Catchup: A Useful Application of Time-Travel in Meetings

Simon Tucker
University of Sheffield
s.tucker@shef.ac.uk

Ofer Bergman
University of Sheffield
o.bergman@shef.ac.uk

Anand Ramamoorthy
University of Sheffield
Anand.Ramamoorthy@Ugent.be

Steve Whittaker
University of Sheffield
s.whittaker@shef.ac.uk

ABSTRACT

People are often required to *catch up* on information they have missed in meetings, because of lateness or scheduling conflicts. Catching up is a complex cognitive process where people try to understand the current conversation without access to prior discussion. We develop and evaluate a novel Catchup audio player that allows 'time-travel'. It automatically identifies the gist of what was missed, allowing people to join the meeting late and still participate effectively. In a lab study, we evaluated people's understanding of meetings they had partly missed, by asking questions about meeting content. We tested whether providing Catchup gist overcomes the potential disadvantage that people must join even later - because catching up takes time. Catchup users understood meetings 70% better than controls who simply joined late. They were more confident of their understanding and indicated a positive attitude towards the tool. We are currently exploring more general applications of the time-travel approach.

Author Keywords

Meetings, Teleconferencing, Catchup, Audio Processing, Gist Extraction, Time Travel.

ACM Classification Keywords

H.4.3 Communications Applications: Computer conferencing, teleconferencing and video conference. H.5.2 User Interfaces Voice I/O User Interfaces: Evaluation/methodology, User-centered design.

General Terms

Algorithms, Experimentation

INTRODUCTION

Meetings are a central part of organizational practice and a key way that organizations generate knowledge and make decisions. However there are numerous productivity and co-ordination issues surrounding meetings. For example, it is not always possible to arrive at meetings on time. Unfortunately, if we do arrive late, then we are often unable to reconstruct the information we missed, and having others summarize prior conversation can be disruptive.

CSCW 2010, February 6–10, 2010, Savannah, Georgia, USA.

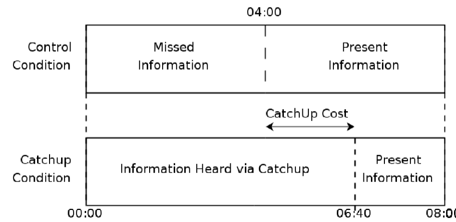

Figure 1: Timeline for the two experimental conditions, illustrating the costs and and benefits of Catchup.

Here we propose and evaluate a technical solution to these problems in the form of a 'Catchup' player for real time speech. The Catchup player identifies and plays the gist of what people missed, allowing them to begin listening any time after the meeting has started and still catch up with the real time meeting.

It is now becoming increasingly common to record meetings, especially teleconferences and Catchup methods operate on these recordings. Until recently, Catchup methods were restricted to using *speed-up*, where recorded speech is played at a faster rate [2]. However speed-up has an inherent upper-bound, as people cannot understand unfamiliar speech played at more than twice real-time [1]. People arriving very late can therefore never catch up with the real time meeting if they use speed-up.

Recently, however, new Catchup methods have been developed that support *speech gisting:* i.e. identifying important and removing unimportant meeting information [7]. These exploit real-time automatic speech recognition (ASR) which generates transcripts with word error rates of 40-60%. These error rates are too high to simply present the ASR transcript *visually* to the user. However prior research has shown that such error rates still allow us to develop effective *audio* gisting algorithms [6]. We apply text processing and information retrieval methods to the transcript to identify important information - which is then played back to the user. A key advantage of gisting over speed-up techniques is that it can remove as much speech as the user desires, and is consequently not subject to the comprehension threshold of speed-up. Thus gisting allows even very late arrivals to catch up.

Gisting has proved effective in other contexts, e.g. in supporting users in extracting the main points of entire meetings they missed [7]. However their application to real-time Catchup is more complex, as there are *trade-offs* involved in using gisting. Although gisting provides a summary of what was missed, this benefit must be offset against the *cost of missing a further part* of the meeting while listening to that summary. This cost is illustrated in Fig 1. Furthermore, gisting necessarily only provides a compressed and partial view of the missed part of the meeting; participants might learn more from joining the meeting immediately and inferring what they have missed.

