Designing Prosthetic Memory: Audio or Transcript, That is the Question

Sandra Trullemans, Payam Ebrahimi and Beat Signer
Web & Information Systems Engineering Lab
Vrije Universiteit Brussel
Pleinlaan 2, 1050 Brussels, Belgium
(strullem,pebrahim,bsigner)@vub.be

ABSTRACT

Audio recordings and the corresponding transcripts are often used as prosthetic memory (PM) after meetings and lectures. While current research is mainly developing novel features for prosthetic memory, less is known on how and why audio recordings and transcripts are used. We investigate how users interact with audio and transcripts as prosthetic memory, whether interaction strategies change over time, and analyse potential differences in accuracy and efficiency. In contrast to the subjective user perception, our results show that audio recordings and transcripts are equally efficient, but that transcripts are generally preferred due to their easily accessible contextual information. We further identified that prosthetic memory is not only used as a recall aid but frequently also consulted for verifying information that has been recalled from organic memory (OM). Our findings are summarised in a number of design implications for prosthetic memory solutions.

CCS CONCEPTS

• **Information systems → Speech / audio search**; • **Human-centered computing → Laboratory experiments**; *User studies*; • **Applied computing → Annotation**;

KEYWORDS

Prosthetic memory; note-taking; speech retrieval; verification; recall; personal information management

ACM Reference Format:
Sandra Trullemans, Payam Ebrahimi and Beat Signer. 2018. Designing Prosthetic Memory: Audio or Transcript, That is the Question. In *AVI '18: 2018 International Conference on Advanced Visual Interfaces, AVI '18, May 29-June 1, 2018, Castiglione della Pescaia, Italy.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3206505.3206545

1 INTRODUCTION

In our daily life we are presented with information and experience situations that we want to keep in memory. Nevertheless, our memory is fallible and therefore we use all kinds of prosthetic memories (PMs) such as photobooks or sticky notes. Prosthetic memory can take various forms and be used for different purposes. For example, in the field of Personal Information Management (PIM), prosthetic memories are often designed to help users organise and re-find documents by allowing them to link related documents as seen in HayStack [1] or OntoPim [13]. Another approach is taken in lifelogging applications such as MyLifeBits [4] which capture a user's interactions with their personal information and provide time-based cues during re-finding activities.

Prosthetic memory is often also designed for recalling information during knowledge tasks performed after meetings or lectures. In these solutions, environmental factors are captured by audio and video recordings and handwritten notes might be captured by advanced technologies such as digital pen and paper solutions [7]. Further, automatic speech recognition (ASR) can be applied to create transcripts of audio or video recordings. The tracked data is often temporally co-indexed in order to allow users to "jump" to a particular temporal position within the prosthetic memory by selecting parts of another PM. For example, by selecting a part of handwritten notes in the Audio Notebook [21], the captured audio recording is played from the position when that specific part of the notes has been taken.

Audio- and transcript-based PM as well as notes that are temporally co-indexed with audio prosthetic memory have been integrated in various tools and the advantages of prosthetic memory usage have been illustrated in a number of studies. However, less is known on how users actually use the provided audio recordings, transcripts and temporally co-indexed notes. In this paper, we investigate how text-based PM consisting of notes and transcripts as well as audio-based PM such as audio recordings are used and whether interaction strategies change when time elapses. We identified the following three main research questions:

RQ1: Is there a difference in the way users interact with notes, transcripts and audio recordings and do the interaction strategies change when time elapses? For example, when the audio recordings and transcripts are temporally co-indexed with the corresponding notes, users can "jump" to a part of a prosthetic memory by selecting a part of their notes. We investigate whether the used strategies are changing over time and why.

RQ2: Is there a difference in accuracy and efficiency between audio and transcript PM over time? The question is whether the answers based on audio and transcripts have the same accuracy and whether co-indexed audio and transcript PM can be accessed with the same efficiency.

RQ3: Do factors such as confidence influence the way users interact with PM? We will investigate whether users only use a prosthetic memory when they are not confident about their answer.

AVI '18, May 29-June 1, 2018, Castiglione della Pescaia, Italy

ACM ISBN 978-1-4503-5616-9/18/05…$15.00
https://doi.org/10.1145/3206505.3206545

We start by discussing related work in the domain of note-taking and introduce our Note4U study platform. The methodology of our study is then outlined and the results of the study are presented. Last but not least, we discuss our findings and derive a number of design implications for future prosthetic memory solutions.

2 BACKGROUND

Nowadays, most people have a smartphone or another kind of recording device at hand. This enables them to easily record situations such as informal conversations, meetings or lectures. While the recording of audio is a simple activity, the use of an audio fragment during re-finding activities can be a burden. Users have to first find the corresponding part of the audio recording and then they often have to carefully listen to the audio fragment multiple times. Over the last decade, major efforts have been made in the domain of spoken content retrieval (SCR) [3, 15, 26]. These technical advances found their way into HCI research and have been applied in various tools for meetings, phone calls and lectures. For example, the Ferret browser [27] allows users to browse meeting data such as video recordings or transcripts and provides communication channels to indicate which person has said what. In addition to the features offered by the Ferret browser, the Portable Meeting Recorder [14] provides automatically extracted metadata such as the location of participants during the recorded meeting. SCANmail [28] allows users to access voicemail messages through the transcript and highlights keywords in these voicemail messages. Finally, in the context of lectures, treemaps can be used to visualise the transcript of the recorded voice in order to facilitate navigation activities through the recorded audio fragments [18]. Note that the discussed features are also applied in commercial systems such as BrightSpace[1] which temporally co-indexes lecture videos, slides and transcripts. Other audio retrieval paradigms are, for example, explored in HyperMeeting [6] where users can create and follow hyperlinks between audio segments of multiple recorded meetings.

Since note-taking during meetings can distract the notetaker from the actual meeting, various tools have been developed where audio and transcripts are used as a base for the note-taking process. For example, Highlighter enables users to highlight parts of the transcript during the meeting while in Hotspots users can click a button to indicate important audio segments during the recording phase [9]. A similar approach has been studied by Carrascal et al. [2] where users can highlight important parts in phone call transcripts.

