

# ClipBoard: Augmenting Movie Entertainment

Nicola Orio

Department of Cultural Heritage, University of Padua, Italy  
nicola.orio@unipd.it

**Abstract.** ClipBoard is a novel system aimed at improving the accessibility of movies. Although subtitles and audio description are already provided for some TV programs and by a limited number of movie theaters, the availability of accessible content still depends on the choice of the film distributor. Through ClipBoard it is possible to synchronize in real-time, using a portable device such as a smartphone or a tablet, the soundtrack which is audible to the audience with an annotated reference signal stored in the device. The goal is to improve movie enjoyment with the synchronous delivery of accessible content, which can be provided by third parties such as associations for disabled, public bodies, and volunteers.

## 1 Motivation

This paper describes a novel system, called *ClipBoard*, that aims at enriching movie enjoyment, with a particular focus on improving the accessibility for persons with visual or hearing impairments. To this end, we developed an automatic tool that allows a person, while watching a movie, to access in real-time to synchronized additional content. This content can consist in subtitles in different languages, audio description, pictures related to the particular scene, additional information regarding the plot, the directory, and the actors, and in general any kind of information that is pertinent to the movie.

First of all, we target accessibility. Watching a movie, either in a theatre or from a TV set, represents a social experience that should be possible disregarding the presence of sensorial impairments. In many European countries, although a number of laws aims at increasing accessibility, the amount of hours per week that are aired with subtitles and/or audio description is still less than 10% of the total. As regards movie theaters, there has been a significant effort by big chains of multiplexes to provide accessible projections yet the total number of accessible theaters is still low.

We believe than one major issue is related to the fact that the accessible content is usually delivered by the same body that projects/airs the movie. While mainstream movies shown in multiplexes or aired in prime time can be considered worth the cost of making them accessible, most of the audiovisual productions may not have enough popularity. We propose to decouple the process of delivering the movie with the one of providing the accessible content. The basic idea is to allow associations for disabled, public bodies or simply volunteers to

provide subtitles and audio description without having to deal with copyright issues. This can be achieved through an automatic real-time synchronization of the movie timeline while it is projected/aired, with the accessible content stored in and shown/played by the portable device of the disabled person.

It is important to notice that an application already exists for the delivery of automatic subtitles based on the identification of the soundtrack [4]. *MovieReading* is based on audio fingerprinting techniques that allows an initial synchronization of the subtitles, which afterwards proceed automatically and thus are based on a precise correspondence between the movie soundtrack and the reference one. Clearly this precision can be achieved only with digital cinema. As described in the following section, ClipBoard aims at overcoming this limitation introducing two additional modules, for fine synchronization and for the automatic detection and recovery of misalignments.

## 2 Audio-based Synchronization

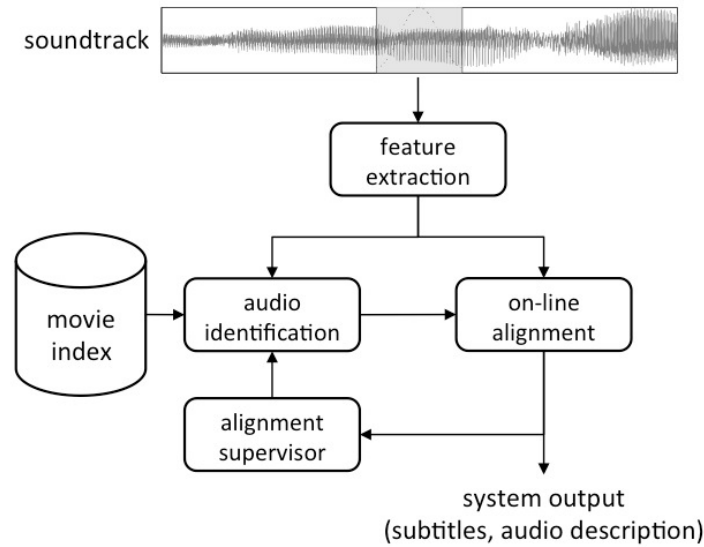
Automatic synchronization between the movie and the subtitles can be based on the soundtrack. Due to its immersive nature, the audio signal can be easily tracked in real-time with mobile devices, even the ones with limited computational resources. The user only needs to start the application, select the movie – or ask the program to identify it from a number of possible candidates – and let the device capture and analyze the soundtrack simply using its built-in microphone.