This paper therefore implements and evaluates Catchup. Specifically, for 41 users in two design meetings we compare users' understanding of the meeting: (a) using Catchup to obtain the gist of missed meeting context at the cost of joining later; with (b) the current default situation of late arrival without access to prior context, but where one joins the meeting immediately.

THE CATCHUP SYSTEM

To extract relevant gist we remove unimportant utterances from the audio recording, by analyzing real-time ASR transcripts of the missed part of the meeting. Our algorithm ranks utterances by *importance*, by comparing them with a corpus of related meeting speech. We retain important elements and excise elements of lesser importance. Although this results in 'gaps' in playback, studies show that the inherent redundancy of speech means that people are able to successfully identify the main points of a compressed recording [6,7].

The technical details of gisting are as follows. We compute an importance score for each word in the ASR transcript using a standard IR technique: TF*IDF, where TF stands for Term Frequency and IDF for Inverse Document Frequency scores [4]. The simplicity of this technique allows our algorithm to work in real time. In the TF*IDF measure, the number of times a term appears in the ASR transcript of the missed region, is divided by the frequency of the term appearing in the comparative corpus of related meeting transcripts, thus the importance *(imp)* of a term, t, appearing in a particular transcript, d, can be computed as:

$$imp_{td} = \frac{\log(count_{td} + 1)}{\log(length_d)} \times \log(\frac{N}{N_t}) ,$$

Where $count_{td}$ is the frequency with which term t appears in transcript d, $length_d$ is the number of unique terms in transcript d, N is the number of transcripts in the corpus and N_t is the number of transcripts in the corpus which contain the term t.

All utterances in the transcript of the missed region are ranked according to their importance. We exclude stop words (words such as *the*, *it*, *of*, etc.) and then use the word importance measures to compute the importance of each utterance, defined as the mean importance of each non-stop word in that utterance. Utterances are selected for inclusion in the Catchup summary, depending on the level of compression specified. Here we used a compression rate of 2.5 times real-time, which means that over half of the utterances were omitted. Prior work suggests this compression rate is efficient for meetings speech [7].

EVALUATION

We evaluated Catchup by asking 41 people to 'participate' in two meetings. We asked them to listen to recordings of two meetings. They were told they were 'late' for both meetings but for one meeting they had access to Catchup that provided the gist of what they had missed. In the other they simply had to use subsequent information to make inferences about what they had missed. After the meetings, we tested their understanding by asking factual questions about different parts of each meeting. We probed understanding of both the parts that they had missed and the parts that they were present for.

METHOD

Materials

Two 8 minute segments were selected from the AMI meeting corpus [5]. The AMI meetings are not scripted, but participants are allocated various roles in a product development team (manager, designer, marketer, UI specialist). Participants are given orienting information associated with that role and participate in multiple meetings in a product design life cycle. Analysis shows the corpus is representative of real meeting interaction [3]. Meetings were chosen so that they realistically represent a meeting to which another participant could arrive late. There are obviously real-life meetings where content and participants are unfamiliar, and this is the situation we examined here. We identified meetings where there was a protracted discussion of a single topic, rather than multiple discrete discussions, because we wanted to explore the effects being late on subsequent understanding.

Both discussions were then segmented into 2 sections (see Figure 1). For Controls, the first four minutes represents the material missed by the latecomer and is not heard by them. Controls join the meeting at this midpoint and hear the remaining four minutes of audio at normal speed. Catchup users hear the gist of the missed section, but listening to this gist takes additional time, which means that unlike Controls they cannot join the meeting at the midpoint. They therefore spend 2 minutes and 40 seconds listening to Catchup audio, before they fully catch up with real-time. Catchup users then hear the remaining 1 minute and 20 seconds of the meeting excerpt at normal speed. Thus both groups listen for four minutes; the controls hear half of the original 8 minute excerpt in real-time and Catchup users hear the first 6 minutes and 40 seconds of the excerpt under compression and the remainder in real time. We used short excerpts here because using longer ones would mean that controls miss entire topics causing them to perform poorly, thus favoring Catchup.

In the test phase, we asked users 6 factual questions for each meeting. Answers were multiple-choice with users selecting one of four possible answers. The questions were of 2 types. *Missed* questions assessed users' ability to answer questions about the part of the meeting they missed because they were late. For Catchup users, this tested the effectiveness of gisting techniques. For controls it assessed their ability to exploit subsequent context to infer what had gone on. *Present* questions referred to information that the control group had heard in real time. *Present* questions assess participants' ability to contextualize the discussion they are present for. They assess whether Catchup gist helped Catchup users to understand what they heard when they were present, or whether controls performed better because they had joined the meeting 2 minutes 40s earlier.