Despite the digital revolution, pen and paper-based note-taking is still a commonly used prosthetic memory. A major disadvantage of notes is that they might become inadequate for the retrieval of detailed information after some time [29]. In order to overcome this issue, applications have been designed which augment paper notes with audio recordings [20, 21, 31, 32]. Similarly, handwritten digital notes can also be temporally co-indexed with recorded audio as realised in Filochat [29] and ChittyChatty [11], or be augmented with pictures [5] or lecture slides as demonstrated in Livenotes [12].

In a collaborative setting such as a meeting, interactions with tracked data of other meeting participants can also be used as prosthetic memory. For example, parts of shared notes which are often accessed by co-workers can be highlighted and serve as a entry point into an audio recording [10]. This idea has been further investigated in the Collaborative Recorded Meeting [19] application which also addresses the navigation in video recordings. A similar approach has been taken in RichReview [33] which studied rich social interactions in the context of collaborative reviewing.

Even though as illustrated earlier the design of prosthetic memory tools is well established, there is still limited research on how, when and why OM is used. In the controlled laboratory study of Kalnikaité and Whittaker [11], the use of notes, a dictaphone and ChittyChatty—a tool where notes and recorded audio are temporally co-indexed—were compared. Participants were given three short stories and at three retention intervals (1 day, 7 days and 30 days) they were asked to answer questions about the particular stories with support of the corresponding prosthetic memories. The results show that the use of a prosthetic memory does not increase at longer retention intervals which is in line with the findings of Lyons [17]. Further, the accuracy and efficiency plays an important role in the choice of a prosthetic memory. When a prosthetic memory is accurate but inefficient to use (e.g. a dictaphone), users are willing to take the risk of only relying on their organic memory. The technique of temporally co-indexing notes with audio has been found the most accurate and efficient design.

3 NOTE4U

We have developed the Note4U note-taking solution which was used in our user study and can, for example, be applied in meetings or lectures. Note4U has been developed as a Microsoft Word add-in and provides the necessary prosthetic memories forming the subject of our analysis. The Note4U graphical user interface shown in Figure 1 consists of the following integrated PMs:

- **Notes:** Users can use a standard Microsoft Word document for note-taking.
- **Audio:** The audio player is a simple media player and visualises recorded audio in a waveform.
- **Transcript:** The transcript of an audio fragment is shown on the right-hand side of the notes.

We have chosen to integrate the required prosthetic memories in Microsoft Word since most users are familiar with this application, which might facilitate text editing and layouting during the note-taking phase of our study.

3.1 Interactions

In Note4U, the notes, audio recordings and transcripts are temporally co-indexed, enabling users to navigate between these three prosthetic memories. For example, when a user selects parts of their notes as shown in (1) in Figure 1, the corresponding part of the transcript reflecting what has been said when the selected part of the notes has been written is highlighted (2). In addition, the audio timeline cursor is moved to the start time of the highlighted transcript block (3). Similarly, users can double click in the transcript and the relevant part of the notes will be highlighted and the audio timeline cursor is moved to the corresponding position. Finally, users can move the audio timeline cursor to a specific time and the corresponding part of the notes and transcript block is highlighted.

[1] https://www.d2l.com/products/capture

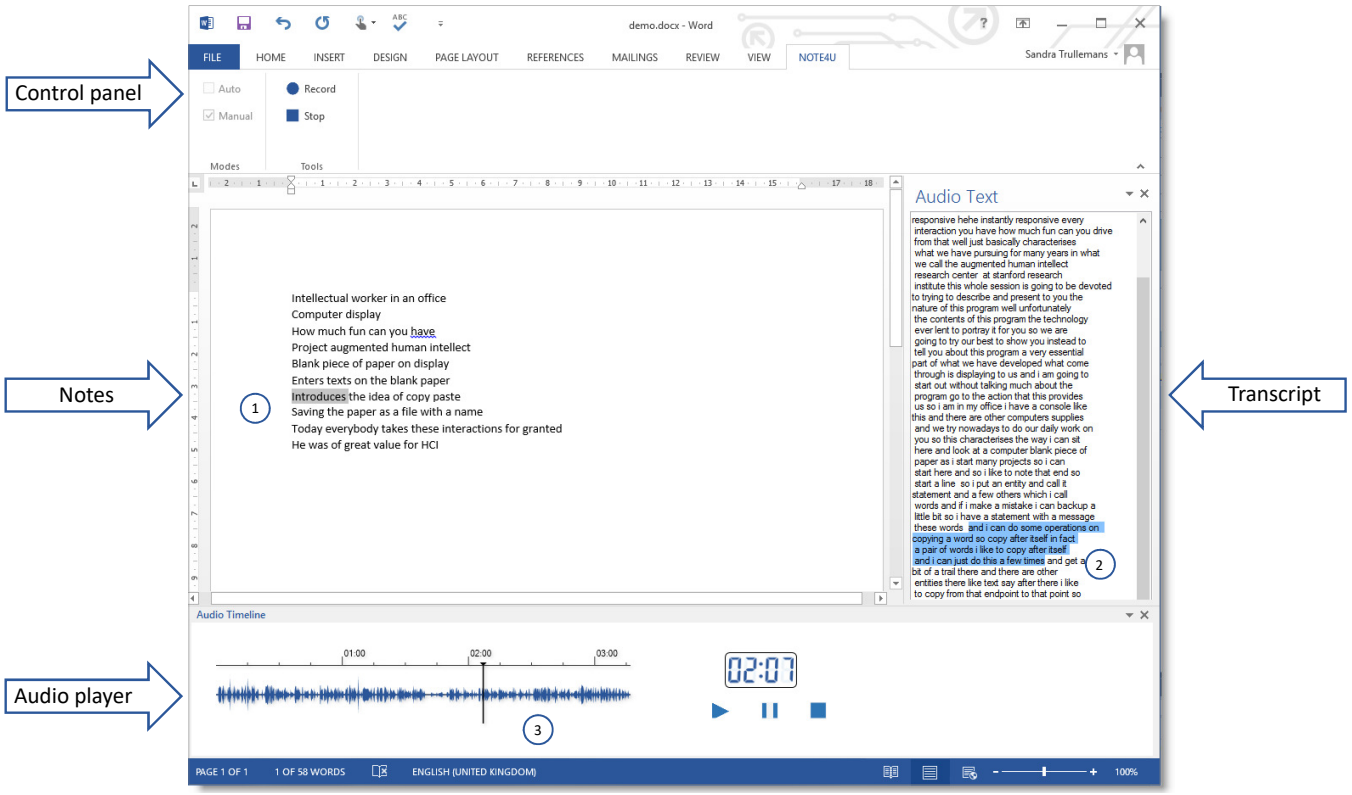


Figure 1: Note4U Microsoft Word add-in. *Control panel*: switch between auto (online) and manual (offline) recording and start/stop the recording. *Notes*: regular Word document to take the notes. *Audio player*: controls for the recorded audio fragment. *Transcript*: panel showing the transcript of the audio recording