In order to be used, the system requires a reference signal, that is a copy of the movie soundtrack which is annotated with time-stamped subtitles and audio description. The actual synchronization is based on three main components, which are shown in Figure 1.

- An *audio identification* module.
- A *on-line alignment* module.
- An *alignment supervisor* module.

The audio identification module based on fingerprinting [1]. Audio fingerprints are low-dimensional audio descriptors that are robust to typical distortions of the audio signal due to lossy compression, additional environmental noise, and small differences between the reference signal and the soundtrack. Audio fingerprinting is a popular technique for music identification [5] and in ClipBoard it has been extended to deal with the audio signal of typical soundtracks: music, speech, and environmental sounds.

On-line alignment is based on a statistical representation of the audio signal, as proposed in [3]; in order to optimize the memory requirements, the audio descriptors for automatic alignment are an higher-dimensionality version of the audio fingerprints and are based on [2]. Also in this case, the approach has been extended to deal with movie soundtracks, where it is possible to have long parts with no dialog and with repetitive environmental noises if not complete silence. One of the main reasons for on-line alignment is that playback speed can vary



**Fig. 1.** The different software modules of ClipBoard and their communication process.

remarkably when the movie is projected/aired at different frame-rates or with analogue devices or when it has been initially recorded in analogue format.

Both modules require a reference signal, which is the complete soundtrack of the movie that has been previously annotated with subtitles and audio description. It is important to note that the reference soundtrack can differ from the one of the movie that is being projected/aired. For instance it can be based on an analogue copy, in VHS format, owned by the body who provides the accessible content, or be recorded from TV with a different frame-rate. Audio fingerprinting and especially on-line alignment techniques are usually robust enough to deal with this kind of differences.

Finally, the alignment supervisor analyses the output of the on-line alignment module and computes the probability of a local misalignment, due for instance to substantial differences between the actual soundtrack and the reference signal. These differences can be due to different cuts of the movie and, in particular, to the insertion of commercials in case of aired movies. The role of the alignment supervisor is to require a new identification in case of a long mismatch, which is typically highlighted by a non-linear advancement along the reference signal. It can be noted that such a module is required because it might not be evident for a disabled person when the system is not correctly synchronized.

### 3 Discussion

Even if audio processing has been already exploited for a number of music application, its potential in movie content access has not been fully explored yet. First of all, the audio descriptors are more compact and low-dimensional compared to the video ones, allowing the application to run also with limited computational power, bandwidth and power supply. Moreover, as already mentioned, the use of compact audio descriptors has the important side effect of decoupling the delivery of a movie, which has tight constraints due to obvious copyright reasons, with the sharing of the accessible content, which should be freely distributed. Each user may obtain a copy of the film through his/her preferred channels – going to a movie theater, buying a DVD, subscribing to a pay TV, watching the movie with inserted commercials – dealing directly himself/herself with copyright issues.

The software presented in this paper is the result of a collaboration with film scholars at the University of Padova who gave the guidelines for the content to be provided in order to improve movie enjoyment. The collaboration with experts in film studies highlighted additional applications of ClipBoard. In fact, it can be applied to the development of current film studies. Much work has been carried out on the description of music content, for instance in music tagging, but its extension to movies description and the possibility to annotate portions of video are still under-investigated.

As it is well known, users can already access to extra content, usually distributed with the official DVD or Blue-Ray of the movie. Yet this is possible only when an official distributor provides them to the public, while ClipBoard enables scholars to make available extra content through a different channel.

### Acknowledgments

The authors wishfully thank Dr. Farah Polato of the Department of Cultural Heritage and Dr. Denis Brotto of the Department of Linguistic and Literary Studies, both at the University of Padova, for the fruitful conversations on the possible applications of ClipBoard.

### References

1. P. Cano, E. Batlle, T. Kalker, and J. Haitsma. A review of audio fingerprinting. *Journal of VLSI Signal Processing*, 41:271–284, 2005.
2. S. Dixon and G. Widmer. MATCH: a music alignment tool chest. In *Proc. of the International Conference on Music Information Retrieval*, pages 492–497, 2005.
3. N. Orio. Automatic identification of audio recordings based on statistical modeling. *Signal Processing*, 90(4):1064–1076, 2010.
4. Universal Multimedia Access. MovieReading. <http://www.moviereading.com/>, 2011.
5. A. Wang. The Shazam music recognition service. *Communications of the ACM*, 49(8):44–48, 2006.