations or photos, to different parts of multimedia lecture recordings. These tags can be used to straightforwardly access different parts of the lecture. The social component of the interface presents information about which tags are most frequently accessed by others: allowing students to infer which parts of the lecture have been of most interest or value to others.

It is well known that students have problems in remembering complex materials presented in lectures, and that their own notes or official handouts are often inadequate for later revision. In our system, a student annotates/tags the lecture in real time, e.g. by taking digital notes/photos when significant points are made. Using the principle of temporal tagging [3, 7], these tags are time aligned with the original recording (see Fig 1). Clicking on a tag (e.g. 'explanation of Nielsen's usability heuristics') will replay what was being said when the note was made or picture taken. This allows the student to easily revisit specific parts of the lecture. The tags are therefore a *visual analogue* to the lecture content - allowing rapid access to key parts of the underlying speech.

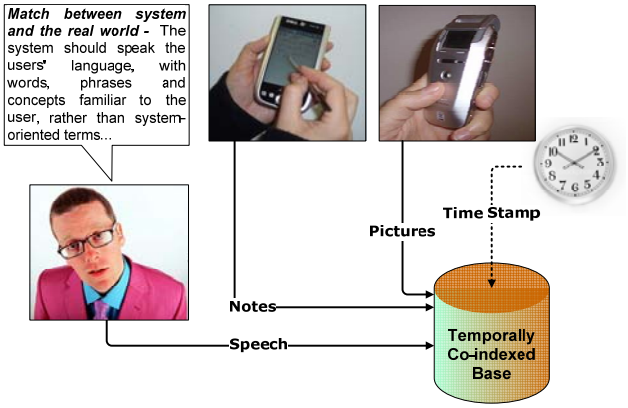

Figure 1: Temporal tagging technique

CSCW'08, November 8–12, 2008, San Diego, California, USA.

We also provide social feedback about which tags (and hence which parts of the lecture) that *other students* found to be useful. Following the principle of social navigation [4], our interface provides visual feedback about frequently accessed tags. Social tagging is a promising technique for multimedia summarization where effective content-based techniques have proved hard to develop [8,9].

The interfaces for pictures and notes are shown in Figs 2 and 3 respectively. Fig 2 shows digital photos taken during a lecture. Each time a student clicks on a photo to access a specific part of the lecture, we make that tag more salient by enlarging it relative to other photos. For digital notes, we adjust their salience by highlighting/colour. Fig 3 shows how notes that have been used most frequently are highlighted in **bold** and in red.

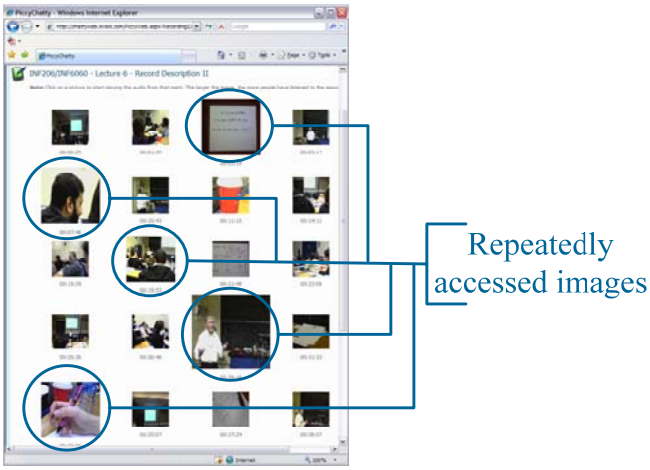

Figure 2: Social Picture Summary

The interface therefore provides aggregate information about which parts of the lecture other users accessed most often. We evaluate social summaries in a naturalistic lecture setting where students used the systems to help with their everyday schoolwork.

