Vocal Sticky Notes: Adding Audio Comments on Paper Documents

Quentin Bonnard

CRAFT, EPFL

CE1530, Station 1

1015 Lausanne, Switzerland

quentin.bonnard@epfl.ch

Andrea Mazzei

CRAFT, EPFL

CE1530, Station 1

1015 Lausanne, Switzerland

andrea.mazzei@epfl.ch

Frédéric Kaplan

CRAFT, EPFL

CE1530, Station 1

1015 Lausanne, Switzerland

frederic.kaplan@epfl.ch

Pierre Dillenbourg

CRAFT, EPFL

CE1530, Station 1

1015 Lausanne, Switzerland

pierre.dillenbourg@epfl.ch

Abstract

In this paper we present a tool to annotate paper documents with vocal comments. This tool does not require specially processed documents, and allows natural and simple interactions: sticking a note to add a comment, and place an object on it to listen to the record. A pilot experiment in which teachers used this tool to annotate reports revealed that vocal comments require an extra effort compared to writing. We discuss future work that could either fix or take advantage of this extra effort.

Keywords

Paper-based interface, document recognition, vocal comments, pedagogical script

ACM Classification Keywords

H.5.2 User Interfaces: Interaction styles, H.5.3 Group and Organization Interfaces: Computer-supported cooperative work

General Terms

Paper-based interface, document recognition, vocal comments, pedagogical script

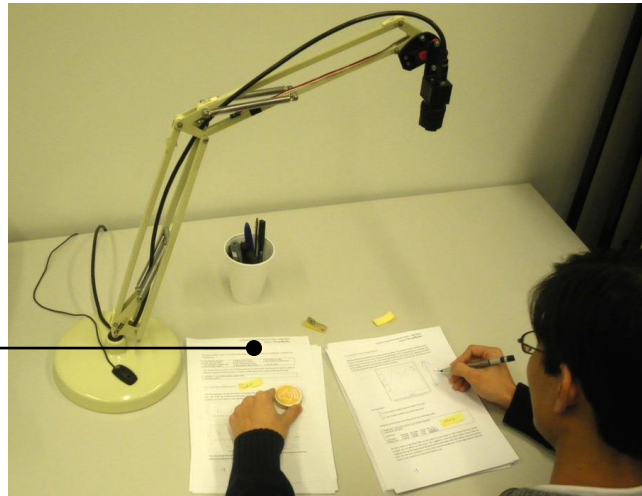Close-up on a sticky note (left) and the *audio controller* (right)

figure 1. Our tool consists of a camera monitoring the work area, and an omnidirectional microphone (on the base of the lamp).

Introduction

Paper documents are still commonly used for a wide variety of educational activities and are now likely to have a digital counterpart. The coexistence of these two forms shows that they both have complementary advantages. On the one hand, physical documents can be manipulated, navigated and annotated more freely. Paper is cheaper, lighter and more autonomous than any mobile technology. On the other hand, digital documents are easier to archive, search, duplicate or share. They can be animated, interactive and hyperlinked. Linking physical and digital documents gives the possibility to accumulate their respective advantages. It is easy to imagine that providing new ways to link the physical and digital version of the same document could enrich current practices.

In our case we propose a tool, shown on Figure 1, which recognizes a page and detects the position of sticky notes: the users can not only write and navigate paper in a natural way, but also associate vocal comments to the sticky notes. These vocal comments exist in the digital world, and can be used in other digital objects, such as a website proposing the digital document and the corresponding vocal annotations.

We use sticky notes to leave a physical marker on the page. It can be removed, as opposed to a stamp or highlighter. We follow the objectives of Klemmer and his colleagues [1] of augmenting an activity without forcing an adoption: it is still possible to leave a written comment on a sticky note. Adding a vocal annotation should be as simple as adding and removing a sticky note. The hardware consists of a consumer webcam on a desk lamp, which subtly integrates on a desktop. We also use sound feedbacks in order to avoid the need of a too invasive screen or projector.

We view this tool as a scriptable artifact [2], i.e. an interactive system (not necessarily a computer), which allows the implementation of a pedagogical scenario designed to trigger productive interaction.

Vocal Sticky Notes

A user can use vocal sticky notes while reading a printed document. When she wants to annotate a page, she places it on an area monitored by our tool. The system notifies the recognition of the page with a sound. Then, the user can place a sticky note on the part of the page she wants to annotate. The system notifies the detection of the sticker with a sound, indicating to the user that she can start talking. Then, the user records her vocal comment, signaling the end

of the recording by placing on a page a special physical token, the *audio controller* (e.g. a monochrome icon printed on a tangible object). The system immediately replays the recorded vocal comment. If it is not satisfactory for the user, it can be erased by simply removing the sticky note from the page. When not recording, users can replay an existing vocal sticky note by placing the *audio controller* over it.
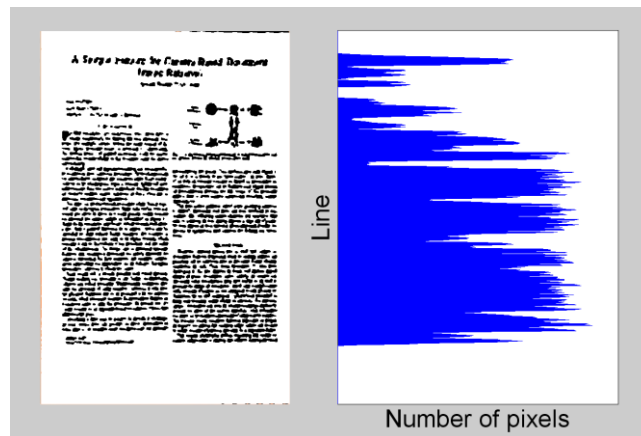

figure 2. Example of the projection profile (right) used as the signature of the image of a document (left).

