On Gamifying the Transcription of Digital Video Lectures

Marco Furini

Dipartimento di Comunicazione ed Economia
Università di Modena e Reggio Emilia
Viale Allegri 9, Reggio Emilia, Italy
EMail address: marco.furini@unimore.it

Abstract

Games can be used to exploit the computational power of humans to perform tasks that are difficult for computers. One of these difficult tasks is the transcription of video lectures. Indeed, the characteristics of the speech that occur in video lectures are not well suited for speech recognition technologies. In this paper we propose ALGA, an ALtruistic GAme, designed to involve students in the production of transcripts. Players challenge each other by listening to short, and randomly selected, pieces of the audio stream, and by submitting the corresponding transcription. When two players (unknown to each other) submit the same version, the transcript of the audio chunk is considered correct and the players gain points. To motivate players, ALGA provides the final transcript to all the players and maintains a high-score list for every video lecture. The evaluation shows that the accuracy of the obtained transcripts is higher than the one obtained by speech recognition technologies and also shows that participants like the game approach. Hence, ALGA can be considered a reasonable, feasible and affordable solution to produce transcripts from video lectures.

Keywords: games with a purpose, automatic transcription, video lectures, accessibility

1. Introduction

Many private and public educational institutes use video lectures to improve the effectiveness of teaching in and out of classrooms and to support

distance-learning students [1, 2, 3, 4], but the accessibility to this material is not as easy as one may think. Indeed, millions of people have difficulties in listening to a lesson (e.g., hearing impaired students), in taking notes (e.g. motion impaired students) or in understanding how a teacher speaks (e.g., English as a second language students) [5].

To increase the accessibility of video contents, many educational institutes provide students with the transcripts of the video lectures. These transcripts are produced either by human beings or by software. The former approach produces accurate transcripts but the process is time consuming and expensive; the latter approach employs ASR (Automatic Speech Recognition) software and takes advantage of the advances in speech recognition technologies that allow to achieve accuracy up to 99% when correctly trained, when used for dictating purposes and when using good quality microphones in a good acoustic environment. Unfortunately, in the classroom scenario the accuracy may drop a lot: recent studies show that the accuracy of speech recognition in a classroom scenario is usually less than 70% [6]. In fact, the classroom scenario is very different from the dictating one, as the spontaneous speech that occurs in a lecture is acoustically, linguistically and structurally different than the one used to create written documents: the speaker talks at different speeds and different volume to emphasize some part of the speech, he/she often uses fillers (e.g. uh, er, um, ah), sometimes he/she hesitates in the middle of a word and does not speak punctuation marks ('comma', 'dot', 'question mark', etc.) [7]. Therefore, in most cases, the transcript produced by a speech recognition application needs to be copy edited. Since professional copy editing might be too expensive, a recent approach involves the use of online task markets, such as Amazon Mechanical Turk, to obtain low-cost and high-accuracy transcripts [8]. However, although less expensive, this solution might be not affordable for many educational institutes [9].

Motivated by the continuous success of social games and applications [10, 11], in this paper we study a possible *gamification* of the transcription process and we propose ALGA, an ALtruistic GAme, designed to involve students in the process of transcript production.

The use of games in the educational scenario is not new: traditional board games and role-play games were used for learning long before the arrival of digital games and several studies agree that games, either digital or not, might play an important role in formal education [12]. Indeed, modern theories of effective learning suggest that computer educational games could provide a rich-resource learning environment and the potential benefits of

using video games as ideal companions to classroom instruction is unquestionable [13, 14]. Furthermore, the use of elements of game to motivate learning is seen as "a serious approach to accelerate the curve of the learning experience, teach complex subjects, and systems of thought" [15]. For these reasons, computer games are more and more used in various learning scenarios, e.g. classroom education, government, financial services, health-care, science, telecommunications, corporate, military training, etc. [16].

In this paper, the game we propose to involve students in the transcription process falls in the area of Game With A Purpose (GWAP) [17]. GWAP games are used to exploit the computational power of humans to perform tasks that are easy for humans, but difficult for computers. An example of GWAP game is ESP [18], a game designed to label pictures: as known, it is very difficult for computers to understand the semantics of an image, but this task is quite easy for humans. Therefore, ESP gamified this process by displaying the same picture to two players (unknown to each other) and by asking them to label the picture. If players submitted the same label for the picture, the label was considered appropriate.

Our proposal aims at gamifying the transcription process of video lectures with an approach that does not require the use of any speech recognition technologies, of professional copy editing and of monetary incentives to motivate players. All ALGA needs is the audio stream of the video lecture and the students playing.

Briefly speaking, ALGA automatically splits the audio stream of the video lecture into several short audio chunks (length in the range 5-20 seconds) and randomly presents one of these audio chunks to a player, asking him/her to submit a transcription of it. If the entered text matches a previous submitted text, the player (and the one who submitted the text that matches) gains points. After playing with a chunk, the player can play with another one and he/she can play as many chunks as he/she wants. A high-score list is maintained for each video lecture and when a player is passed in the ranking, he/she is informed (e.g., through e-mail) in order to encourage him/her to play again. The transcript is made available to all the players when the percentage of audio chunks correctly transcribed is above a threshold (defined by the game administrator, usually the video lecture speaker).

In addition to transcription, ALGA allows to play with links to external resources: players can suggest resources that contain materials to deepen the study of the topic(s) covered in the video lecture. Also in this case, if the entered link matches a previous submitted link, the player (and the one who

submitted the link that matches) gains points.

The evaluation of the proposed game shows that the game approach is appreciated by students and also shows that the accuracy of the produced transcript is higher than the one obtained through ASR technologies. Therefore, ALGA can be considered a reasonable, feasible and affordable solution to produce transcripts from video lectures.

The remainder of the paper is organized as follows: Section 2 presents related work in the area of transcript correction; Section 3 shows details of the ALGA proposal, whereas Section 4 shows the ALGA evaluation study. Conclusions are drawn in Section 5.

2. Related Work

The gamification of the transcription process of digital video lectures involves different fields: the use of games in education, the gamification approach, the transcription of digital lectures and the game with a purpose scenario. In the following, we overview studies in these different fields.