We also asked subjective questions probing how confident people were about each of their answers, how hard Catchup was to understand, whether it was easier to understand the meeting with Catchup or control, and whether having Catchup made them less anxious about being late.

We should note that gisting was carried out completely independently of the process of question generation. It therefore turned out that when we analyzed the Catchup clips that the gisting algorithm had removed information that was directly relevant to one of the missed questions, making our evaluation a stringent test of the technology.

Participants and Procedure

41 participants, all fluent English speakers, took part: 29 were male and 12 were female. None reported any hearing problems. The purpose of Catchup was explained. Before starting the experiment proper, there was a short familiarization phase where people listened to a 1 minute example of Catchup. Once they were confident of the procedure, they heard one of the meetings in either the Catchup or Control condition. They were then asked the missed and present questions about the meeting, followed by subjective questions. They then heard the other meeting excerpt in the complementary condition which was then followed by questions. The presentation ordering of both meetings and conditions was counterbalanced across participants. At the end of the experiment, we asked further subjective questions and encouraged participants to voice any final thoughts about Catchup.

RESULTS

Our hypotheses and results were as follows:

Overall Effects

H1: Catchup participants will perform better overall than controls on questions about the meetings; Catchup not only provides a summary of what was missed, it also allows people to better interpret information when they are present.

We compared the overall percentage of correct answers to both types of factual questions in the two conditions, i.e. we combined responses to missed and present questions. A paired t test showed a significant difference, $t(40)=6.08$, $p<0.01$. Using Catchup resulted in a large positive effect on the percentage of correct answers raising it from a mean of 37% (standard deviation $\pm$15%) for the control condition to 63% ($\pm$20%) for the Catchup condition. These results indicate that Catchup users understood meetings 70% better than controls who simply joined late. We then compared effects for *missed* and *present* sections independently.

Missed Effect

H2: Participants who use Catch-up will perform better on questions about the 'missed' portion of the meeting. Catchup provides access to the gist of what they missed, whereas controls have to infer what they missed.

Another paired t test showed a significant difference in the percentage of correct answers for 'missed' information, $t(40)=4.18$, $p<0.01$. The average percentage of correct results was 34% ($\pm$23%) for the control group and 60% ($\pm$27%) for the Catchup condition. We also conducted a One-Sample test for controls and found that the percentage of correct answers for the 'missed' results (34%) was significantly greater than chance (25% could be achieved by guessing as there are four possible answers to each multiple choice question) $t(40)=2.51$, $p<0.05$. Thus, controls were partially able to use subsequent information to make inferences about the missed parts of the meeting. However the use of Catchup almost doubled their percentage of correct answers.

This result is by no means trivial, as Catchup gist is a partial representation of what was missed and it may not provide users with a good view of the early part of the meeting, e.g. if the automatically generated summary omits critical information.

Present Effect

H3: Participants using Catch-up will perform better on questions about portions of the meeting where they are present. This is because the missed part provides a context for what comes later. Despite being present for more of the meeting, controls do not have access to the initial context, and so should find it harder to interpret what goes on when they are present.

Results show an average of 37% ($\pm$26%) correct answers in the control condition compared to 59% ($\pm$37%) for the Catchup condition. A paired t test was significant $t(40)=2.88$, $p<0.01$. Again we conducted a One-Sample t test and found that the results of the control condition are significantly above the 25% chance percentage $t(40)=2.81$, $p<0.01$. This shows that controls showed a limited ability to make sense of what they were hearing 'live'. However, Catchup considerably increased this understanding.

Subjective Evaluation

H4: People should feel that they understood the meeting better when they use Catchup, and that Catchup is better than trying to infer what went on just from what was said. We further predicted that they would be more confident of their answers with Catchup.

We asked our participants questions, regarding their experience with Catchup using a 1-3 Likert scale. Table 1 presents the distribution of their answers.