3.2 Implementation

The Note4U Microsoft Word add-in has been implemented in C#. Currently, the Microsoft Speech Recognition Engine with the corresponding dictation grammar is used for the speech-to-text processing. Note that we have foreseen an interface to easily integrate future speech-to-text frameworks. After the recording phase, the audio, transcript and timestamped notes are forwarded to the Note4U engine which performs the temporal co-indexing. Further, the temporal data, audio recording as well as the transcript are stored as part of the XML structure of the underlying Word document (Office Open XML File Format). This self-contained approach allows a user to move and share their document together with the supplementary recordings. The audio timeline and digital clock have been implemented based on the WPF Sound Visualization Library[2].

4 METHODOLOGY

In the presented study, we investigated how users use text- and audio-based prosthetic memory and which factors influence the used interaction strategies. The used methodology is similar to the methodology used by Kalnikaité and Whittaker [11] in their study inquiring when and why people use prosthetic memory.

The procedure consists of two phases. First, participants were given three note-taking tasks where they had to listen to three audio fragments. In the second phase, participants had to answer questions about the audio fragments at three different time intervals (immediately after listening to the audio fragments as well as 7 and 30 days later). For each of these intervals, we measured the variables discussed below. Participants could use the provided PM but they were also allowed to use their organic memory. The detailed tasks and questions can be found in [22].

4.1 Participants

We recruited 14 participants with the profile of a knowledge worker. Our participants professionally work in domains where note-taking activities have to be performed on a regular basis, including academic staff, employees in the private sector as well as university students. The age of our participants was between 23 and 58 with an average of 37. Further, each participant received a 10 Euro shopping gift card for an online store given that they participated in all three sessions (one for each time interval). We are aware of the relatively small number of participants and therefore used non-parametrised tests for the analysis of the results.

[2]https://wpfsvl.codeplex.com

4.2 Prosthetic Memories and Their Interactions

For our study we have used Note4U with its audio and transcript PM allowing users to apply the following basic interaction strategies observed as the main interactions in previous work [19, 30]:

- **Navigation:** Users jump to a place in the transcript or audio recording by selecting a part of their notes.
- **Skimming:** Users parse their notes or transcript, or move through the audio recording without using another prosthetic memory as an entry point.

We have manually entered the audio fragments and transcripts used in this study into the Note4U solution. In addition, the transcripts were manually extracted from the audio fragments in order to prevent any bias from potentially inaccurate automatic transcription. Nevertheless, the transcripts do not contain punctuations and are just reflecting what people said in the audio fragment. An example of a transcript is highlighted with (2) in Figure 1. In the future, we might deploy a field user study with automated extracted transcripts to determine the influence of errors in the transcript on the use of prosthetic memory.

4.3 Research Method

The study took place in a controlled laboratory setting and followed a within-subjects research design. Since we were interested in the interaction differences of audio and transcript PM, we defined three cases. In the first case, participants were allowed to use their notes and audio PM only. In the second case, they could use their notes and the transcript PM only and in the third case they were allowed to use all three prosthetic memories (i.e. notes, audio and transcript). We defined the first two cases in order to measure and observe the interaction strategies and usage of both audio and transcript PM in isolation. The third case allowed us to investigate how these prosthetic memories are used together and in which situations users only use one. For the first two cases, we either removed the transcript PM or the audio PM from the Microsoft Word graphical user interface.

4.4 Audio Fragments and Questions

We have used three audio fragments (i.e. each participant was given another audio fragment for each case). The three fragments are informal interviews where the interviewer calls a person to gain more information about a topic. These topics were about an event in Brussels, giving Dutch classes to prisoners and someone who runs a marathon every day. An example of a part of an audio fragment (English translation) is given below:

"*And there is a special Brusselicious tram driving across Brussels right? Yes, the tram experience and there is actually a two hours dinner on the tram while driving through Brussels. The dining is also very special. All menus are made by two star chefs, selected from a pool of two Walloon, two Flemish and two Brussels chefs. The menus also change every season and actually it is about dining and exploring Brussels at the same time.*"

Note that Dutch audio fragments were given to our native Dutch-speaking participants. Although we could have used English audio fragments, it might have introduced some bias since the participants were all non-native English speakers. We have selected the audio fragments and constructed the questions in collaboration with two teachers who teach Dutch classes at high school level. The chosen fragments have been taken from the open access Taalblad[3] teaching platform. We opted for this procedure in order to have audio fragments which were clear to understand and did not contain any dialect or difficult to understand terminology. Finally, the fragments had an average duration of 4:45 minutes.

For each audio fragment, we constructed three sets of questions, each containing six questions. Furthermore, each set of questions consisted of three gist and three verbatim questions. For instance, a gist question for the above-mentioned audio fragment was "*What is the tram experience about?*" and a verbatim question was "*Where are the six two star chefs coming from?*". The full transcripts and question sets can be found in [22].

4.5 Procedure

The study took place in two phases. In a first phase, participants were asked to take notes while listening to the three audio fragments. For each audio fragment they used a new Microsoft Word document. Before the note-taking task, they were notified which prosthetic memory (i.e. audio only, transcript only or both) they could later use in the second phase. The second phase consisted of answering a set of questions for each audio fragment the day the note-taking took place as well as 7 days and 30 days later. The questions were presented to the participants via a web-based form. During each interval, participants could make use of the prosthetic memory which was allocated to the particular audio fragment at the beginning. For example, if a participant was given fragment one in the case where they could only use audio PM, they received questions about fragment one and they could only use audio PM in the second phase. In order to prevent bias from the audio fragments and questions, we have used a counterbalanced design with the audio fragments, question sets as well as the order of prosthetic memory use as factors.