2.1. Serious Games in Education

In the past years, different studies showed that well-designed computer educational games provide undeniable benefits and might be well suited for active learning as they provide a learning environment able to foster the higher order knowledge and the skills of students [13, 14, 15, 16]. Indeed, most computer games are active, experiential and interactive and these features are those that most influence effective learning [13]. Therefore, it is not surprising that games are being employed more and more into learning environments (e.g. classroom education, financial services, healthcare, military training, etc.) [16, 19]. For instance, Corti [20] analyzed the use of serious games in the learning environment and found that the benefits of using video games is unquestionable. Hwang et al. [21] showed that games are perceived as a means to engage players in enjoyable activities to accomplish some challenging objectives. Wrzesien and Alcaiz Raya [22] indicated three main reasons for the ever-increasing use of serious games in education: (a) they use actions rather than explanations and create personal motivation and satisfaction, (b) they accommodate multiple learning styles and abilities, and (c) they foster decision-making and problem-solving activities in a virtual setting. Gentile and Gentile [23] observed that the use of computer games in

the learning environment brings happiness and sense of achievement to learners, thus helping them to improve their learning results and stimulate them to think; these advantages suggest that computer games could be applied to improve traditional methods of teaching. Ebner and Holzinger [24] found that students who use games in the learning environment produce learning results that surpass other methods of traditional teaching. Finally, we observe that researchers have developed computer games to improve learning for diverse disciplines, such as mathematics, computer science and linguistic (e.g., [25, 26, 27], just to name a few).

However, it is worth mentioning that the use of serious games in the educational scenario is also subject to a critical thinking. Guillen et al. [16] highlithed that there is still little consensus on the process by which games engage learners and on the types of learning outcomes that can be achieved through game play. Connelly et al. [13] observed that it is difficult to understand the effects of games. Lin et al. [19] observed the difficulties in developing effective games for the learning environment.

Far from settle the debate, we observe that a recent trend in education is the usage of the so-called "gamification" approach.

2.2. The Gamification approach

Since the initial proposal of using computer games in the educational scenario, the use of games for learning purposes has evolved and recently the "gamification" process has gained significant attention [28]. This process refers to "the use of game design elements in non-game contexts" and is applied not only to education, but also to many other scenarios like healthcare or production.

With respect to the educational scenario, the gamification approach is increasingly used for teaching students, training people, engaging users and balancing difficulties and abilities [29, 30, 31]. In particular, Barata et al [32] explored how gamification can be applied to education in order to improve student engagement: they gamified a college course and compared it against a non-gamified one. Results showed that students who attended the gamified course had significant improvements in terms of attention to reference materials, online participation and proactivity. Moreno [33] investigated the gamification of programming and showed that students who used a video game to improve their programming skills, performed 12% better in their final exam than students who did not. Iosup and Eperma [15] experimented a

teaching technique that used social gaming elements to deliver higher education: they found that that gamification is correlated with an increase in the percentage of passing students, and in the participation in voluntary activities and challenging assignments. Furthermore, they found that gamification also fostered interaction in the classroom and triggered students to pay more attention to the design of the course. Sepehr and Head [34] examined the effect of gamification techniques in engaging students in a teaching context; in particular they analyzed the influence of competition and results showed that competition is a key element that highly motivates students to engage in the gamification tasks. Gene' et al. [28] analyzed the gamification in Massive Online Open Courses and proposed a model to motivate MOOC's students based on game elements: they proposed to rank students, to show course progress, to provide a course certification and to introduce the "like" features in the MOOC courses. O'Donovan et al. [35] discussed positive and negative aspects of gamifying a university course and they showed evidences that gamification improves student engagement and understanding.

2.3. Transcription of Digital Lectures

The Web is full of video lectures, tutorial, reviews, reports, interviews that are not accessible to a large part of the population (e.g., hearing impaired people, dyslexic students or non-native speaker) to machine (e.g., search engines) and to anyone who needs to search, review or translate what is said in the video [36]. With no doubt, the availability of transcripts would provide several benefits (e.g., improve comprehension, reduce deficiencies, improve indexing, etc.) and currently there are two main approaches to produce transcripts: manual or automatic. The former approach involves professional people who usually guarantee accuracies as high as 99%, for fees as low as $1 per minute of transcribed text [8], whereas the latter approach requires the usage of ASR software that may achieve very high accuracy (up to 99%) in a ideal scenario (e.g., good acoustic environment and microphone quality), but when used in the classroom scenario the accuracy may drop a lot (less than 70%) and therefore manual correction is required [6].

Knight and Almerot [37] proposed AutoCap, a software designed to produce timecoded transcripts through ASR technologies, but the approach requires experts to correct the transcript produced by the ASR. Wald [7, 36] described a tool that facilitates crowdsourcing correction of speech recognition captioning errors. Starting from a time-coded transcript, the tool allows

users (one at a time) to correct possible errors in the transcript and the system stores the modified versions of the transcript. Luz et al. [38] proposed a 3D "transcription game" that displays sentence transcription candidates through animated 3D representations of word lattices generated by speech recognition. The user can interact with these sentence representations by selecting the correct paths as the words move towards the background. Novotney and Callison-Burch [39] suggested to partition audio files into 5-second segments and to post tasks on the Amazon Turk platform, where Turk workers transcribe the audio segment for a small fee. The best transcripts were 13% lower in word accuracy. Liem et al. [8] used a similar approach to improve existing transcripts, but instead of using the Amazon Turk Platform, they involved students. They proposed to partition the audio track into 10-second segments and to process these segments independently. The mechanism uses ASR software to generate the transcript and asks students to correct it.

Recently, different proposals introduced the game approach in the transcription process, making these proposals to fall into the Game With A Purpose category.

2.4. Games With A Purpose

Games With A Purpose (GWAP) are games that exploit the computational power of humans to perform tasks that are easy (and somehow entertaining) for humans, but difficult for computers. There are many examples of these tasks, from labeling pictures to decoding the code for genetic diseases, from understanding the perceived colors of a picture to unfolding RNA molecules.

The first known example of GWAP game is ESP, a game designed to label images [18]. This task is very difficult for computers, but quite easy for humans. Therefore, the idea of the game is to make entertaining the labeling process, by displaying the same picture to two players, randomly picked and unknown to each other, and by asking them to reach an agreement on a label for the picture. Another example of difficult task for computers is the ranking of images and Lux et al. [40] proposed to introduce game elements in this task.