This scenario relies on two technological components, which we describe shortly hereafter: page identification and sticky notes detection. The easiest way to achieve page identification would be to use visual markers, such as (2D) barcodes. However, this restrains the usage to specially designed documents: if the marker was not included while printing, the document will not be recognized. Various computer vision-based approaches have been developed to recognize documents using a simple camera [e.g. 2]. We compute a simple signature

based on the projection profile of a document. In a nutshell, we preprocess the image, detect the bounding box and orientation of the page, keep only the edges, and count the pixels on each line. Figure 2 shows an example of the signature of a page. To compare and match two signatures, we consider them as signals and compute their maximum cross-correlation. Intuitively, this corresponds to the best offset to match the variation of two signals relatively to their average.

Following the signature computation, extract the sticky notes by a color subtraction approach. Pixels too different from background and foreground colors of the document are considered as candidates. They are clustered according to their position and color. Temporally stable clusters corresponding to the size and color of sticky notes are labeled as sticky notes. We use the same process to detect clusters of another color, labeled as the *audio controller*.

Pilot Experiment
An annotation can have different functions: feedback, personal comment, bookmark, summary, paraphrasing, etc.; there are various motivations behind using audio. For example, private annotations can be more fluent with free speech; paraphrasing can be more effective if the medium changes (written to oral). In our case, written feedbacks in limited space are hard to read and understand compared to direct speech. Recorded speech is obviously more persistent than a face-to-face meeting, which decreases the risk of being forgotten.

We applied our tool to document reviewing in a pedagogical context. We designed a pilot study with 2 teachers and 25 students of a real, master level course. The students, in groups of 3 or 4, had to write a 10

pages report on their statistical analysis of an experiment. Traditionally, the teachers would correct the reports and summarize their feedback in an email. We asked teachers to give this feedback with vocal comments using our tool.

A qualitative analysis of this study highlighted an important difference between written and vocal comments: vocal annotations have to be produced in one shot, while writing allows for pauses even in the middle of sentences. The complete vocal comment has to be clearly and fully formulated in the users' mind before starting to express it. For this reason it can be argued that the elaboration of vocal annotations requires more cognitive effort than for written comments. This is especially true in the context of this experiment, as the correction of these reports is a challenging task: the content is technical, and the teacher-student relationship implies formal annotations.

Discussion
Preliminary results from our pilot study suggest that vocal sticky notes make the effort of the reviewer bigger than for written ones. This could be seen as an obstacle for using this tool. We could reduce this effort by designing a more sophisticated *audio controller* that could emulate the possibility of written annotations to be stopped and resumed. We could also consider application domains for which annotations are used in a less formal way. For example, as vocal bookmarks to summarize paragraphs on a book would remain private, they could be produced in more "relaxed" manner.

A different approach to deal with this extra effort is to actually use it as a relevant pedagogical feature. This corresponds to the goal of Collaborative Learning to foster knowledge by triggering learner's interaction. Practically, we can imagine a scenario where students have to produce a document, review their peers' work with vocal comments, and use the feedbacks for improvement. In fact, the effort dedicated to the elaboration of the comments can trigger an internalization of the reviewing capacity. This could provide students with a greater ability to improve their own work in other contexts. This last point could be especially relevant for pedagogical scripts where vocal elaboration has been shown to be crucially important for learning, for example for improving reading comprehension (e.g. reciprocal tutoring scripts where students have to produce periodically a vocal summary of what they just read [4]).

References
[1] S. R. Klemmer, J. Graham, G. J. Wolff, and J. A. Landay. Books with voices: paper transcripts as a physical interface to oral histories. *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM, 2003, pp. 89–96.

[2] F. Kaplan, and P. Dillenbourg. Scriptable classrooms in K. Makitalo-Siegl, J. Zottmann, F. Kaplan, and F. Fischer (eds), Classroom of the future: Orchestrating collaborative spaces. Sense Publishers, 2010.

[3] T. Nakai, K. Kise, and M. Iwamura. Use of affine invariants in locally likely arrangement hashing for camera-based document image retrieval. *Document Analysis Systems VII*, vol. 3872/2006, pp. 541–552, 2006.

[4] A.S. Palincsar, and A.L. Brown Reciprocal Teaching of Comprehension-Fostering and Comprehension-Monitoring Activities. *Cognition and Instruction*, 1 (2), pp. 117-175, 1984.