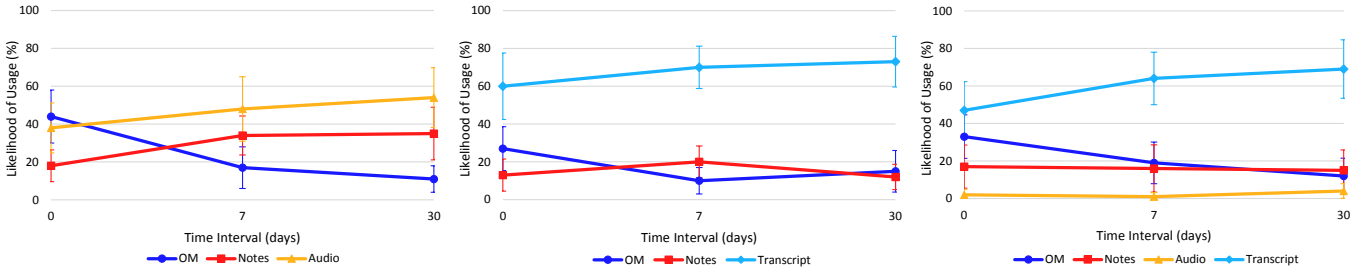
Before starting the first phase of the study, participants were introduced to Note4U. Similar as to the study procedure, they first listened to an audio fragment and took notes followed by answering two questions for each of the three prosthetic memory cases. They only started the study when they succeeded in this task and when they did not have any remarks or questions left. Note that we did not foresee any additional training sessions during the retention intervals since the Note4U application is straightforward to use and provides basic transcription and audio functionality in a familiar note-taking environment. We also did not observe any issues concerning memorability or learnability after the training session.

After the two study phases, we performed an open interview with each participant in order to gain more insights about their personal interaction behaviour. Finally, each participant took part in a formative usability evaluation of Note4U based on the USE-questionnaire [16]. Note that we did not set up a separate study for the evaluation since the evaluation methodology would require a similar approach as the presented laboratory user study.

4.6 Dependent and Independent Variables

In each case (i.e. audio only, transcript only and the combined case), we measured certain variables:

[3] http://www.taalblad.be/zoekresultaat



(a) Audio only case (b) Transcript only case (c) Audio and transcript combined case

Figure 2: Likelihood of usage for OM, notes and PM for the three cases including audio only, transcript only as well as both audio and transcript over the three time intervals. The error bars define the confidence interval (CI).

Usage: During the observations, for each time interval we noted down which prosthetic memory participants used to answer a question. In case that a participant used more than one prosthetic memory to answer a question, we counted both as being used. Nevertheless, we made a distinction between successful and failed use. In order to make this distinction, we applied the think-aloud method where users told their strategy during the recall tasks.

Accuracy: Similar as to the methodology of Kalnikaité and Whittaker [11], we graded each answer against a correction scheme. Each answer was assigned a score between 0 and 5 based on the contained keywords and context. For example, an answer which contained all keywords (or synonyms) and the necessary context received a score of 5. When all keywords were given but no context, it received a score of 4 and scores between 1 and 3 were assigned to answers with parts of the keywords or context. A score of 0 was given whenever participants did not fill in an answer or if all keywords and context were wrong. The accuracy of a prosthetic memory after a certain time interval is the average of these scores.

Efficiency: The efficiency of a PM was measured by monitoring the time (in seconds) that users spent on answering a question when using a particular PM. We used the time they effectively used a particular PM to define the efficiency.

Confidence: The study of Kalnikaité and Whittaker shows that users make more use of a prosthetic memory when they are less confident that their answer retrieved from their organic memory is correct. Therefore, we can also investigate these results for audio PM and transcript PM. Confidence is measured by asking participants for each question they answered how confident they are about their answer when only relying on organic memory. The confidence level is indicated by means of a 5-point Likert scale.

5 RESULTS

The results are structured based on our three research questions RQ1, RQ2 and RQ3. For each research question, we elaborate on the quantitative and qualitative findings from the logged data, observations and open interviews. In order to statistically analyse the results, we have used the Wilcoxon signed-rank test (z) with a Bonferroni correction and the Friedman test (X^2) to see a significant difference in distribution of factors. The statistical results are considered significant if $p < 0.05$.

RQ1: Is there a difference in the way users interact with notes, transcripts and audio recordings and do the interaction strategies change when time elapses?

Overall Usage

The results of the usage of the provided prosthetic memories are illustrated in Figure 2. Both the use of audio and transcripts (i.e. regardless the used interaction strategy) follow a similar pattern when they are used in isolation. Moreover, we could not find a difference in the use of audio ($X^2(2) = 3.68, p = .159$) and transcripts ($X^2(2) = 3.01, p = .216$) over the time interval of 30 days. We also observed that audio is significantly less likely to be used than transcripts after 7 and 30 days as shown in Table 1. In the audio and transcript case where users are given the option to use both prosthetic memories, as expected we see that audio is used much less than in the case where transcripts were not available. In addition, in 66% of the cases where audio was used it was due to a previously failed search in the transcript. This behaviour indicates that transcripts seem to have an advantage over audio.

	Same day	7 days	30 days
Audio only - Transcript only	$z = -1.6$ $p = 0.113$	$z = -2.7$ $p = 0.007$	$z = -2.4$ $p = 0.017$
Audio only - Combined	$z = -3.2$ $p = 0.001$	$z = -3.3$ $p = 0.001$	$z = -3.3$ $p = 0.001$

Table 1: The Wilcoxon Signed Rank value (z) and significance (p) with regard to the statements

Finally, when audio and/or transcripts are provided to the user, the use of notes stays stable when time elapses (Audio only case ($X^2(2) = 6.38, p = .061$); Transcript only case ($X^2(2) = 3.13, p = .344$); Combined case ($X^2(2) = 0.49, p = .976$)). This is in line with previous findings where it has been observed that the use of notes, dictaphone and ChittyChatty did not increase over time intervals [11]. We can conclude that prosthetic memories are significantly used from the first day on and their usage does not change over a period of 30 days.