With respect to the transcription process, Kacorri et al. [9] proposed a caption editing system based on crowdsourced work. The audio is first transcribed with the IBM Attilia speech recognition engine and then is split into 2-10 seconds segments. Users have to identify the correct captions for each

video segment. To make the task more entertaining, users watch the video segment and has to enter the correct caption playing against a countdown timer. To obtain the final caption, the proposal aligns and merges all the captions submitted by users. Preliminary results with 42 participants and 578 short video segments showed that the accuracy increased of 4% with respect to the one achieved by the sole ASR software.

3. The ALGA Proposal

To produce transcripts from video lectures without the use of speech recognition technologies, of professional copy editing and of monetary incentives to motivate non-professional workers, we consider the GWAP approach and we propose the gamification of the transcription process of video lectures. In particular, we propose ALGA, an ALtruistic GAme designed to involve students in the transcription of video lectures.

In ALGA, students can challenge each other with two different play modes: textual and link. In the textual mode the goal is to produce the video lecture transcript, whereas in the link mode the goal is to associate to the video lecture external resources that may help students to deepen the study of the topic(s) covered in the lesson. In particular, in the textual mode, the player has to listen to a piece of the audio lesson and has to submit the corresponding transcription. The length of the audio chunk is determined by an audio segmentation process that splits the audio according to the audio characteristics (note that, from an experimental investigation, see Section 4, the length of the audio chunk varies between 5 and 20 seconds). The splitting is done to ease the transcription process, as players only need to listen and transcribe a short piece of audio. Indeed, since a person speaks 150-170 words per minute, the transcription of an audio chunk of 5-20 seconds requires the player to write around 13 -50 words. When the entered text matches a previous submitted text, the player (and the one who submitted the text that matches) gains points. After playing with an audio chunk, the player can play with another one and can play as many chunks as he/she wants. In the link mode, the player has to submit a URL to an external resource that he/she thinks might be useful to explore the topic covered in the lesson. When the entered URL matches a previous submitted URL, the player (and the one who submitted the URL that matches) gains points.

Since a game needs to be interesting and engaging [41], ALGA motivates students with two different rewarding schemes: i) the entire transcript of the

video lecture and ii) a high-score list for every video lecture. If the transcript of a lecture can be certainly useful to students, the use of high-score lists is due to the fact that these are felt like a reward [21, 42, 43]. For this reason, when a player reaches the first position, he/she is nominated the current "Master of the video lecture' and, at the same time, the previous "Master" is informed through e-mail that he/she is no longer at the top of the high-score list. This works as a stimulus to play again.

ALGA also provides the figure of a director for each course (e.g., a group of video lectures), a special user (e.g., the one who speaks in the video lecture, or a manager appointed by the content provider) that may transcribe or may provide links without the need to be confirmed. Also, the director has to specify the threshold (e.g., 90%) beyond which it is possible to release the transcript to the players.

In the following, we present the details of the game rules, the way the game engine operates and an example of the game. An ALGA prototype will be presented in the next section.

3.1. The Rules of The Game

Object of the game

Become "Master of the video lecture" by collecting more points than other players. Points are gained when a correct transcription or a useful link is entered. To be considered correct, the transcript must be submitted by two different players (or submitted by a player and confirmed by the director). The same applies to the link to be considered useful.

Game Reward

The video lecture transcript.

Game Setup

After login (only registered users may play), a player has to select the course topic (e.g., Linguistics, Communication Technology, Computer Networks), the specific video lecture he/she wants to play with and the play modality. The course topic and the video lecture can be selected among the ones available, whereas the play modality may be either Textual or Link.

1. **Textual**: a player has to listen to an audio chunk (randomly selected by the system) and has to write the textual transcription. If the submitted transcription matches a transcription submitted by another player, both players gains one point. If the transcription does not match any

other transcriptions, but it is confirmed by the director, the player gains one point.

2. **Link**: a player has to suggest a link to a Web resource for the entire video lecture he/she is playing with. The player gains ten points if the submitted link matches a link submitted by another player or if the director confirms the link is useful for the topic(s) covered in the video lecture.

3.2. The Game Engine

ALGA considers a video lecture as composed of three components: the audio stream (A), the audio transcript (T) and the links (L). In particular, the audio file is logically seen as composed of a sequence of N audio chunks: $A = \{a_1, a_2, ..., a_i, ..., a_N\}$; the transcript T is considered as composed of a sequence of N pieces: $T = \{t_1, t_2, ..., t_i, ..., t_N\}$, and the link L contains all the external resources associated with the video lecture $L = \{l_1, l_2, ..., l_i, ..., l_K\}$. Initially, T and L are empty and will be filled by the players during the play.

Each t_i contains the following information:

- s_time_i: the initial audio point of the audio chunk a_i with respect to the entire audio stream;
- e_time_i: the final audio point of the audio chunk a_i with respect to the entire audio stream;
- $text_i$: the textual transcription of what is said in chunk a_i;
- $state_i$: the current state of the chunk t_i;
- $player_j$: is a $t-uple$ containing the ID of the players who entered $text_i$ (i.e., $Player_j = (p_1, p_2)$)

Each link l_i contains the following information:

- url_i: the url suggested for the video lecture;
- $state_i$: the current state of the link l_i;
- $player_j$: is a $t-uple$ containing the ID of the players who entered url_i (i.e., $Player_j = (p_1, p_2)$)

The state of t_i can be one of the following and the transitions between states are depicted in Figure 1:

Stand-by: the transcript of the audio chunk finds no confirmation and therefore, for the game point of view, the actual transcript of chunk a_i is still to be determined;

Checked: two players submitted the same transcript of the audio chunk, or the transcript submitted by a player was confirmed by the director, or the director directly wrote the audio chunk transcription. In all these cases, for the game point of view, *text* contains what is actually said in the audio chunk a_i;

The state of l_i can be one of the following:

L-Stand-by: the link finds no confirmation and therefore, for the game point of view, the link contains material not yet verified either by other players or by the director.

Enriched: two players submitted the same link, or the link submitted by a player was confirmed by the director, or the director directly specified a link for the video lecture. In all these cases, the link is included into the $Link_i$ $t-uple$.

Depending on the game modality, the game engine operates as follows:

1. **Textual Mode**: the game server randomly selects t_i among the ones that are not in the *checked* state and presents the corresponding audio chunk a_i to the player in its aural form. The player has to listen to the audio and has to submit the transcript. The submitted version is compared against all the versions in the *stand-by* status. If a match is found, the players that provided the same version gain one point, the status of the entry is changed from *stand-by* to *checked* and all the other versions of t_i with *stand-by* status (if any) are removed. If there is no match, the state of t_i is labeled as *stand-by* and the player does not gain points. Note that, the director may change from *stand-by* to *checked* the status of any entry.

