

On-Line Certification for All: The PINVOX Algorithm

<http://dx.doi.org/10.3991/ijet.v7i3.2123>

E. Canessa, C. Fonda and M. Zennaro

The Abdus Salam International Centre for Theoretical Physics (ICTP), Trieste, Italy

Abstract—A prototype algorithm: PINVOX (“*Personal Identification Number by Voice*”) for on-line certification is introduced to guarantee that scholars have followed, *i.e.*, listened and watched, a complete recorded lecture with the option of earning a certificate or diploma of completion after remotely attending courses. It is based on the injection of unique, randomly selected and pre-recorded integer numbers (or single letters or words) within the audio trace of a video stream at places where silence is automatically detected. The certificate of completion or “*virtual attendance*” is generated on-the-fly after the successful identification of the embedded PINVOX code by a video viewer student.

Index Terms—Education technology, education innovation, on-line learning, software engineering.

I. INTRODUCTION

The possibility of awarding “*Certificates of Virtual Attendance*” for academic activities is being considered by renowned scientific institutions, such as the ICTP in Trieste, Italy [1]. The idea is to motivate more scholars to watch educational videos and (some of) the scientific material recorded during conferences, workshops and seminars by Nobel prize laureates and others. A certified virtual attendance would allow, for example, ICTP to enlarge its scientific audience, to offer more opportunities to anyone interested in science, and to enforce its institutional mandate to support science and education specially in developing countries. To achieve this goal it is necessary to implement an *ad-hoc* algorithm that allows to verify (to then certify) the real vision of the recorded material by students.

On-line, self-paced education is capturing lots of interest by scholars world-wide —specially through a combination of interactions via chats, VoIP, quizzes, *etc.* (*e.g.*, using Skype, Moodle CMS) and the passive viewing of video lectures (*e.g.*, ICTP recordings of Physics & Math lectures [1] and MIT's OpenCourseWare in many other subjects [2]). New open programs like edX [3] by the partnership MIT-Harvard University, aim at delivering soon on-line course materials by also adding rich-media content to their (MITx and Harvardx) portfolios from other interested Universities. In joining these ventures, scholars will be able to get a certificate upon completion of the on-line courses requirements. The first course by MITx: *Circuits & Electronics*, enrolled about 120 thousand students of which about ten thousand students completed the midterm exam [4].

These novel on-line educational programs that rely on both, watching guided video content and interactive

experiences, raise some new questions though. One of these is to figure out how to assess the progress of a student at a distance. This problem is based on the fact that there is not a common pattern to evaluate students in a variety of disciplines. In natural sciences, for instance, mathematical knowledge and physics-based reasoning is required. In humanities, essays is the method used for demonstrating subject mastery. For clinical case discussions, face-to-face encounters are essential, however, internet interaction in this case enables a more relaxed environment which can greatly benefit the learning process [5].

From the technical point of view there are some important challenges to consider when implementing on-line education with certification. The main concern in on-line assessment tests is *i)* to avoid students guessing from a random itemized selection of answers, and *ii)* to fully identify the candidate presence across the net. Honesty needs to be assumed by default on the student's side. On the other hand, there are no techniques available so far to verify if the vision of a recorded lecture was successfully completed [6,7].

We introduce here a prototype algorithm: PINVOX, “*Personal Identification Number by Voice*” (“*vox*” in Latin) for trusted, on-line certification which ensures that viewers do indeed follow, *i.e.*, listen and watch, a complete video lecture to a reasonable extent. The method is based on the random insertion of pre-recorded integer numbers (or letters or single words [8]) on video lectures. These injected audio PIN codes *i)* are composed of few numbers (depending of the length of each video), *ii)* are unique for every video download, since these are automatically created and inserted on-the-fly, *iii)* are (to be) identified by students when they follow the whole video only, and *iv)* are superimposed into the audio trace of the lecturer at places where (in principle, but not necessarily) silence is detected by the algorithm below some chosen noise threshold in order not to disturb the main contents of the vision. An on-line certificate is generated after the successful identification by the learners of the assigned acoustic PIN injected into the videos.

II. CERTIFICATION BEYOND THE PHYSICAL CLASSROOM

TV broadcasters have tried to assure the loyalty of their viewers by using a system called “*Watch&Win*”. For example you “watch a live TV show and keep your eyes peeled for the secret word. When you've got the word right, you enter the code in a web form and then you are part of the competition entries. The code may appear

superimposed on the screen several times to give more chances to spot it”.

Based on this TV *Watch&Win* method, we have designed an alternative algorithm to assure loyalty in the vision of academic video recordings named PINVOX. The main difference with respect to *Watch&Win* word encoding, is that PINVOX injects a random audio PIN code in recorded videos that can be fast forwarded and rewinded. PINVOX avoids the use of numeric values superimposed on the video frames which may be easily spotted by students.