Different Ways of Interacting with Prosthetic Memory

Our results show that navigation is the most used interaction strategy as illustrated in Figure 3 and Table 2. We can also notice that when transcripts are used, the interaction strategy changes over the time intervals and after 30 days, navigation and skimming are used more or less equally.

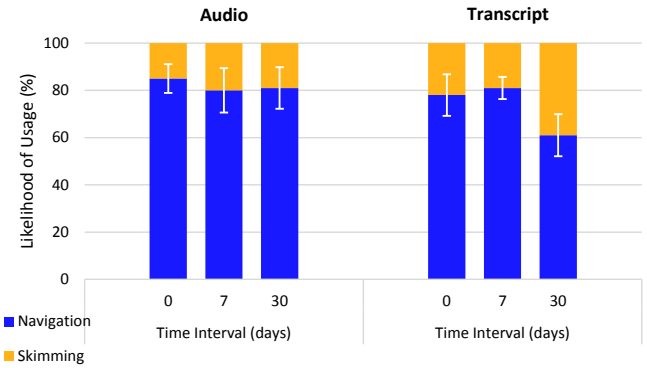


Figure 3: Mean (SEM) of the used interaction for the audio only and transcript only case

During open interviews, participants mentioned the issue that notes did not provide enough contextual cues anymore for using navigation during the last session after 30 days. In cases where they could not find an appropriate keyword in their notes, they thought that it was more efficient to just skim the transcript rather than trying out various keywords for navigation. In contrast, when only audio was available, they tried out various keywords until they found the necessary information. As one participant mentioned "*It will take me too long to play the audio recording forward and backward so I tried to determine the position of the answer by navigating from different parts of the story. But my technique has cost me a lot of (mental) effort.*" By investigating different positions in the audio recording, they tried to catch the contextual overview of the story in the audio environment. In addition, most of the participants who applied this approach mentioned the extra cognitive effort that the search imposed compared to the use of the transcript PM. In general, navigation is a last resort interaction strategy after 30 days in the case that a prosthetic memory does not foresee a contextual overview as it is the case with audio PM.

	Same day	7 days	30 days
Audio skimming - navigation	$z = -3.8$ $p = 0.003$	$z = -2.6$ $p = 0.009$	$z = -2.5$ $p = 0.013$
Transcript skimming - navigation	$z = -2.0$ $p = 0.045$	$z = -2.8$ $p = 0.005$	$z = -0.9$ $p = 0.325$

Table 2: The Wilcoxon Signed Rank value (z) and significance (p) with regard to the statements

Different Purposes of Prosthetic Memory Use

During our observations we identified that participants used a prosthetic memory even if they had written down an answer and were confident about its correctness. Therefore, we monitored this behaviour and defined two purposes for using a prosthetic memory:

Recalling: Users do not know the answer and make use of a PM to find it. We measured the purpose for recall by annotating the cases were users did not or only partly know the answer and had doubts about correctness of their answer.

Verifying: Users know the answer but make use of notes, audio or transcripts in order to verify their answer. In this case users might or might not be confident about their answer. We measured the purpose of verifying by annotating the cases were PM was used after an answer had been written down.

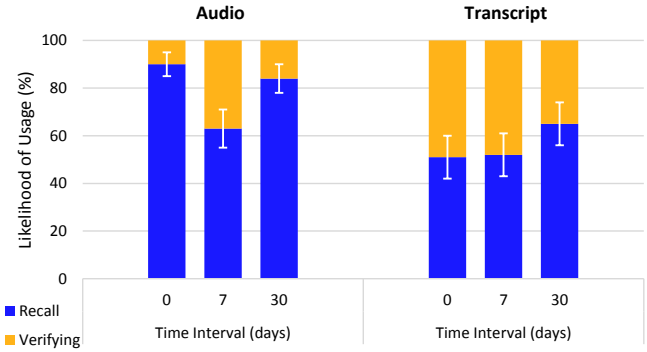


Figure 4: Mean (SEM) of the purpose for PM interaction for the audio only case and transcript only case

Our results shown in Figure 4 indicate no general difference in the purpose of using the transcript PM ($X^2(5) = 7.53, p = .184$). Participants used the transcript for recalling and verifying to a similar degree and their behaviour did not change when time elapsed. In contrast, the purpose of use of audio PM shows that audio is more often used for recalling than for verifying ($X^2(5) = 36.63, p = .000$). Our open interview results indicate that when audio is used for verifying purposes, users make a cost-benefit analysis and only use audio for verification if they have somewhat an idea where the information is situated within the audio fragment and are not very confident about their answer from organic memory. For the use of transcripts, the cost-benefit analysis is less of a burden since users where normally confident about the efficient retrieval of the corresponding information.

RQ2: Is there a difference in accuracy and efficiency between audio and transcript PM?

Accuracy

The Friedman test shows no significant differences in accuracy for audio and transcript use ($X^2(5) = 5.51, p = .358$). The overall mean of the accuracy for audio is 4.17 ($SE = 0.082$) and for transcripts it is 4.058 ($SE = 0.129$). The accuracy did also not change over time which means that the previously mentioned change in strategy when using transcription (i.e. going over from navigation to skimming after 30 days) does not influence the accuracy and can be seen as successful. In addition, during our observations we have seen that users often first used navigation to jump to a position in the

audio recording. Although we start the audio playback 5 seconds before the co-indexed timestamp, users moved the timeline cursor slightly back on the timeline before starting to listen to the audio. The amount of seconds that they moved the cursor back depended on how much additional information they preferred to re-listen. By applying this strategy, users were given the same contextual information as with the transcript. For example, a participant stated the following during the open interview sessions: "*I always moved the cursor a centimetre to the front since I need to know what was said before. It is then easier to formulate my answer. Anyway, it was just a hack to get the same what you get for free with the transcript.*"

	Same day	7 days	30 days
Skimming audio - transcript	$z = -1.3$ $p = 0.180$	$z = -1.1$ $p = 0.285$	$z = -0.4$ $p = 0.686$
Navigation audio - transcript	$z = -1.8$ $p = 0.071$	$z = -0.4$ $p = 0.694$	$z = -1.7$ $p = 0.084$

Table 3: The Wilcoxon Signed Rank value (z) and significance (p) with regard to the statements

Efficiency

Similar to the accuracy, we did not find significant differences in the efficiency between audio and transcripts for both, skimming and navigation as shown in Table 3. The overall efficiency of audio and transcript for the time intervals is highlighted in Figure 5. Although there is no significant difference in efficiency when using audio or transcript PM, users do have the subjective feeling that audio is less efficient than transcripts.