2. **Link**: the player may suggest a link to an external web resource that he/she thinks is worth reading to deepen the topic(s) covered in the entire video lecture. Every submitted link is compared against all the links in the *L-stand-by* status. If a match is found, the players that

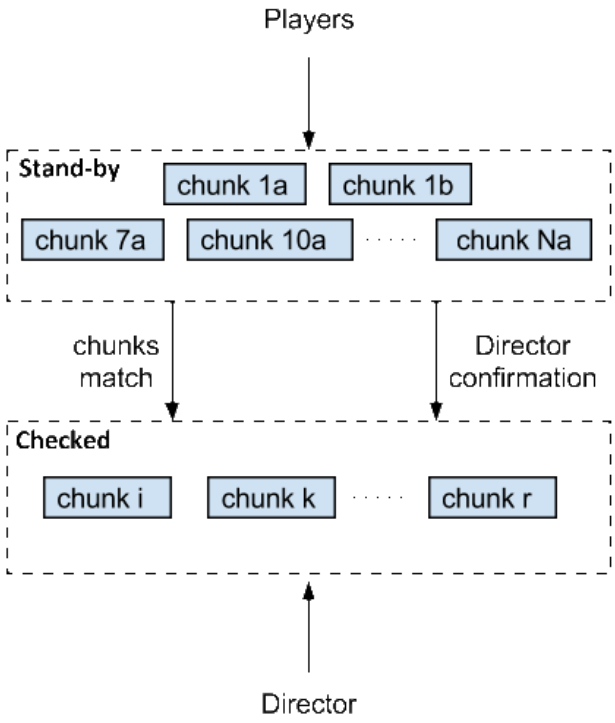

Figure 1: States and transitions of an audio transcript t_i.

provided the same version gain ten points, the state of the entry is changed from *L-stand-by* to *enriched*. If there is no match, the state of the suggested link is labeled as *L-stand-by* and the player does not gain points. In addition, the director may change the status from from *L-stand-by* to *enriched* of every suggested link that he/she thinks is worth reading to deepen the topic(s) covered in the entire video lecture. Also in this case, the player who suggested the link gains ten points.

3.3. A Game Example

Let us suppose that the selected original audio chunk a_i contains: "The Web is an Internet application that links millions of web pages" and suppose that the game is played by players P_a, P_b and P_c.

Scenario 1: P_a plays in the Textual mode and writes "The web is an Internet application that links millions of web pages". P_b plays in the Textual mode and writes "The web is an application that links billions of web pages".

If there is no match with the entries labeled as *stand-by*, the following entries are stored in the DB:

- (..., "The web is an Internet application that links millions of web pages", *stand-by*, ..., (P_a))
- (..., "The web is an application that links billions of web pages", *stand-by*, ..., (P_b))

In this scenario, no points are given to the player.

Scenario 2: P_c plays in the Textual mode and the DB stores the entries of the Scenario 1. P_c writes "The web is an Internet application that links millions of web pages". Since the P_c version matches the one of P_a, the version is considered correct. As a consequence, the state of the entry is changed from *stand-by* to *checked*, whereas all the other versions labeled as *stand-by* (if any) are removed:

- (..., "The web is an Internet application that links millions of web pages", *checked*, ,..., (P_a, P_c))

In this scenario, one point is given to players P_a and P_c.

4. Experimental Assessment

To evaluate ALGA, we developed a prototype version of the game and we set-up an experimental scenario composed of twelve different 45 minutes video lectures taken from three different courses (Linguistic, Videocommunication Lab and Communication Technology). We involved 59 students and we asked them to participate to the experimental scenario.

4.1. The ALGA Prototype

The prototype is based on a web architecture, see Figure 2, so that the game can be enjoyed by players using any type of device. The back-end of the game is composed of a game engine that accesses to the audio stream of the video lecture, splits it into several chunks and implements the rules of the game. A MySQL DB is used to keep track of the players actions.

One of the most important component of back-end is the one that splits the audio of the video lecture into several chunks. To this aim, we developed an audio splitting mechanism based on the audio energy investigation as this

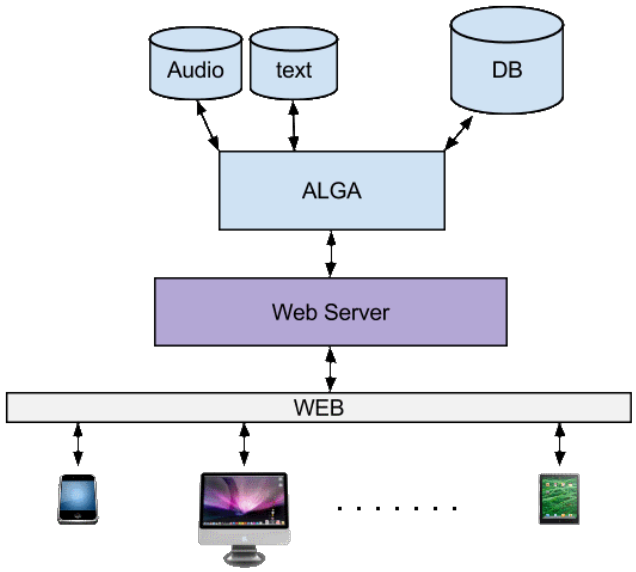

Figure 2: ALGA architecture: it is designed as a Web-based game in order to improve accessibility.

technique is easy to implement and has been widely used in literature (e.g. [44, 45, 46, 47]. In particular, for the sake of simplicity, we considered the audio stream as a sequence of N *virtual frames*, $f_1 f_2 ... f_N$, where the time length of each virtual frame is fixed and equal to 30 milliseconds (a common time length) and then we computed the sound energy of each virtual frame with the formula:

$$Energy_i = \frac{\sqrt{\sum_{n=1}^{N} pcm_n^2}}{N} \tag{1}$$

where, N is the number of audio samples within the frame and pcm_i is the $i-th$ audio sample of the considered frame[1].

The goal of our mechanism is to identify locations where to split the audio stream. These locations must correspond to silence periods. The audio energy investigation computes all the frames that can be considered

[1]Note that since the number of audio samples per second may be considerable, we considered sub-sampling and we computed one audio sample out of ten.

