



Lecture instrumentation based on synchronous speech transcription

Vincent Bettenfeld, Christophe Choquet, Claudine Piau-Toffolon

► To cite this version:

Vincent Bettenfeld, Christophe Choquet, Claudine Piau-Toffolon. Lecture instrumentation based on synchronous speech transcription. 18th International Conference on Advanced Learning Technologies, Jul 2018, Bombay, India. pp. 11-15. hal-01850439

HAL Id: hal-01850439

<https://hal.science/hal-01850439>

Submitted on 27 Jul 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Lecture instrumentation based on synchronous speech transcription

Vincent Bettenfeld, Christophe Choquet, Claudine Piau-Toffolon

LIUM - IEIAH

Le Mans Université

Laval, France

{name.surname}@univ-lemans.fr

Abstract—Speech transcription can now be achieved quickly enough to generate subtitles in real-time. Research presented in this paper aims to explore the potential of synchronous transcription in instrumenting learning environments, specifically in the context of hybrid lectures. Considering available technologies, we designed a toolset in order to answer the existing needs of students and teachers. This toolset is built around the transcription of the teacher’s speech, which acts as a support for interactions and provides data needed by other functionalities. We present here the different interfaces and their components, as well as the results of the user test carried out in real life conditions.

Learning Web Environment; User Needs Assessment; User Interface

I. PROJECT OVERVIEW

PASTEL is a research project aiming to explore the potential of synchronous speech transcription, and its applications, in specific teaching situations. This technology allows to generate a textual version of the teacher’s or the students’ speech, and immediately use it in order to help them in their teaching or learning activities. Other project interests include an editorial toolset for teachers to save, edit and reuse the material produced or gathered in class. This edited material can be exploited during future sessions of the course, or become the support of another course of different format such as an online class.

The project focuses on pedagogical situations such as lectures and group work, mainly in tertiary education. These situations can take place either in the classroom, online using videoconferencing systems, or in hybrid classes. These situations are complex and need to be instrumented with flexible and adaptive tools, proposing various configurations for diverse actors.

Speech transcription allows human actors to access the textual version of a sentence a few seconds after it was said, and browse the whole text as they wish. This tool can help students solving comprehension problems caused by hearing [5], or allow them to use down time to read again a more complex section. Having access to the recording of a lecture decreases students’ stress: they express trust in these recordings, which leads them to take fewer notes and concentrate on the teacher’s explanations [4]. In addition, their performances can be improved if they are accustomed to using such a transcription in their learning activity [6]. In addition to these functionalities, the transcription can also be

the support of teacher-student interactions. Facilitating these interactions can increase students’ engagement as well as their performances [2].

As part of this project, we will additionally test usability of tools based on other technologies which uses have for now been limited in synchronous pedagogical situations. Course material recommendation and text segmentation are two of the technologies explored to design tools for facilitating students’ and teachers’ tasks, as editing a summary, accessing other complementary resources, etc. Our whole toolset is designed to facilitate the aggregation of content and the replay of the lecture at a later time, as these are assets of learning systems based on recording [1].

The development of this project follows iterative cycles. Through successive iterations, more pedagogical situations will be considered and more features will be added to cover needs derived from these new situations.

II. TARGET AUDIENCE AND USER NEEDS

The first step of the project consisted in studying the existing uses in considered pedagogical situations before any instrumentation, as well as possible uses for tools planned to be developed during the project.

We conducted a study [3] based on a twofold approach, both quantitative and qualitative. First, we built a questionnaire and distributed it to 94 students enrolled in a computer science master’s degree. Sixty-two questionnaires were collected and exploited for this study. Meanwhile, we conducted a semi-directive interview with an associate professor giving lectures as part of the computer science master’s program at Le Mans Université. For facilitation the development and tests of a first prototype of these tools, we’ve chosen to work with a small students’ cohort, already invested in transcription technologies and aware of research activities.

In regard to lectures, among students encountering difficulties, 66% report experiencing downtime at a point during class, as well as difficulties in understanding the class concepts (reported by 61% of them). Only a minority is using external resources given that 75% of students do not search for additional material such as definitions, texts or graphics on the Internet during class.

Considering the course instrumentation based on transcription, students are mainly interested by the possibility to communicate to the teacher which points were not understood. A majority of students also estimate the

synchronous transcription is useful, particularly the possibility of reading and exploiting the text after class. Seventy-two percent of students estimate that the possibility of adding comments to the transcription is moderately to very useful.