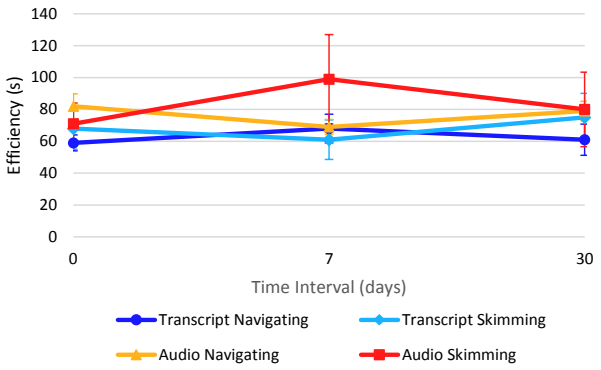


Figure 5: Mean (SEM) of efficiency for navigation and skimming when using audio and transcript

During the open interviews, participants mentioned the subjective inefficiency of audio as a reason for not using it when both prosthetic memories are available. We can conclude that the accuracy and efficiency of a prosthetic memory does not play a crucial role in the choice of a prosthetic memory.

RQ3: Do factors such as confidence influence the way users interact with PM?

Our results do not show a significant difference in the use of the transcript or audio PM for each level of confidence. Moreover, the cost-benefit analysis which users do when using the audio recording for verifying purposes does not take into account their confidence about the answer. Finally, during the open interviews we have asked participants to give us examples from their own private or working environment where they are confident but would still apply a verification strategy. From these discussions we can conclude that there is a significant trade-off between costs (in terms of time) and accuracy when placing the results in practice. Moreover, participants would do the verification effort when information is of critical nature. Nevertheless, our results show that if users are unaware of time costs, they frequently use prosthetic memory for verifying purposes even if they are confident about their answers.

5.1 Usability Evaluation of Note4U

Finally, we present a formative evaluation of the Note4U prototype that we have developed for our study. The evaluation was done after the last session of our study. All participants filled in the USE questionnaire [16] which measures the usefulness, ease of use, ease of learning and satisfaction on a 7-point Likert scale. The results from the USE questionnaire show a positive evaluation for all four factors as shown in Figure 6. Participants mentioned that it was nice to have the audio recordings and transcripts integrated in Microsoft Word since they are familiar with the environment and that they would not easily switch to another note-taking tool.

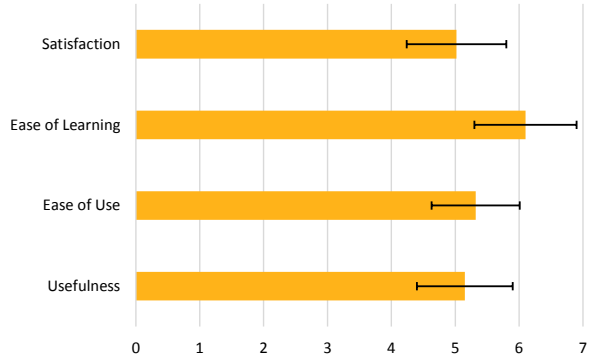


Figure 6: Mean values of the four usability factors with error bars representing the standard deviation

6 DESIGN IMPLICATIONS

The results presented in this paper further allow us to derive the following three design implications:

Designing for Individual or Combined Audio and Transcript
Whenever audio is integrated in the design of a prosthetic memory application, we must take into account that there is a difference in prosthetic memory usage when audio PM is provided together with a transcript. In case that audio is not augmented with a transcript, our results indicate a similar behaviour as for the interaction

with transcripts where navigation is used as the main interaction strategy. However, when time elapses, navigation becomes more of a last resort interaction since notes loose their context. Therefore, when designing prosthetic memory for audio or transcript only, it is important to address the fail of navigation after longer time periods. A design alternative might be to augment notes with contextual information enabling users to keep using their notes for navigation. For example, notes could be augmented with information about who was speaking when the note was taken or the topic of the note. On the other hand, when a user interface offers audio combined with transcripts, we have seen that audio is rarely used due to a lack of context in terms of the surrounding audio information. In this case, designers should focus on the design of the transcript component and see audio as a secondary on-demand prosthetic memory which is mainly used when search in the transcript fails.

PM is Not Only Chosen Based on Efficiency and Accuracy
Previous work on the use of prosthetic memories shows that the choice of a prosthetic memory depends on the level of its efficiency [11]. In contrast to the subjective perception of audio PM being slower than transcript PM, our results indicate that audio and transcript PM offer the same efficiency and accuracy. However, transcript PM has been used significantly more often than audio PM. Our qualitative results show that the mental effort required to retrieve contextual information about the retrieved information from a prosthetic memory plays a significant role in the choice of a prosthetic memory and that this mental effort is lower for transcript PM. When navigating, users skim a few sentences before and after the target position in the transcript in order to get the overall context. In order to retrieve the same information from audio, users first have to make a mental model of the range which might contain interesting contextual information before processing the information itself. It is this additional mental effort for audio PM—and not its efficiency—which makes audio a secondary prosthetic memory in user interfaces offering audio recordings together with their transcripts. Future prosthetic memory design could provide features to reduce the mental effort required for retrieving contextual information from audio. For example, we could indicate the time when the presenter started with an explanation of a certain topic during a lecture. For both, audio and transcripts, additional contextual information might be provided in order to reduce the mental effort.

Designing for Recall and Verification
Besides the use of prosthetic memory for recall, we have seen that prosthetic memory is also used for verifying information retrieved from organic memory. When designing applications offering a transcript component, we should therefore take this verification behaviour into consideration. Although there are no differences in how users currently use prosthetic memory for both purposes, we can facilitate the interaction with transcripts for verification purposes. For example, if a user is writing a report about a topic discussed in a previous meeting, we might analyse the report and notify the user when detailed information is missing. In contrast to transcripts, audio PM is used much less for verification. When designing for audio PM, we have to take into consideration that users perform a more in-depth cost-benefit analysis than for recall purposes. From our results, we see that factors such as the awareness of the position of the information to be verified and a user's level of confidence play an important role in deciding whether audio PM should be used for verification.