Silence Lenght	Avg # of chunks per video lecture	Avg chunk length	Avg chunk words	Avg # of truncated words
30 ms	371	7.30 s	18	59
60 ms	267	9.83 s	24	37
90 ms	240	10.67 s	27	18

Table 1: Characteristics of the audio chunks produced by the audio splitting mechanism. The minimum audio chunk length is set to 5 seconds and the silence length varies. The number of truncated words refers to a video lecture of 45 minutes.

silence. An important parameter that may affect the identification of the locations where to split the audio stream is the length of the silence. Indeed, if we consider very short silence, the mechanism will likely identify as "silence periods" the short silence located between syllables of the same word and this will likely cause our mechanism to split the audio in the middle of a word. For this reason, in our experimental scenario we consider different values of the silence length: from 30 milliseconds (shorter silences are not worth investigating as they will likely correspond to silence between syllables of the same word) [48] to 90 milliseconds.

Another important parameter required by our mechanism is the length of the audio chunk that players will need to transcribe. To this aim, we performed an experimental assessment asking participants to listen and transcribe chunks of different lengths (5, 10, 15 and 20 seconds). After that, we asked them the preferred length: results showed that the majority of them (87%) agree on the 5 seconds length.

We applied the developed audio splitting mechanism to the video lectures by varying the silence length and the characteristics of the obtained chunks are reported in Table 1. These characteristics show that the shorter the silence length is, the shorter the average audio chunk length is. The reason is that short silences are more frequent than longer silence and therefore it is possible to split the audio more frequently (in this case close to the 5 seconds length suggested by participants). Note that the average number of words per chunk is statistically computed by considering that a speech has around 150-170 words per minute.

With respect to the front-end, ALGA requires user to login and to select the course he/she wants to play with and the play mode (i.e. Textual or

Link). This is done by selecting the course degree (left column of Figure 3) and the course name (central part of Figure 3). After that, a pop-up window asks user the play modality. Once the course and the play mode are selected, the player has to select the video lecture he/she wants to play with (left part of Figure 4). Once selected, the game is ready to play.

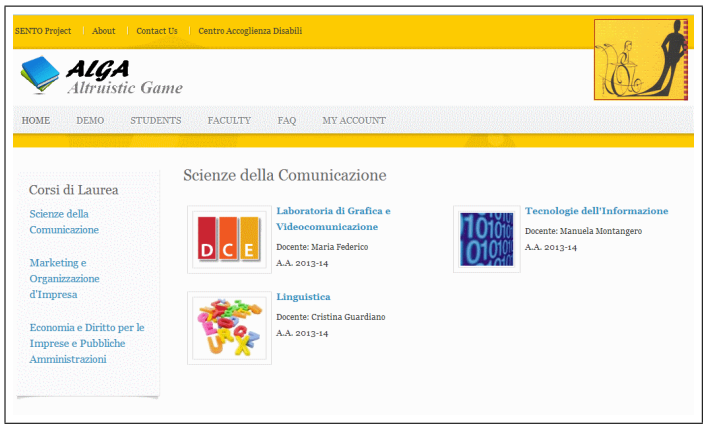

Figure 3: Players can select the course they want to play with by selecting the course degree (left) and the course name (central).

Figure 4 shows an example of the textual mode: the player has to listen to the audio chunk and has to write the transcription in the box. After that, the player is notified if he/she gained the point, he/she is informed about the current position in the high-score list and is asked if he/she wants to play again.

Figure 6 shows an example of the personal profile page: here the player can keep track of his/her position in the various video lectures. For instance, in the ranking page he/she can observe the current ranking in all the video lectures he/she played with.

4.2. Level of Participation

After one week we observed that the level of participation was considerable. In particular, each chunk obtained with a silence length of 30 milliseconds was played 3.09 times; each chunk obtained with a silence of 60 milliseconds was played 3.2 times; each chunk obtained with a silence length of 90 milliseconds was played 3.4 times. On average, every day each player played 28 chunks.

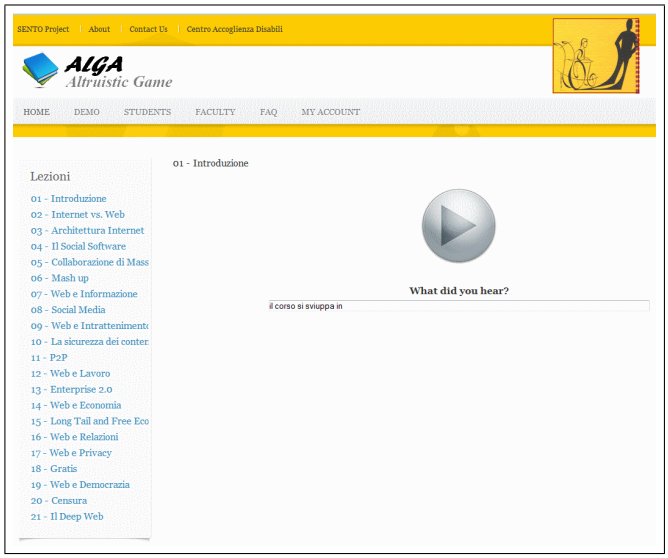

Figure 4: Players can listen to the audio chunk and can enter the transcript. Then, the player is notified if he/she gained points, of the current position in the high-score list and is asked if he/she wants to play again.

With respect to the link mode, on average, 17 different links to external resources per video lecture were submitted and 1.5 (18 links in the 12 video lectures) was considered useful to deepen the topics covered in the video lecture.

4.3. Participants' Experience

To investigate the participants' experience, we considered a Mean Opinion Score (MOS) evaluation and we asked them to fill a questionnaire using a 5-point Likert scale. In particular, we asked them four different questions: Q1: Evaluate the game in general. Q2: Was the game fun to play? Q3: Was the final transcript useful? Q4: Did the final transcript meet you expectations?

Results, presented in Figure 6, show that the game approach was considered interesting. The only score below 3 is the one related to the expectation of the final transcript. We asked participants to motivate their score and most of them expected a book-chapter with figures and formatted text, rather than plain transcript. Hence, the low score was mainly due to a misunderstanding of the term "transcript".