The teacher estimates that the access to the transcribed speech is a good thing, albeit student should still be taking notes, as he views writing down content as a part of the learning process. He does not want students to be distracted from their learning activity, so the resource recommendation system usage should be occasional and brief, in order to prevent cognitive overload.

III. SCENARIO

Though diverse pedagogical situations will be considered in the project, current development has focused on lecture situations, either in presence of students or in hybrid configuration. Traditional, face-to-face situations have been considered first due to their relative simplicity. The teacher interacts with his audience as a whole and student-to-student interactions are very limited. Hybrid situations add, on top of this context, the possibility for remote students to attend to the lecture entirely by using their remote interface to watch the professor giving their lecture. Interactions are similar, yet some information is lost: teachers may evaluate the degree of understanding or motivation by reading the students' posture or watching their note-taking activity. This evaluation considerably harder is the given context, students being scattered through different spaces and not necessarily being

filmed. By considering these situations, we must find a way to convey information needed by the teachers during their task. Furthermore, as teachers traditionally use slides as a visual medium for their lectures, we have to adapt to the constraint of diffusing a slideshow to the audience as support of the learning process. Additionally, we concentrate on courses teaching topics complex but known enough so that external resources are available.

In the next section, we describe the interface through two use cases: a student using the tool to attend a distant lecture in real time, and a teacher giving the lecture to students connected to the tool situated both in class and in a remote location.

IV. STUDENT'S TOOL INTERFACE

A. Architecture of the device

The solution was conceived as a Moodle plugin. Moodle is an open source Learning Management System and as such our learning tool can be dropped into existing courses. The course, as well as the data generated while using the plugin, are hosted on a Moodle server. The video recording of the lecture, the corresponding transcription, the notes taken by each student, and the traces of their activity are part of this data.

Students connecting to the PASTEL Moodle plugin integrated to the course, are first asked if they wish to access to the remote interface or the face-to-face interface. As the first one contains more features, we will focus on a student choosing the remote interface. They are then redirected to the corresponding page, and their connection logs are stored for