7 DISCUSSION AND FUTURE WORK

The results of our longitudinal user study show that when audio and transcript PM are used on their own there is no overall difference in use at the different retention intervals. In contrast, when audio and transcripts are combined in a single user interface, audio PM is mainly used when transcripts fail. While previous research would argue that this is caused by the fact that audio PM is slower than transcript PM, we could not observe a difference in efficiency between the two PMs. However, our qualitative results show that the preference for transcript PM lies in the fact that audio PM asks for an additional mental effort when contextual information should be extracted. It is therefore important that we improve the accessibility of the corresponding contextual information. Further, a lack of positional overview for audio PM contributes to the increased mental effort.

The use of contextual and spatial cues has also been observed in re-finding behaviour in the field of Personal Information Management (PIM) [25]. Furthermore, in PIM applications such as PimVis [24] or SOPHYA [8], users re-find documents by using navigation. They select a search result in the PIM application and as a result the corresponding digital or physical artefact containing the document is shown. For example, in digital space the File Explorer might open the folder containing the searched document whereas in physical space a filing cabinet might be augmented with LEDs in order to highlight the file containing a printed document [23]. In these settings, we can inform the design by pointing to the fact that a spatial reference is not enough and users might want to have additional contextual information as observed when navigating audio and transcript PM. Note that our results might also inform the design of other environments where the same information is provided in textual as well as auditive form, including news websites, websites with recipes, blogs or webcasts. When re-visiting these environments for recall or verification, it is likely that audio will not be used.

8 CONCLUSION

We have presented a longitudinal user study investigating the differences between audio and transcript prosthetic memory in terms of their usage, the applied interaction strategies as well as their accuracy and efficiency. Our results do not confirm the subjective perception that audio PM is less efficient than transcript PM. The main reason why users prefer transcript PM over audio PM is rather the fact that audio demands for a higher mental effort to use the spatial and contextual cues. We have shown that PM is used for verifying information retrieved from OM to a similar degree as PM is used for recall purposes. Finally, we have provided a number of design implications for audio as well as transcript PM and highlighted the applicability of our findings to other environments.

ACKNOWLEDGEMENTS

The research of Sandra Trullemans is funded by the Agency for Innovation by Science and Technology in Flanders (IWT).

REFERENCES

[1] Eytan Adar, David Karger, and Lynn Andrea Stein. 1999. Haystack: Per-User Information Environments. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM 1999)*. Kansas City, USA, 413–422. https://doi.org/10.1145/319950.323231

[2] Juan Pablo Carrascal, Rodrigo De Oliveira, and Mauro Cherubini. 2015. To Call or to Recall? That's the Research Question. *ACM Transactions on Computer-Human Interaction (TOCHI)* 22, 1 (2015). https://doi.org/10.1145/2656211 Article 4.

[3] Ciprian Chelba, Timothy J. Hazen, and Murat Saraçlar. 2008. Retrieval and Browsing of Spoken Content. *IEEE Signal Processing Magazine* 25, 3 (2008), 39–49. https://doi.org/10.1109/MSP.2008.917992

[4] Jim Gemmell, Gordon Bell, Roger Lueder, Steven Drucker, and Curtis Wong. 2002. MyLifeBits: Fulfilling the Memex Vision. In *Proceedings of the ACM International Conference on Multimedia (Multimedia 2002)*. Juan-les-Pins, France, 235–238. https://doi.org/10.1145/641007.641053

[5] Tobias Giesbrecht, Tino Comes, and Gerhard Schwabe. 2015. Back in Sight, Back in Mind: Picture-Centric Support for Mobile Counseling Sessions. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW 2015)*. Vancouver, Canada, 486–495. https://doi.org/10.1145/2675133.2675169

[6] Andreas Girgensohn, Jennifer Marlow, Frank Shipman, and Lynn Wilcox. 2016. Guiding Users through Asynchronous Meeting Content with Hypervideo Playback Plans. In *Proceedings of the ACM Conference on Hypertext and Social Media (HT 2016)*. Halifax, Canada, 49–59. https://doi.org/10.1145/2914586.2914597

[7] Adriana Ispas, Beat Signer, and Moira C. Norrie. 2011. An Extensible Digital Ink Segmentation and Classification Framework for Natural Notetaking. In *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS 2011)*. Pisa, Italy, 231–240. https://doi.org/10.1145/1996461.1996528

[8] Matthew G. Jervis and Masood Masoodian. 2010. SOPHYA: A System for Digital Management of Ordered Physical Document Collections. In *Proceedings of the ACM International Conference on Tangible, Embedded, and Embodied Interaction (TEI 2010)*. Cambridge, USA, 33–40. https://doi.org/10.1145/1709886.1709894

[9] Vaiva Kalnikaité, Patrick Ehlen, and Steve Whittaker. 2012. Markup as You Talk: Establishing Effective Memory Cues While Still Contributing to a Meeting. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 2012)*. Savannah, USA, 349–358. https://doi.org/10.1145/2145204.2145260

[10] Vaiva Kalnikaité and Steve Whittaker. 2008. Social Summarization: Does Social Feedback Improve Access to Speech Data?. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 2008)*. San Diego, USA, 9–12. https://doi.org/10.1145/1460563.1460567

[11] Vaiva Kalnikaité and Steve Whittaker. 2007. Software or Wetware?: Discovering When and Why People Use Digital Prosthetic Memory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2007)*. San Jose, USA. https://doi.org/10.1145/1240624.1240635

[12] Matthew Kam, Jingtao Wang, Alastair Iles, Eric Tse, Jane Chiu, Daniel Glaser, Orna Tarshish, and Jhon Canny. 2005. Livenotes: A System for Cooperative and Augmented Note-Taking in Lectures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2005)*. Monreal, Canada, 531–540. https://doi.org/10.1145/1054972.1055046