In order to give fair certification beyond the physical classroom soon after the complete vision of a recording, within PINVOX the user selects a video lecture catalogued on a web page (*e.g.*, from ICTP.tv). This link then opens an on-line form where two fields needs to be fulfilled, namely: the student e-mail address and a CAPTCHA (to ensure that these responses are generated by a person). After entering this minimum information requested, the algorithm automatically triggers an e-mail reply to the indicated e-mail account. Prior to sending this e-mail, the algorithm carries out some automated processes in order to create the PINVOX audio code and mix it into the original audio of the video. This can be done automatically in only few steps:

- A unique session ID is given first to the selected video.
- From this video, the audio stream is separated from the original audio stream (using, *e.g.*, the open source "ffmpeg" program). An example of audio stream is shown in Fig. 1 (see upper curve).
- General information about the video and audio is analysed, such as audio/video bitrate, number of audio channels, video duration, video format, *etc.*
- Pre-recorded integer numbers (or letters or words [8] in different languages) anticipated with a "gong" are selected randomly from a database and copied into a Log file associating them to each video ID.
- Depending on the length of the original audio, a second audio trace of identical length is created in which the selected random audio PIN numbers are intercalated with streams of silence. Figure 1 shows and example of this second audio track (with five-digit random audio PIN codes labelled a,b,c,d,e).
- By using, *e.g.*, "sox", the cross-platform (Windows, Linux, MacOS X, *etc.*) command line for sound eXchange or the Audacity graphics tool for audio manipulation, we then mux in one file both, the original audio retrieved out of the video with the second digital audio track containing the unique PINVOX code plus spaces of silence. In doing this, the info about bitrate and duration, audio channels, *etc.* is used to maintain the same audio/video quality than the original.
- Finally, the new mixed audio track is embedded back into the video recording. In the example of Fig. 1, this PINVOX audio track with digits generated randomly is superimposed to the original audio track as shown by the bottom curve.

In long audio spectrum, there are sound signals of many different frequencies and amplitudes which can be plotted and analysed to detect “periods of silence”, background noise, audience noise, *etc.* from foreground speech. These

can easily manipulated digitally to place labels, reduce threshold percentages, adjust levels at specific points, *etc.* Within PINVOX it is necessary to track up the periods of silence to inject the random audio PIN codes (as shown in Fig. 1) in order not to interrupt the normal flux of a speech or seminar talk and avoid cutting in two some words within the original audio recording.

For illustrative purposes, an analysis of areas of sound that are separated by silence is shown in Fig. 2 (asuming the same minimum threshold), where we split original audio recordings into several tracks at specific points of silence of different durations (2 to 10 seconds). The curves in this log-log plot correspond to four different speakers giving pre PhD lectures of physics and mathematics in the ICTP Diploma Courses [1].

It is interesting to note that the number of audio files split out from areas of silence follows closely a linear curve. This fact allows to simplify choosing the places where to place automatically the audio PINs correctly. All the above processes are reasonably fast to carry out on any desktop running Linux Ubuntu. For a 90 minutes video we found, *e.g.*, it can only take less than five minutes to complete.

The main concept behind PINVOX is represented in simple terms by Fig. 3. For instance, a random multi-digit

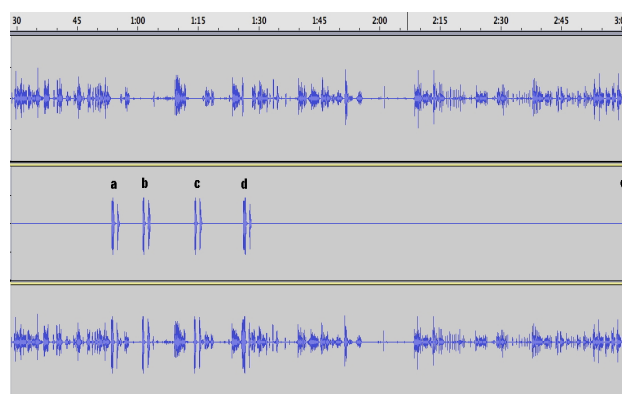


Figure 1. Illustration of audio PIN injection: *Upper curve*: original audio stream of a video recording. *Center*: PINVOX track of five digits generated randomly (labelled a,b,c,d,e). *Bottom curve*: final mixed audio stream.

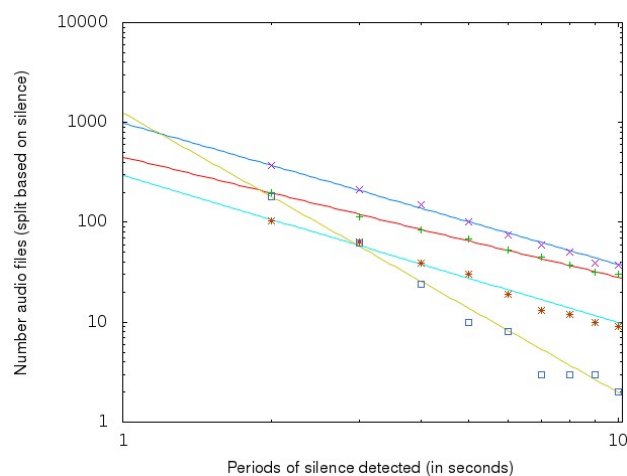


Figure 2. Split of original audio stream from 4 different ICTP Diploma video lectures and lecturers (taken from ICTP.tv) into several tracks at specific points of silence of different durations.