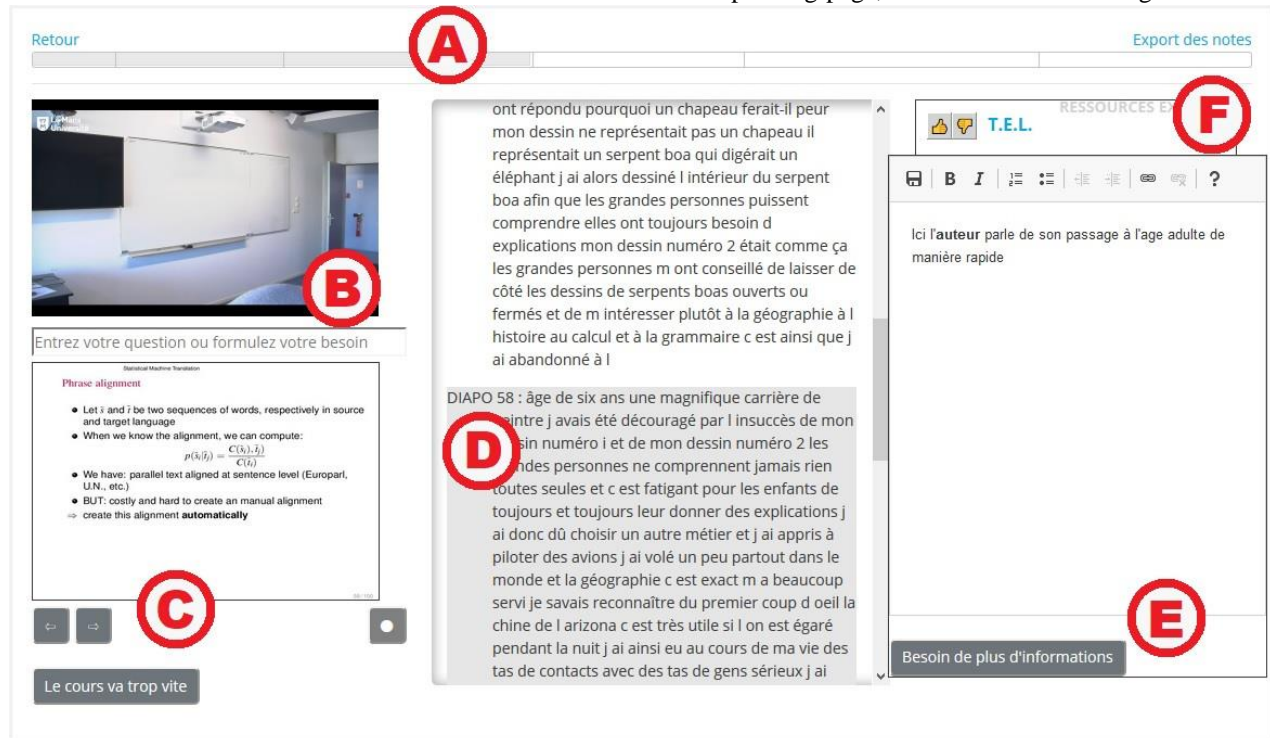


Figure 1. Example of the student's interface: (a) Timeline (b) Webcam flux and question field (c) Slideshow and control panel (d) Transcription (e) Notes & alert panel (f) Material recommendation.

eventual later consultation by the teacher. The page is divided in three columns and a screen-wide timeline.

B. Visualization and interaction with the classroom

To the left (see Fig. 1 B), students can watch the teacher's video stream in real-time. Under this video feed, they can enter their questions or their needs in a text field and submit it by pressing enter (the default label states "Enter your questions or express your needs"). The text is then sent to the teacher, and also analyzed by the resource recommendation system (described later in this section). Since the teacher has no time during the session to read long argumentation, the text field is by design small, to encourage concise demands.

The slideshow display (see C) is positioned in the bottom left, and is by default synchronized with the slide projected in the physical classroom. The students can thus read the slide's content as they listen to the professor's speech. If the content is too small, they can zoom on the slide by clicking on it, opening a pop-up with a larger version of the picture displayed.

If a student was not able to keep pace with an explanation or if a concept was explained too abruptly by the professor, typing it in the open question field could make him lose more time. This is why a button (in the bottom left corner, labelled "The lecture is going too fast") was specifically designed to send an alert to the professor expressing that the lecture's pace goes too fast for the student. An issue drawn from the users' needs assessment was the teacher's inability to read body language or watch directly the note-taking activity of the students, if said students are not situated in the classroom. This button was added so that the professor would be able to control the lecture's pace despite the absence of these usual clues.

C. Transcription display

The transcription display area is located in the center of the screen (see D) and is updated synchronously. It is divided in several paragraphs, each of them being associated to the slide projected at the moment they were transcribed. Each time the teacher projects a new slide, a new paragraph is created and the transcription of the current speech is added to it. Each of these paragraphs can be selected (its background color changes as shown in figure 1-D), which causes the note-taking panel (see E) to open.

Using this panel, two actions concerning the paragraphs are available to the students. If the transcription is not complete enough and they want to write down additional information, the text they type and format is saved in the Moodle plugin database for later use. If the teacher's comment on a given slide is not clear enough for the student, they can send an alert concerning that particular slide using the button labelled "[I] need more information" on the bottom. This allows the teacher to identify which section of his lecture needs further explanation. The material recommendation system also takes in consideration this alert, since the students are expressing their need for more information. The automated system analyses the concepts

being explained or cited in the paragraph to provide relevant information.

D. Recommendation system

This automated system suggests a set of links on the right part of the screen (see F), evolving in real time, redirecting to external resources moderated by the teacher or the course designer. Each of these resources has a title and a very short preview. The first five links suggested are relevant to the topic currently discussed. The list is updated every time the teacher changes the slide displayed. Another three are suggested specially to answer the demand for more information expressed by the students, and are updated as students express their needs by typing questions or clicking the button asking for more information.

Two buttons, one of which shows a thumb up and the other a thumb down, constitute a system of evaluation for each resource. After consulting a resource, the students can evaluate it in regard of its usefulness in the learning activity. In the current version of the plugin, the ratings of the class are only stored on the server for asynchronous use. After class, students can find their favorite resources. The teacher can find feedback on the quality of resources, and potentially decide which one could be integrated in the lecture if the ratings are unanimous. A synchronous usage of these ratings (described in section VI) is planned in future iterations.

E. Navigation system

The material proposed during a lecture can vary depending on time, and the students may want to finish reading a slide before the teacher moves on to the next one. This is why the tool provides two modes of content browsing: synchronous and asynchronous. In synchronous mode, slide display is coordinated in real-time with the projection in the classroom (by the way of a timestamp event). In asynchronous mode, even though data are received and stored, the display keeps focused on a given slide or point of the transcription, allowing the student to spend a longer time reading.