[13] Vivi Katifori, Antonella Poggi, Monica Scannapieco, Tiziana Catarci, and Yannis Ioannidis. 2005. OntoPIM: How to Rely on a Personal Ontology for Personal Information Management. In *Proceedings of the Semantic Desktop Workshop (SDW 2005)*. Galway, Ireland, 258–262. http://ceur-ws.org/Vol-175/25_poggi_ontopim_poster.pdf

[14] Dar-Shyang Lee, Berna Erol, Jamey Graham, Jonathan J. Hull, and Norihiko Murata. 2002. Portable Meeting Recorder. In *Proceedings of the ACM International Conference on Multimedia (Multimedia 2002)*. Juan-les-Pins, France, 493–502. https://doi.org/10.1145/641007.641111

[15] Lin-shan Lee, James Glass, Hung-yi Lee, and Chun-an Chan. 2015. Spoken Content Retrieval – Beyond Cascading Speech Recognition with Text Retrieval. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 23, 9 (2015), 1389–1420. https://doi.org/10.1109/TASLP.2015.2438543

[16] Arnold Lund. 2001. Measuring Usability with the USE Questionnaire. *Usability Interface* 8, 2 (2001), 3–6. https://www.researchgate.net/publication/230786746_Measuring_usability_with_the_USE_questionnaire

[17] Kent Lyons. 2003. Everyday Wearable Computer Use: A Case Study of an Expert User. In *Proceedings of the Fifth International Symposium on Human Computer Interaction with Mobile Devices and Services (MOBILE HCI 2003)*. Udine, Italy, 61–75. https://doi.org/10.1007/978-3-540-45233-1_6

[18] Cosmin Munteanu, Ronald Baecker, Gerald Penn, Elaine Toms, and David James. 2006. The Effect of Speech Recognition Accuracy Rates on the Usefulness and Usability of Webcast Archives. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2006)*. Monreal, Canada, 493–502. https://doi.org/10.1145/1124772.1124848

[19] Mukesh Nathan, Mercan Topkara, Jennifer Lai, Shimei Pan, Steven Wood, Jeff Boston, and Loren Terveen. 2012. In Case You Missed It: Benefits of Attendee-Shared Annotations for Non-Attendees of Remote Meetings. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 2012)*. Seattle, USA, 339–348. https://doi.org/10.1145/2145204.2145259

[20] Beat Signer, Moira C. Norrie, Michael Grossniklaus, Rudi Belotti, Corsin Decurtins, and Nadir Weibel. 2006. Paper-based Mobile Access to Databases. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD 2006)*. Chicago, USA. https://doi.org/10.1145/1142473.1142581

[21] Lisa Stifelman, Barry Arons, and Chris Schmandt. 2001. The Audio Notebook: Paper and Pen Interaction With Structured Speech. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2001)*. Seattle, USA, 182–189. https://doi.org/10.1145/365024.365096

[22] Sandra Trullemans. 2018. *Enabling and Informing the Design of Cross-Media Personal Information Management Solutions*. Ph.D. Dissertation. Vrije Universiteit Brussel.

[23] Sandra Trullemans, Payam Ebrahimi, and Beat Signer. 2018. Crossing Spaces: Towards Personal Cross-Media Information Management User Interfaces. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI 2018)*. Grosseto, Italy.

[24] Sandra Trullemans, Audrey Sanctorum, and Beat Signer. 2016. PimVis: Exploring and Re-finding Documents in Cross-Media Information Spaces. In *Proceedings of the International Working Conference on Advanced Visual Interface (AVI 2016)*. Bari, Italy, 176–183. https://doi.org/10.1145/2909132.2909261

[25] Sandra Trullemans and Beat Signer. 2014. From User Needs to Opportunities in Personal Information Management: A Case Study on Organisational Strategies in Cross-Media Information Spaces. In *Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries (DL 2014)*. London, UK, 87–96. https://doi.org/10.1109/JCDL.2014.6970154

[26] Nigel G. Ward and Steven D. Werner. 2013. Using Dialog-Activity Similarity for Spoken Information Retrieval. In *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH 2013)*. Lyon, France, 1569–1573. http://www.cs.utep.edu/nigel/abstracts/interspeech13.html

[27] Pierre Wellner, Mike Flynn, and Maël Guillemot. 2005. Browsing Recorded Meetings with Ferret. In *Proceedings of the First International Workshop Machine Learning for Multimodal Interaction (MLMI 2004)*. Martigny, Switzerland, 12–21. https://doi.org/10.1007/978-3-540-30568-2_2

[28] Steve Whittaker, Julia Hirschberg, Brian Amento, Litza Stark, Michiel Bacchiani, Philip Isenhour, Larry Stead, Gary Zamchick, and Aaron Rosenberg. 2002. SCANMail: A Voicemail Interface That Makes Speech Browsable, Readable and Searchable. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2002)*. Minneapolis, USA, 275–282. https://doi.org/10.1145/503376.503426

[29] Steve Whittaker, Patrick Hyland, and Myrtle Wiley. 1994. FILOCHAT: Handwritten Notes Provide Access to Recorded Conversations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 1994)*. Boston, USA, 271–277. https://doi.org/10.1145/191666.191763

[30] Steve Whittaker, Simon Tucker, Kumutha Swampillai, and Rachel Laban. 2008. Design and Evaluation of Systems to Support Interaction Capture and Retrieval. *Personal and Ubiquitous Computing* 12, 3 (2008), 197–221. https://doi.org/10.1007/s00779-007-0146-3

[31] Lynn Wilcox, Bill Schilit, and Nitin Swahney. 1997. Dynomite: a Dynamically Organized Ink and Audio Notebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 1997)*. Atlanta, USA. https://doi.org/10.1145/258549.258700

[32] Ron Yeh, Chunyuan Liao, Scott Klemmer, François Guimbretiére, Brian Lee, Boyko Kakaradov, Jeannie Stamberger, and Andreas Paepcke. 2006. ButterflyNet: A Mobile Capture and Access System for Field Biology Research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2006)*. Montreal, Canada, 571–580. https://doi.org/10.1145/1124772.1124859

[33] Dongwook Yoon, Nicholas Chen, François Guimbretiére, and Abigail Sellen. 2014. RichReview: Blending Ink, Speech, and Gesture to Support Collaborative Document Review. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST 2014)*. Honolulu, USA, 481–490. https://doi.org/10.1145/2642918.2647390