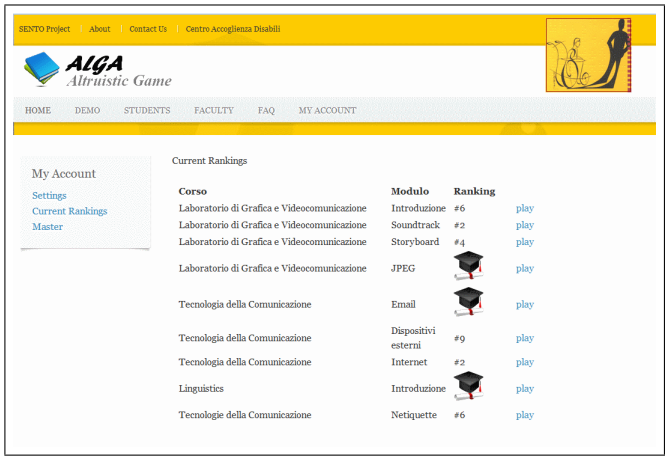

Figure 5: A player can keep track of his/her position in any module he/she played. In addition to the current ranking position, a short cut to play the module is available. A special icon appears when the player is in first position.

4.4. Transcripts Accuracy

The accuracy of the transcripts produced with ALGA was compared against a transcription automatically obtained through the Dragon Naturally Speaking v.11, one of the most accurate ASR (Automatic Speec Recognition) application available among the off-the-shelf ones.

On average, the accuracy achieved while playing the chunks obtained with a 30 ms silence-length was 89% (against 83% achieved by the ASR); the accuracy achieved while playing the chunks obtained with a 60 ms silence-length was 78% (against 73% achieved by the ASR); the accuracy achieved while playing the chunks obtained with a 90 ms silence-length was 77% (against 72% achieved by the ASR). It is to note that these percentages include the number of words truncated by the automatic splitting mechanism.

4.5. Summary of Results

From the evaluation process, it emerged that: i) participants liked the game approach; ii) the Textual mode has been largely played by participants, and iii) the accuracy of the produced transcripts is higher than the one obtained by using the ASR application.

It also emerged that the audio splitting mechanism should consider silence of 30 milliseconds length. In fact, if on the one side the 30 ms silence increased the number of truncated words (see, Table 1), on the other side, the accuracy

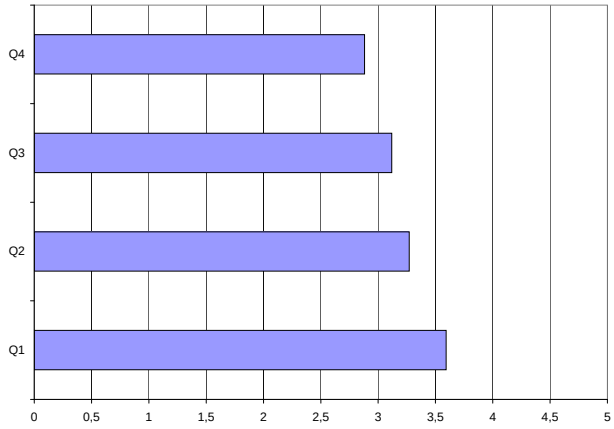

Figure 6: Results of the MOS investigation. Q1: Evaluate the game in general. Q2: Was the game fun to play? Q3: Was the final transcript useful? Q4: Did the final transcript meet you expectations? Standard deviation was in the range [0.7-1.1].

obtained in the transcription process is much higher than the one obtained with longer silences.

5. Conclusions and Future Work

In this paper we proposed ALGA, an altruistic game designed to ease the process of video lecture transcription. In particular, our approach aims at involving students in this difficult task. Throughout the paper we detailed the rules and the details of the game, as well as the characteristics of the prototype we developed in order to investigate the effectiveness of our proposal. Results obtained from the experimental assessment showed that participants liked the game approach and also showed that the accuracy achieved is higher than the one obtained by speech recognition technologies. According to these results the game approach can be considered a promising and easy solution to increase the accessibility of digital video lecture contents and suggested that the gamification approach can be a right direction towards the transcription of video lectures.

Although the obtained results were interesting, it is worth noting that ALGA is just a prototype and hence different entrenchments may be introduced. For instance, it is web-based but the design of a specific app may improve the students experience in the mobile scenario. Note that this does

not require any modification to the game architecture or to the game engine, but it simply requires designing a more usable interface for mobile devices. Another enhancement regards the students participation and engagement. Currently, the prototype employs two rewarding schemes: i) high-score list, and ii) availability of the transcription. Although these schemes contributed to achieve acceptable participation level, they were designed to give us some metrics on the applicability of the gamification approach. Therefore, these schemes may be unfit for long-term engagement. In future versions it would be interesting to investigate the benefits of other rewarding schemes designed to increase both the participation level and the long-term engagement in the game. For instance, we intend to incorporate and evaluate schemes that may be more rewarding (we recall here that ALGA does not consider monetary rewards):

- **Social Media**. In the social age, updates and recommendations play a fundamental role. Social friends are invited to play specific games and are informed about game achievement. Therefore, ALGA should be linked to social media platforms in order to update (either automatically or manually) player's friends about the ALGA score and in order to let them know that he/she is playing the ALGA game. The resulting post/tweet may trigger friends to try the game (among those who don't know the game) or to play again the game (among those who already played the game).

- **Credits by Play**. A student may earn credits for his/her participation to the game. Indeed, the game director should set a threshold (or a set of thresholds) that corresponds to the number of earned credits (e.g., above 70% of chunks played for course X you get 1 credits). Similarly, the threshold (or a set of thresholds) may correspond to the number of questions a student may skip at the written exam (e.g., above 70% of chunks played for course X you may skip 3 questions)

- **Homework by Play**. A student may skip a number of homework if he/she plays a predefined number of chunks. Also in this case, the threshold should be up to the game director (e.g., above 70% of chunks played for course X you may skip 1 homework assignment).

To understand the most suitable scheme we think it is necessary to dynamically understand what can motivate students. Indeed, a rewarding

scheme may be valid this year, but may be unfit the next year (students and their habits change). Therefore, ALGA should ask its players to fill a questionnaire about the preferred rewarding schemes. This questionnaire should be submitted just once to any player and should be submitted either after playing some matches or after a silent period (i.e., a period of time where the student does not play with the game; in this case an invitation through email is necessary to invite him/her to compile the questionnaire). The questionnaire should ask students to rate the reasons to participate to the game (e.g., transcript availability, homework by play, credits by play, etc.) and/or the reasons for having stopped playing with ALGA. The use of one or more of these schemes should increase the participation level and the long-term engagement in the game.