At the top of the screen (see A), a wide timeline of the lecture is displayed. It is a navigation tool designed to facilitate browsing the slideshow and the transcription window quickly. The white bar taking the entire width of the screen symbolizes the current timespan of the lecture, and similarly to a video player, the darker bar represents the duration of time before the moment being browsed. To facilitate access to a given point in time, vertical lines are displayed to show slides changes.

If a given student wants to read again a slide that was showed previously during the first quarter of the lecture, he can do so by clicking on the first quarter of the timeline. The slideshow and the transcription displays are then updated to show the text and slide projected at that moment. If the student didn't click the correct moment on the timeline, for example if the lecture was long and many slides were used, the sections of the timeline guide him and allow to find at what point the previous slide was used, and the relevant paragraph in transcription is synchronously displayed.

Under the slide display is a control panel allowing asynchronous browsing of content already presented during the lecture. The student clicking on the arrows can exploit a short downtime in the professor's speech to go back to a previous slide and view its content again. Once he has finished and wishes to resume the real-time display of the slideshow visible in the classroom, he hits the button with a circle symbol, that then lights red akin to a "recording" icon. The slideshow keeps being updated in real time until he presses one of those buttons again.

When the lecture is over, students can click the link in the top right corner to export a file containing the slides used, the teacher's speech corresponding to each slide as well as their notes. After saving this file, they can go back to the course's Moodle page using the link in the top left corner.

V. TEACHER'S TOOL INTERFACE

If the Moodle user account is listed as a teacher's account for this class, then the user is redirected towards a dedicated page (as shown on figure 2). This page acts as a dashboard designed to assist the teacher in their task, and there is no particular need to share it with the entire class.

A. Feedback of the students' view

First, this screen gives visual feedback of the camera, displaying the same video stream as the students receive (see A). The teachers can accordingly place themselves in the camera's field of view.

In the central column, the resources recommended to students are displayed in real time, in a table (see B). The appearance is more concise as a teacher do not need to rate, or get a preview of the resources they selected. Only the titles are displayed, allowing him to recommend a particular resource to students, or even projecting it during the

presentation if most of the audience is located in the classroom. However, this usage is not extensive, and this list is primarily there to ensure that teachers have feedback on what is displayed to their students.

B. Feedback of the student's activity

Under this table is presented a set of indicators (see C), two of them shown as bar graphs. The first one displays the number of alerts sent by students estimating the lecture's pace was too quick relatively to the total number of students. If this bar grows too large, the teachers can slow down their explanation, to ensure most of the audience can keep up, or interrupt it for explanative feedbacks. The second bar displays the proportion of students browsing previous content, and thus not keeping up with the lecture. If this bar fills up, a teacher may have lost his audience at a previous point, and could react by adapting his speech accordingly. For instance, students could be looking at a big table of data or confused by a definition. If the teacher wants to know precisely where the problem lies, he can then ask student to type it or look at which slide they asked for more explanation.

Three of these slides are listed in a table at the bottom center of the screen (see D). Each row contains the sequence number of the slide dealing with a topic needing more explanations, as well as the number of students requiring them. The three slides numbers shown are those marked by the biggest numbers of students.

The right part of the screen is dedicated to a text feed showing open-ended questions asked by the audience in real time (see E). The most recent question is stacked on top of the list, and is displayed anonymously (even though the sender can be identified by searching the Moodle plugin database, to prevent eventual counter-productive behavior)

The screenshot shows the teacher's interface with several components labeled A through F:

- A**: Webcam feedback showing a classroom view.
- B**: Material recommendation review table.
- C**: Indicators display showing two bar graphs.
- D**: Table of slides needing more explanations.
- E**: Real-time questions feed.
- F**: Link to the slideshow control page.

Table B: T.E.L. Algorithm

Speech Transcription	Convergence
Expectation	Maximization
EM Models	Perplexity

Table D: Étudiants exprimant une difficulté

Diapositive concernée	Étudiants
47	3
45	1
44	1

Figure 2. Example of the teacher's interface (a) Webcam feedback (b) Material recommendation review (c) & (d) Indicators display (e) Real-time questions feed (f) Link to the slideshow control page.

as soon as it is typed and sent. The teacher can ask a question and watch answers be displayed in real time in this area. The question's purpose can be the evaluation of the student's understanding, but can also take the form of organizational discussion if needed.

This interface is designed to assist the teacher throughout the lecture, and is not the teaching support shared with students. By clicking on a link (see F), he can open a second page displaying the slides, destined to be projected to the audience in the classroom. He can operate the slideshow using the mouse or the left and right arrow keys, updating all the students' displays in the process.