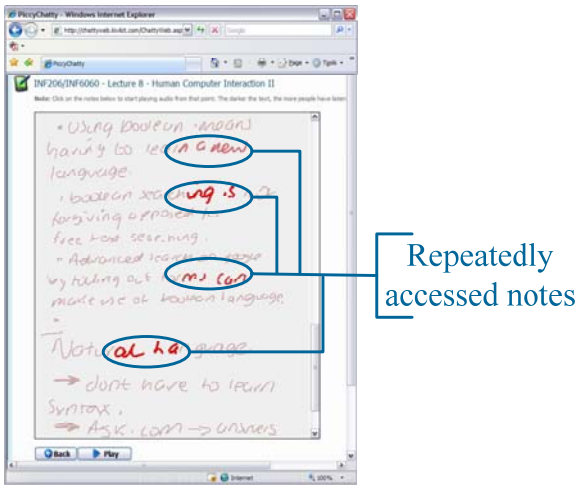

Figure 3: Social Notes Summary

Our main research questions were:

- Do users make *greater use* of systems offering social feedback? While there are potential advantages to providing information about others' behaviours in supporting access to critical parts of lectures, there are also disadvantages: students may find the social interface confusing or want to make their own decisions about what is important. We evaluate these trade-offs by comparing use of system versions that provided social feedback, with versions that did not provide such feedback. We also examined the benefits of social feedback *within each lecture* by looking at whether people preferred to access indices that had previously been accessed by others, and whether use of such social information became more prevalent over time.

- Which *types* of social tags are most useful for retrieving lecture materials? Are photos or notes more useful as indices? We compared the utility of social summaries that presented digital *handwritten* tags with those that used *photos* to tag the underlying speech.

- *What* are the *benefits* of using this type of system? Does using the system help people achieve better overall grades or is it mainly used by students who have difficulty following lectures? We evaluated this by examining the relationship between system usage and coursework scores.

METHOD

Systems

For each of 10 2-hour lectures in a course on Information Retrieval, initial tags were created by giving two volunteer students: (a) a pen-based digital note-taking tool or (b) a modified digital camera (see Fig 1). In both cases tags were time-indexed to an underlying speech recording. There is normally a time delay between hearing an important part of the lecture and creating the tag. Based on prior work [5] we therefore offset the tag by 1.5s to allow for this delay.

Volunteers were instructed to take notes as they normally would in a lecture. In the case of photos they were told to capture the images they thought would be most useful as retrieval indices. On average students took 174 words as notes, and 60 pictures. Fig 3 shows that digital notes were similar to regular handwritten notes. This is consistent with prior research [5]. Student comments in post hoc interviews revealed three main types of photo capture (see Fig 2): (a) *key slides* to capture verbal discussion associated with these, (b) *people* who asked questions, (c) *contextual images* of surrounding people or the lecture theatre to recreate the context of the lecture. One annotator commented: *"[I] tried to take pictures of people asking questions"* and *"when listening to what the lecturer was saying [I] tried to capture reactions of students."*

We used these tags to create 4 different versions of the system. Two versions included social feedback: Social Picture Summary (Fig 2), and Social Notes Summary (Fig 3). The other two *Asocial* versions (one with photos and one with notes) provided the same overall set of tags but without social feedback information. Thus, in Asocial versions, all photos were the same size and handwritten

notes were presented at the same level of greyscale and colour salience.

Usage Data

Our study ran over the last 6 weeks of the course. 25 students took part in the course, and had access to all 4 systems. In the first lecture of the course, students were given a detailed description and demonstration (using materials collected during a prior course) of the 4 systems, and reminded throughout the course that they could use them to access prior lectures. It was explained that the aim of the systems was to help retrieve information from lectures, e.g. if they had found lecture content to be confusing when it was initially presented, or if they simply wanted to revisit the entire lecture.

Access to the systems was provided via a webpage, which was prominently referenced on the course website. The webpage provided tutorials about each of the four systems, as well as links to access each of the systems. Users were allowed to freely choose which version of the system they wanted to access.

We collected naturalistic data about system usage, logging the number and duration of access sessions for each student; which version of the system they accessed, which tags they used, and how long they listened to speech associated with each tag. We also elicited qualitative feedback via an email survey. Finally we collected overall coursework scores to explore the relation between system usage and student performance.

RESULTS

Do people make more overall use of tools that present social feedback?