References

[1] T. Liu, J. Kender, Lecture videos for e-learning: current research and challenges, in: Multimedia Software Engineering, 2004. Proceedings. IEEE Sixth International Symposium on, 2004, pp. 574–578. doi:10.1109/MMSE.2004.48.

[2] M. Furini, Secure, portable, and customizable video lectures for e-learning on the move, Informatica 33 (1) (2009) 77–84.

[3] P. E. Dickson, D. I. Warshow, A. C. Goebel, C. C. Roache, W. R. Adrion, Student reactions to classroom lecture capture, in: Proceedings of the 17th ACM Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE '12, ACM, New York, NY, USA, 2012, pp. 144–149. doi:10.1145/2325296.2325334.
URL http://doi.acm.org/10.1145/2325296.2325334

[4] I. N. Toppin, Video lecture capture (vlc) system: A comparison of student versus faculty perceptions, Education and Information Technologies 16 (4) (2011) 383–393. doi:10.1007/s10639-010-9140-x.
URL http://dx.doi.org/10.1007/s10639-010-9140-x

[5] M. Federico, M. Furini, Enhancing learning accessibility through fully automatic captioning, in: Proceedings of the International Cross-Disciplinary Conference on Web Accessibility, W4A '12, ACM, New York, NY, USA, 2012, pp. 40:1–40:4. doi:10.1145/2207016.2207053.
URL http://doi.acm.org/10.1145/2207016.2207053

[6] M. Federico, M. Furini, An automatic caption alignment mechanism for off-the-shelf speech recognition technologies, Multimedia Tools and Applications 72 (1) (2014) 21–40. doi:10.1007/s11042-012-1318-3.
URL http://dx.doi.org/10.1007/s11042-012-1318-3

[7] M. Wald, Crowdsourcing correction of speech recognition captioning errors, in: Proceedings of the International Cross-Disciplinary Conference on Web Accessibility, W4A '11, ACM, New York, NY, USA, 2011, pp. 22:1–22:2. doi:10.1145/1969289.1969318.
URL http://doi.acm.org/10.1145/1969289.1969318

[8] B. Liem, H. Zhang, Y. Chen, An iterative dual pathway structure for speech-to-text transcription.

[9] H. Kacorri, K. Shinkawa, S. Saito, Introducing game elements in crowdsourced video captioning by non-experts, in: Proceedings of the 11th Web for All Conference, W4A '14, ACM, New York, NY, USA, 2014, pp. 29:1–29:4. doi:10.1145/2596695.2596713.
URL http://doi.acm.org/10.1145/2596695.2596713

[10] M. Furini, Mobile games: What to expext in the near future, in: Proceedings of GAMEON Conference on Simulation and AI in Computer Games, EuroSis Society, 2007.

[11] B. Kirman, S. Björk, S. Deterding, J. Paavilainen, V. Rao, Social game studies at CHI 2011, 2011, pp. 17–20. doi:10.1145/1979742.1979590.
URL http://doi.acm.org/10.1145/1979742.1979590

[12] S. Egenfeldt-Nielsen, The Educational Potential of Computer Games, Continuum International Publishing Group Ltd, 2007.

[13] T. M. Connolly, E. A. Boyle, E. MacArthur, T. Hainey, J. M. Boyle, A systematic literature review of empirical evidence on computer games and serious games, Comput. Educ. 59 (2) (2012) 661–686. doi:10.1016/j.compedu.2012.03.004.
URL http://dx.doi.org/10.1016/j.compedu.2012.03.004

[14] R. Van Eck, Six ideas in search of a discipline, The design and use of simulation computer games in education (2007) 31–56.

[15] A. Iosup, D. Epema, An experience report on using gamification in technical higher education, in: Proceedings of the 45th ACM Technical Symposium on Computer Science Education, SIGCSE '14, ACM, New York, NY, USA, 2014, pp. 27–32. doi:10.1145/2538862.2538899.
URL http://doi.acm.org/10.1145/2538862.2538899

[16] V. Guillén-Nieto, M. Aleson-Carbonell, Serious games and learning effectiveness: The case of it's a deal!, Comput. Educ. 58 (1) (2012) 435–448. doi:10.1016/j.compedu.2011.07.015.
URL http://dx.doi.org/10.1016/j.compedu.2011.07.015

[17] L. von Ahn, L. Dabbish, Designing games with a purpose, Commun. ACM 51 (8) (2008) 58–67. doi:10.1145/1378704.1378719.
URL http://doi.acm.org/10.1145/1378704.1378719

[18] L. von Ahn, L. Dabbish, Labeling images with a computer game, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '04, ACM, New York, NY, USA, 2004, pp. 319–326. doi:10.1145/985692.985733.
URL http://doi.acm.org/10.1145/985692.985733

[19] H.-W. Lin, Y.-L. Lin, Digital educational game value hierarchy from a learners' perspective, Comput. Hum. Behav. 30 (2014) 1–12. doi:10.1016/j.chb.2013.07.034.
URL http://dx.doi.org/10.1016/j.chb.2013.07.034

[20] K. Corti, Games-based learning; a serious business application, PIXE-Learning Limited.

[21] G.-J. Hwang, P.-H. Wu, C.-C. Chen, An online game approach for improving students' learning performance in web-based problem-solving activities, Comput. Educ. 59 (4) (2012) 1246–1256. doi:10.1016/j.compedu.2012.05.009.
URL http://dx.doi.org/10.1016/j.compedu.2012.05.009

[22] M. Wrzesien, M. A. Raya, Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the e-junior project, Computers & Education 55 (1) (2010) 178 – 187. doi:http://dx.doi.org/10.1016/j.compedu.2010.01.003.
URL http://www.sciencedirect.com/science/article/pii/S0360131510000060

[23] D. Gentile, J. Gentile, Violent video games as exemplary teachers: A conceptual analysis, Journal of Youth and Adolescence 37 (2) (2008) 127–141. doi:10.1007/s10964-007-9206-2.
URL http://dx.doi.org/10.1007/s10964-007-9206-2

[24] M. Ebner, A. Holzinger, Successful implementation of user-centered game based learning in higher education: An example from civil engineering, Computers & Education 49 (3) (2007) 873 – 890. doi:http://dx.doi.org/10.1016/j.compedu.2005.11.026.
URL http://www.sciencedirect.com/science/article/pii/S0360131505001910