VI. USER TEST IN REAL-LIFE CONDITIONS

A first user test took place December 12, 2017 in hybrid configuration, both on Le Mans and on Nantes campuses. The teacher was situated on the former location accompanied by three students while the remaining 7 students were located on the other campus. In both places, participants were filmed, their screen's activity was recorded and they agreed to take part in a post-experiment interview.

Since the recommendation system is in development, material was selected and tagged by the team designing the system, and associated to a particular slide. However, the system did analyze questions asked by students during the session and automatically considered their demand for more information through the interface. Resources fetched were collected from the web in real time.

Collected traces and interviews have still to be analyzed thoroughly, but general tendencies are described in this section. The quality of the microphone impacted negatively on the transcription's fidelity. In order to instrument other real-life situations, we will need to take into account the quality of the recording device. Students' behavior during the lecture was mainly concentrated on the teacher and slides, and their usage of the transcription and resources was more occasional. To navigate between different slides, students mostly used the previous and next slide display buttons. Their usage of the whole timeline to browse content was limited. During the interview, they expressed that manipulating the timeline requires additional thinking, albeit short, they cannot invest time and concentration in. The students also made explicit their frustration about the substance of recommended material. The content was not adapted to their master's degree level: a lot of suggested information was considered trivial at this point of the cursus, such as definitions of basic concepts used during the lecture.

VII. ISSUES AND FUTURE WORK

As several class configurations are possible, and for each of them a great amount of teacher pedagogical practices and student personal preferences can exist, our goal is to propose a flexible interface to match the diversity of practices. This is getting more important as we add functionalities and take into account configurations differing from the studied case.

After asking a question, suggested resources can be picked from a larger pool than the limited subset moderated

by the teacher. Future work on the system include training the system so that it can extract relevant concepts from the questions asked by learners and fetch corresponding online resources

Currently, post-session asynchronous access to course content is limited to an ex-port of the text posted, generated and typed, and to watching the recorded video as it is stored on the server. We plan to develop a browsing system allowing students to retrieve all this content on a single interface where the different elements are synchronized together, and allow them to watch a real-time replay of the lecture. More than controlling only the video, setting the replay to a particular point in time would display the corresponding piece of transcription and resources. Due to the large amount of content, readability and browsing can become a tedious task; a thematic segmentation system will be implemented in order to divide the material into smaller chapters, easier to label and to browse individually.

As for the teacher, post-session use of traces and generated content is for now limited to the display of the contents of the Moodle database. Future work involves the conception of an editorial toolset allowing to retrieve more easily and more clearly the content produced and the feedback collected during a particular lecture. This information could become more useful once summarized as indicators helping the teacher revise or keep unchanged certain parts of the course depending on the feedback. Finally, the opportunity to extract parts of the content, especially video chapters, is a big part of the project. This would allow the teacher to easily produce and share online pedagogical material relevant to a particular theme, greatly reducing the time investment required to set up a SPOC or a flipped classroom configuration.

ACKNOWLEDGMENT

The current work is supported by the ANR project PASTEL <ANR-16-CE38-0007>. The authors would like to thank the others participants of the project for their collaboration.

REFERENCES

- [1] Ashwini, B. V., & Rananavare, L. B. (2014) Enhancement of Learning using Speech Recognition and Lecture Transcription: A Survey. *International Journal of Computer Applications* (0975-8887).
- [2] Caldwell, J. E. (2007). Clickers in the large classroom: Current research and best-practice tips. *CBE-Life sciences education*, 6(1), 9-20.
- [3] Crétin-Pirolli, R., Pirolli, F., Bettenfeld, V. (2017) Analyse des Besoins - Document de synthèse, livrable du projet PASTEL.
- [4] Heilesen, S. (2010). Podcasts for efficient learning, In L. Dirckinck-Holmfeld, et. al. (Ed.), *Proceedings of the 7th International Conference on Networked Learning* (pp. 971-972).
- [5] Ranchal, R., Taber-Doughty, T., Guo, Y., Bain, K., Martin, H., Robinson, J. P., & Duerstock, B. S. (2013). Using speech recognition for real-time captioning and lecture transcription in the classroom. *IEEE Transactions on Learning Technologies*, 6(4), 299-311.
- [6] Shadiev, R., Hwang, W. Y., Chen, N. S., & Yueh-Min, H. (2014). Review of speech-to-text recognition technology for enhancing learning. *Journal of Educational Technology & Society*, 17(4), 65.