Fig 4 shows the total usage time across all users for each of the 4 systems. There was almost no use of the Asocial systems, but significant use of the Social systems (respective mean usage times/user 10.9 and 462.1 secs, SD = 449.8, Wilcoxon's $W^1(25) = -213$, $z = -3.2$, $p<0.001$). As one user said: "*…when I missed a lecture, I briefly looked at the [social summary tools] just to get a general idea of what happened in the lecture.*"

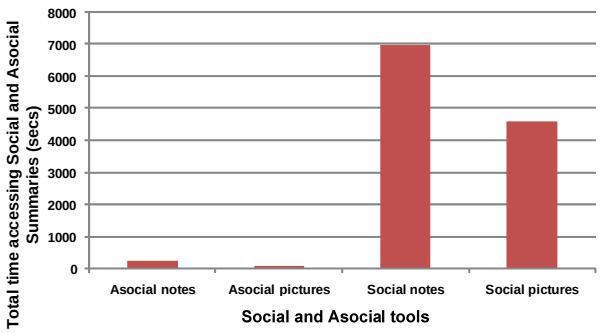

Figure 4: Social and Asocial Summary Use

We also evaluated the use of social information *within* each lecture. Within a given lecture, we can distinguish between accesses that use tags previously accessed by others (i.e. popular tags), from those not previously accessed by others. Overall, the benefits of social feedback were clear: almost half (49%) of all accesses were for *popular* tags. This is well above baseline which would predict 3% repeat accesses for pictures and 1% for words if users had simply chosen tags at random.

Furthermore, this reliance on popular tags became more prevalent over time. If popular tags are indeed useful, then we should expect greater use of these over time. We divided the 6 week test into two halves. We compared the proportion of popular accesses (i.e. number of accesses using tags that had previously been selected by someone else, divided by the total number of accesses) in the first versus the second half of the trial. Table 1 shows that for the Social Summaries there was a global shift to using popular tags in the second half of the trial. Social Pictures, shifted from 0 selections of popular tags in weeks 1-3 to *all* in weeks 4-6 (respective means = 0, and 0.25, SD = 0.5 Wilcoxon's $W^1(25) = 91$, $z = 3.2$, $p<0.001$). Similarly, there was also a significant shift in popular tags for Social Notes in the second half (respective means = 0.15, and 0.23, SD = 0.37 Wilcoxon's $W^1(25) = 36$, $z = 2.2$, $p<0.03$).

Tool	Social Notes Summary	Social Pictures Summary	Overall
Weeks 1-3	44%	0%	35%
Weeks 4-6	56%	25%	65%

Table 1: % of accesses of popular tags (those previously selected by someone else), for the first versus the second half of the trial

Which tags are better: pictures or words?

We also analyzed whether people tended to make more use of Social Notes versus Social Pictures. As Figure 5 indicates, there was greater overall usage of Social Notes, when we compared mean number of clicks for each type of tag for each participant (respective means = 88.83, and 9.2, SD = 156.4 Wilcoxon's $W^1(25) = 133$, $z = -3.1$, $p<0.002$) Why was this? It seemed users preferred notes because they provided a finer granularity of access (recall that there were almost three times as many word tags as pictures). Furthermore, some of the picture tags tended to be taken to provide contextual information (e.g. photos of those sitting close to the student in a given lecture), and such contextual information may have been less useful for retrieval.

As one user said: "*…I did not find the pictures themselves valuable…I found it easier to find the relevant part of the lecture using [social notes].*"

[1] Non-normal distributions led us to use non-parametric statistics throughout.

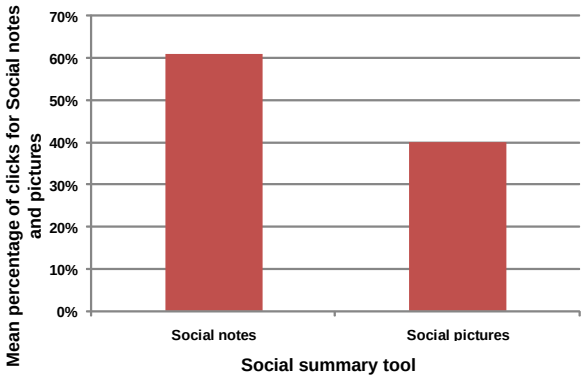

Figure 5: Comparing Note and Picture Indices

Does Social Summary use predict higher coursework scores?