[25] M. Papastergiou, Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation, Comput. Educ. 52 (1) (2009) 1–12. doi:10.1016/j.compedu.2008.06.004.
URL http://dx.doi.org/10.1016/j.compedu.2008.06.004

[26] T.-Y. Liu, Y.-L. Chu, Using ubiquitous games in an english listening and speaking course: Impact on learning outcomes and motivation, Comput. Educ. 55 (2) (2010) 630–643. doi:10.1016/j.compedu.2010.02.023.
URL http://dx.doi.org/10.1016/j.compedu.2010.02.023

[27] F. Ke, Computer-game-based tutoring of mathematics, Comput. Educ. 60 (1) (2013) 448–457. doi:10.1016/j.compedu.2012.08.012.
URL http://dx.doi.org/10.1016/j.compedu.2012.08.012

[28] O. B. Gené, M. M. Núñez, A. F. Blanco, Gamification in mooc: Challenges, opportunities and proposals for advancing mooc model, in: Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality, TEEM '14, ACM, New York, NY, USA, 2014, pp. 215–220. doi:10.1145/2669711.2669902.
URL http://doi.acm.org/10.1145/2669711.2669902

[29] S. Deterding, D. Dixon, R. Khaled, L. Nacke, From game design elements to gamefulness: Defining "gamification", in: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek '11, ACM, New York, NY, USA, 2011, pp. 9–15. doi:10.1145/2181037.2181040.
URL http://doi.acm.org/10.1145/2181037.2181040

[30] S. Deterding, M. Sicart, L. Nacke, K. O'Hara, D. Dixon, Gamification. using game-design elements in non-gaming contexts, in: CHI '11 Extended Abstracts on Human Factors in Computing Systems, CHI EA '11, ACM, New York, NY, USA, 2011, pp. 2425–2428. doi:10.1145/1979742.1979575.
URL http://doi.acm.org/10.1145/1979742.1979575

[31] S. de Sousa Borges, V. H. S. Durelli, H. M. Reis, S. Isotani, A systematic mapping on gamification applied to education, in: Proceedings of the 29th Annual ACM Symposium on Applied Computing, SAC '14, ACM, New York, NY, USA, 2014, pp. 216–222. doi:10.1145/2554850.2554956.
URL http://doi.acm.org/10.1145/2554850.2554956

[32] G. Barata, S. Gama, J. Jorge, D. Gonçalves, Improving participation and learning with gamification, in: Proceedings of the First International Conference on Gameful Design, Research, and Applications, Gamification '13, ACM, New York, NY, USA, 2013, pp. 10–17. doi:10.1145/2583008.2583010.
URL http://doi.acm.org/10.1145/2583008.2583010

[33] J. Moreno, Digital competition game to improve programming skills., Educational Technology & Society 15 (3) (2012) 288–297.
URL http://dblp.uni-trier.de/db/journals/ets/ets15.html#Moreno12

[34] S. Sepehr, M. Head, Competition as an element of gamification for learning: An exploratory longitudinal investigation, in: Proceedings of the First International Conference on Gameful Design, Research, and Applications, Gamification '13, ACM, New York, NY, USA, 2013, pp. 2–9. doi:10.1145/2583008.2583009.
URL http://doi.acm.org/10.1145/2583008.2583009

[35] S. O'Donovan, J. Gain, P. Marais, A case study in the gamification of a university-level games development course, in: Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference, SAICSIT '13, ACM, New York, NY, USA, 2013, pp. 242–251. doi:10.1145/2513456.2513469.
URL http://doi.acm.org/10.1145/2513456.2513469

[36] M. Wald, Concurrent collaborative captioning, 2013.
URL http://eprints.soton.ac.uk/354312/

[37] A. Knight, K. Almeroth, Fast caption alignment for automatic indexing of audio, International Journal of Multimedia Data Engeneering and Management 1 (2) (2010) 1–17.

[38] S. Luz, M. Masoodian, B. Rogers, Supporting collaborative transcription of recorded speech with a 3d game interface, in: Proceedings of the 14th International Conference on Knowledge-based and Intelligent Information and Engineering Systems: Part IV, KES'10, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 394–401.
URL http://dl.acm.org/citation.cfm?id=1893971.1894019

[39] S. Novotney, C. Callison-Burch, Cheap, fast and good enough: Automatic speech recognition with non-expert transcription, in: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, HLT '10, Association for Computational Linguistics, Stroudsburg, PA, USA, 2010, pp. 207–215.
URL http://dl.acm.org/citation.cfm?id=1857999.1858023

[40] M. Lux, M. Guggenberger, M. Riegler, Picturesort: Gamification of image ranking, in: Proceedings of the First International Workshop on Gamification for Information Retrieval, GamifIR '14, ACM, New York, NY, USA, 2014, pp. 57–60. doi:10.1145/2594776.2594789.
URL http://doi.acm.org/10.1145/2594776.2594789

[41] S. Egenfeldt-Nielsen, What makes a good learning game?: Going beyond edutainment, eLearn 2011 (2). doi:10.1145/1943208.1943210.
URL http://doi.acm.org/10.1145/1943208.1943210

[42] W. H. Z.O. Toups, A. Kerne, Motivating play through score, in: CHI Workshop Engagement by Design, ACM, New York, NY, USA, 2009.

[43] E. Z. Katie Salen, Rules of Play: Game Design Fundamentals, The MIT Press, 2003.

[44] M. Furini, V. Ghini, An audio-video summarization scheme based on audio and video analysis, in: Proceedings of IEEE Consumer Communications & Networking (CCNC2006), 2006, pp. 1209–1213.

[45] M. Furini, Fast Play: A Novel Feature for Digital Consumer Video Devices", IEEE Transaction on Consumer Electronics 54 (2) (2008) 513–520.

[46] S.-K. Kim, D. S. Hwang, J. Y. Kim, Y.-S. Seo, An effective news anchorperson shot detection method based on adaptive audio/visual model generation, in: Proceedings of the International Conference on Image and Video Retrieval (CIVR), 2005, pp. 276–285.

[47] T. Kemp, M. Schmidt, M. Westphal, A. Waibel, Strategies for automatic segmentation of audio data, in: Proceedings of the International IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2000, pp. 1423–1426.

[48] K. Johnson, Acoustic and Auditory Phonetics, 3rd Edition, Wiley-Blackwell, Malden, 2011.
URL `get-book.cfm?BookID=58781`