Figure 3. Principal steps for on-line certification via PINVOX to ensure that scholars have followed complete video lectures.

code is generated (3-7-1-9 in this example). This numerical code is transformed in an audio track, which is formed by silences plus the sound of a bell followed by the english reading of a number (“THrē”, “Sevən”, “wən”, or “nīn”). Then by identifying the right code after the full vision of the video recording, a certificate is automatically issued.

Finally, to make PINVOX more robust, the initial e-mail message delivered to the User can contain *i)* an encrypted URL pointing to the selected video having the audio PIN code injected randomly, *ii)* the unique ID for the video to be examined, and *iii)* another encrypted URL pointing to a second web form used to retrieve the award certificate after the correct identification of the associated PINVOX code. Soon after, the certificate is also sent via e-mail to the User's address provided initially.

III. REMARKS

Evaluation mechanisms of students following on-line courses have been extensively discussed within e-learning, distance education and mobile learning programs [5], whereas verified video viewing for on-line certification has received no attention [6,7]. Both types of assessments are necessary to have a reliable mechanism that leads to obtain an academic certification on-line. The issue of ensuring at least the virtual attendance to the video lectures is the one tackled in this work. Our goal is to ensure that a scholar do indeed follow, *i.e.*, listened and watched completely, a video lecture.

So far as we know, there is no another way to guarantee that students have gone through the whole video. PINVOX uses an audio numeric code superimposed to the original video audio track. The code is random both in length as in content, so the students are obliged to listen to the whole video lecture in order to know the code. When he/she has the correct code, at least the attendance to the lecture is ensured. As the code is generated randomly and on the fly, two students will have different codes. Once a code is used (*i.e.*, entered in a web form as in Fig. 3), the audio PIN track is immediately deleted from the database.

In summary, PINVOX is based on the random insertion of pre-recorded integer numbers (or letters or single words [8]) on a video.

1. Audio PIN codes are composed of few numbers (depending of the length of each video).
2. These are unique generated on-the-fly at each video downloaded.
3. Audio PIN codes are only identified when watching the whole video only.
4. AudioPIN codes are superimposed into the audio trace of the lecturer at places where (in principle, but not necessarily) silence is detected by the algorithm after some chosen noise tolerance.
5. The certificate is awarded on-line after the successful identification of the unique acoustic PIN injected into the video.

In this way, the numeric audio values superimposed on the original video (which can be fast forwarded and rewinded), cannot be easily spotted by students.

One can also imagine many other creative applications for PINVOX to verify video viewing. For example, it could also be implemented for a more accurate pre-selection of interested participants to scientific events or for on-line marketing with prize awards.

ACKNOWLEDGMENT

The authors acknowledge Prof. F. Quevedo, ICTP Director, for supporting the PINVOX project.

REFERENCES

- [1] ICTP phys & math diploma courses on-line: <http://www.ictp.tv> and <http://itunes.ictp.tv>
- [2] MIT OpenCourseWare: <http://ocw.mit.edu>
- [3] edX partnership between the Massachusetts Institute of Technology (MIT) and Harvard University: <http://www.edxonline.org>
- [4] Forbes magazine, article by J.M. Crotty, “New Equation For Free Education: MITx + HARVARDx = edX”, <http://www.forbes.com> published: 2012/05/06
- [5] N. Martin, O. Martínez Lazalde, C. Stokes and D. Romano, “An evaluation of remote communication versus face-to-face in clinical dental education”, *British Dental Journal* **212** (2012) 277. <http://dx.doi.org/10.1038/sj.bdj.2012.226>
- [6] E. Canessa, C. Fonda and M. Zennaro, “One year of ICTP diploma courses on-line using the automated EyA recording system”, *Computers & Education* **53** (2009) 183. <http://dx.doi.org/10.1016/j.compedu.2009.01.011>
- [7] P. Gorissen, J. Van Bruggen and W. Jochems, “Usage reporting on recorded lectures using educational data mining”, *Int. J. of Learning Tech.* **1** (2012) 23. <http://dx.doi.org/10.1504/IJLT.2012.046864>
- [8] Pre-recorded audio library: <http://evolution.voxeo.com/library/audio/prompts/home.jsp>

AUTHORS

E. Canessa, C. Fonda, M. Zennaro are with the Science Dissemination Unit of the International Centre for Theoretical Physics (ICTP), Trieste, Italy (e-mail: sdu@ictp.it).

Received 20 May 2012. Published as re-submitted by the authors 7 August 2012.