Finally, we looked at how system use relates to academic performance. There was a significant correlation between coursework marks and number of Social sessions (Spearman's $rho^1(25)=0.72$, $p<0.001$), but not Asocial sessions ($rho^1(25)=0.42$, $p>0.05$). It is possible that more motivated students are more likely to make greater use of our systems. However, this would lead us to expect an increase in coursework scores for *both* Social and Asocial systems. Instead the elevated coursework marks are *only* associated with Social system use, suggesting learning benefits that are specific to Social systems. One student described the benefits: *"Hearing the lecture notes whilst reading the lecture [social summary] notes helps greatly with revision. Today I was scanning through all the [social summary] notes whilst listening to them, and this helps enormously with revision."*

CONCLUSIONS

We have both implemented and evaluated the utility of a novel working system that presents Social Summaries, derived from the analysis of tag popularity. Our fieldwork results show in a real-life practical setting that this is a promising technique for accessing complex multimedia materials. The results are consistent with other research on social computing [1, 4], as well as collaborative note-taking [2, 3, 6] in demonstrating the utility of social information. They also promise to make an important contribution to multimedia access of conversational data, where effective automatic summary techniques have proven hard to develop [8, 9]. Our approach also finesses a weakness of many social computing applications in motivating users to tag. Here tags are generated as a *side-effect* of existing user practices, such as note-taking and, more recently, taking photos of presentations. At a practical level, the tool seemed to help the educational process with greater use predicting higher course marks.

There are a number of possible extensions of the approach to related applications. For example, social tags might be combined for large scale social events such as rock concerts or sporting events where many people attending the same event might upload photos to construct a large scale social summary – allowing group access to favourite moments from the event. And the same techniques might be used to create shared tags for meeting records for the purposes of creating high quality minutes [3]. We might also explore the relations between content-based and social summary techniques.

Another direction might be to develop techniques for presenting *configurable* summaries. Instead of supporting end user browsing through a set of tags, we might precompile an audio summary of a lecture based around the audio clips associated with, say, the 10% most popular indices. As with other temporal compression techniques [8, 9] users could specify a desired compression level, and the system would compile the appropriate summary.

Finally these interfaces present interesting social challenges that need further empirical investigation. Less dedicated students may increasingly rely on the efforts of more conscientious classmates: e.g., waiting for others to access and 'mark up' the main parts of the lecture before they begin their own revision. And of course there is a potential problem of 'overfitting': students may become so focused on popular tags that they neglect major parts of the lecture. Despite these potential problems, we believe the approach shows demonstrable promise in an area where it has been difficult to develop effective techniques.

ACKNOWLEDGEMENTS

We thank our course participants for their help.

REFERENCES

1. Brin, S., Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *In Proc. of the seventh international conference on World Wide Web*, 107-117.
2. Brotherton J. A., Abowd G. D. (2004). Lessons learned from eClass: Assessing automated capture and access in the classroom. *ACM Trans. on CHI04*, 121 - 155.
3. Davis R. C, Landay J. A., et al (1999). NotePals: lightweight note sharing by the group, for the group, *In Proc. CHI99*, ACM Press, NY, 338-345.
4. Dieberger, A., et al (2000). Social navigation: techniques for building more usable systems. *Interactions*, 7(6), 36-45.
5. Kalnikaite, V., and Whittaker, S. (2007). Software or Wetware? Discovering When and Why People Use Digital Prosthetic Memory. *In Proc. CHI07*, ACM Press, NY, 71-80.
6. Pimentel, M. G. C., Abowd, G. D., et al (2000) Linking by interacting: a paradigm for authoring hypertext. *In Proc. of Hypertext and Hypermedia*, ACM Press, NY, 39–48.
7. Steifelman L. J., et al (1993). VoiceNotes: a speech interface for a hand-held note taker. *In Proc. of ACM CH1'93*, ACM Press, NY, 179-186.
8. Tucker S., and Whittaker S. (2006). Time is of the Essence: An Evaluation of Temporal Compression Algorithms. *In Proc. of ACM CH1'00*, ACM Press, NY, 329-338.
9. Whittaker, S., Tucker, S., et al (2008). Design and Evaluation of Systems to Support Interaction Capture and Retrieval. *Pers Ubiquit Comput* 12:197–221.