The results imply that although Catchup introduced 2.5 times compression, the majority of participants didn't find

Question	Yes	Neutral	No
I found the Catch-up playback hard to follow	4 (10%)	5 (12%)	**32 (78%)**
Catchup helped me understand the meeting better	**32 (78%)**	7 (17%)	2 (5%)
I would use Catchup if it were available	**31 (76%)**	5 (12%)	5 (12%)
Using Catchup was better than trying to work out what went on without it.	**34 (83%)**	5 (12%)	2 (5%)
I felt calmer trying to understand the meeting with Catchup	**20 (49%)**	14 (34%)	7 (17%)

Table 1: Distributions of answers to subjective questions.

it hard to follow the sped up conversation. The 4 participants who found Catchup hard to follow commented that it was *"jumpy"* and that *"the sound in the playback could be improved"*. The majority of participants also indicated that Catchup helped them understand the meeting better and that they would use it if it was available. However only about half the participants indicated that they felt calmer when trying to understand the meeting using Catchup.

We also measured people's subjective confidence that they had given a correct answer for each of the questions using a 1-5 Likert scale. The average confidence for the Control condition question was 2.75 ($\pm$0.72) while Catchup increased participants' level of confidence to 3.54 ($\pm$ 0.49). A paired t test showed this increase was significant $t(40)=6.44$, $p<0.01$.

Qualitative Feedback

The majority of spontaneous comments were positive. Seven participants found Catchup to be "*useful*". One participant commented: *"Catchup is potentially a useful technology/tool. The utterance excision method is very useful. Removal of giggles, okays and other inconsequential words makes the Catchup playback easy to focus on and not annoying",* while another exclaimed that "*Catchup works*!". Interestingly three participants independently suggested adding video to Catchup. Only one participant made negative comments about Catchup, being concerned that "*Catchup may not be useful for capturing fine details*".

DISCUSSION & CONCLUSIONS

One way of thinking about Catchup is that participants are *time-traveling* in meetings – being physically present but listening to another part of the meeting. We have recently implemented generalizations of this time-travel idea. Another common problem with meetings is that we wish to revisit an earlier part of the meeting because we realize it is relevant to the current conversation. However to stop the meeting for all participants to revisit this point would be highly disruptive. Our new application therefore gists the entire meeting. This allows participants to drop out of the current conversation and relisten to the recording of the prior point. They can then use Catchup to gist the period missed while they were relistening.

Of course there are clear social issues surrounding the use of this technology. It may mean that participants are unaware of who is currently attending to the meeting and who is listening to an earlier part of it. These potential problems need to be evaluated against the benefits of being able to quickly revisit what one failed to understand. However these issues are not dissimilar from current meeting experiences. It is commonplace for participants to be physically present in meetings but to fail to pay attention to the conversation, spending the majority of their time processing email on their laptops.

Similar issues also arise in the practical deployment of Catchup. We imagine the technology would be best deployed in the context of phone conferences, where audio recording often already takes place. Current conferencing systems already indicate when new participants join the conference, but with Catchup we might generate such a signal only when the participant is fully caught up and ready to fully participate.

Overall, our results are very encouraging showing that Catchup improves participants' understanding, both of what they missed, as well as what they hear when they join the meeting. Catchup appears to provide a solution to a repeatedly occurring practical problem faced by people who arrive late for meetings. We also believe the more general concept of 'time-travel' has other interesting applications that we are currently exploring.

REFERENCES

1. Arons, B. Techniques, perception, and applications of time-compressed speech, In American Voice I/O Society, September (1992), 169–177.
2. Hejna, D. Real-time time-scale modification of speech via the synchronized overlap-add algorithm, Master's Thesis, M.I.T., 1990.
3. Hsueh, P. and Moore, J. (2009). Improving Meeting Summarization by Focusing on User Needs: A Task-Oriented Evaluation. In Proceedings of IUI, 2009.
4. Jones, K.S. A statistical interpretation of term specificity and its application in retrieval, J. Doc., vol. 28 (1972), pp. 11–21.
5. McCowan et al., The AMI meeting corpus. In Proc. of the 5th International Conference on Methods and Techniques in Behavioral Research, 2005.
6. Tucker, S., Kyprianou, N. and Whittaker, S. Time-Compressing Speech. In Popescu-Belis, A. and Stiefelhagen, R. (Eds.) Machine Learning for Multimodal Interaction, LNCS 5237, Springer-Verlag, p. 226-235, 2008.
7. Tucker, S. and Whittaker, S. Time is of the Essence: Proceedings of CHI, 2006, 329-338.