# TV Advertisements Detection and Clustering based on Acoustic Information

## Abstract

*Detection and clustering of commercial advertisements plays an important role in multimedia indexing as well as in the creation of personalized user content. Its aim is at detecting individual commercials within a broadcast and grouping together all repetitions of the same commercial over time. Several algorithms are found in the literature to tackle the detection task using either video and audio or only video cues, but none has been found for clustering. In this paper we present an acoustic-only system to perform both the detection and clustering of commercials. On the one hand, detection is done in three steps, incrementally refining an initial coarse energy detection. On the other hand clustering is later done over all previously detected commercials to find out how many times each commercial appears. Our detection system achieves* $82$*% precision and recall using only acoustic information. For the clustering step, three algorithms are compared, obtaining best results using a modified dynamic Time Warping approach, which achieves* $100$*% recall and* $99$*% precision.*

## 1 Introduction

Even though TV commercials are never very well considered by the audience, they have a great influence in nowadays broadcasting industry. Automatic detection and clustering of such content has many applications both in the professional and personal areas. On one hand advertised companies have an interest in monitoring how many times and when their advertisements have been aired. On the other hand personal users or archival centers are interested in eliminating such advertisements when recording the content on their digital media centers. This also opens a new way for audiovisual advertisements distribution by being able to detect and substitute commercials targeted to the user's personal preferences.

Currently most of the advertisements detection and clustering for monitoring purposes is performed by human professionals in a way that becomes tedious and time consuming. Being able to automatize this process with sufficient performance would ease its processing and allow for many applications like the ones described in the first paragraph. Currently, to the authors' knowledge, there does not exist any published system to perform both the detection and the clustering (detection of repetitions) for commercials. In this paper we propose a system that is able to detect start-end points for broadcasted advertisements and to find how many times they are repeated over time. The implementation of personalized advertisements is an evolution of the current system and will be presented in a future publication.

In order to detect commercials on TV some efforts have been already made using either video or audio+video. When using video alone [4] [6] [10] a combination of rules identifying the dynamics of commercials insertion by the broadcasting companies and image features are used, for example searching for black frames or shot-cuts rate average. These systems are usually computationally expensive and cannot achieve the performance of systems using audio features. Other authors [5] [8] [2] [3] propose combined audio-visual methods. In [3] they exploit the repetition of commercials over time using video and refine the results using audio features while in [2] both audio and video features are analyzed for repetitions. Such approaches fail whenever non-commercial segments are repeated (for example in news programs). In [8] black video frames and audio energy are used together with a rule-based decision algorithm, with several fine-tuned thresholds. In [5] a set of visual and acoustic-based features are combined with an SVM (Support Vector Machine) classifier for every detected video shot. In doing so they consider that all commercials contain common audio-video features that tell them different from regular content, which is not necessarily true in all cases.

The proposed system performs a detection and clustering of commercials using only acoustic information. In doing so it results in a much more computationally effective system than using some sort of video analysis while still achieving a good performance. The system initially performs a detection of possible commercial edges via detection of strong average signal energy depression. It then ap-

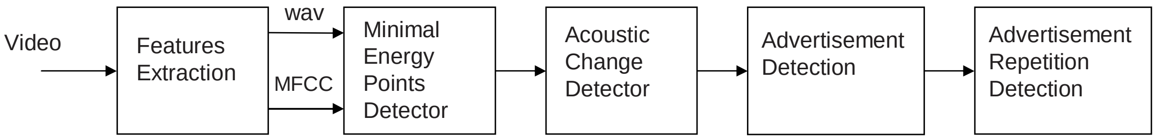


**Figure 1. Schematic diagram of advertisements detection and clustering.**

plies two filtering steps to eliminate false alarms and return the detected commercials. In a first step the Bayesian Information Criterion (BIC) distance [9] is used to filter out changes found due to video edits where the same speaker or acoustic conditions appear in both sides. A second detection step enforces certain possible commercial durations which were found to be most common in TV broadcasts.

For commercials clustering, the detected advertisement is searched on a database created from advertisements previously found in other recordings. If no similar advertisement is found the new commercial is included on the database. In order to determine wether two commercials are the same three possible algorithms are evaluated. Two of them are based on the Dynamic Time Warping (DTW) algorithm [7] and a third one uses cross-correlation.

The paper is structured as follows. Section 2 introduces the acoustic-based advertisements detection system and section 3 describes the clustering system. Section 4 describes the experimental database and Section 5 explains the web database application developed to show the system analyssis and classification results. Finally some conclusions are drawn.

## 2 Acoustics-based Advertisement Detection

The advertisement detection system presented here is based only on the analysis of the acoustic signal and detects commercials on television broadcasts based in two facts.

On one hand, advertisement breaks are usually isolated from actual programme material by a decrease in the audio signal occurring before and after each individual advertisement. Usually these silences last from $10$ to $30$ milliseconds and are digital nulls when advertising agencies and broadcasters use digital equipment. However, it is possible, and maybe quite probable, that these energy drops also occur during the valuable material of the programme itself. For example, they are not uncommon when news programmes go back and forth from anchor person to news reports, during scene changes within a soap opera or in shortened interviews.

On the other hand, advertisements usually have standard, defined lengths, they last $5$s, $10$s, $15$s, $20$s ... Although there are some exceptions, like TV channels self-promotions, very long TVShop-like commercials, etc... In this study the lengths $10$s, $20$s and $30$s have been considered for detection as they correspond to more than $88\%$ of the total number of advertisements labeled in $14$h $50$min of broadcasted data. The total length distribution on the database can be seen in Table 1

In order to efficiently locate an advertisement, after extracting the acoustic signal and its MFCC parameters from the video file, a three-stage approach is used, as depicted in Figure 1. First the minimum energy points within the audio signal are found as hypothetical commercial start/end changes. Then a validation of the candidates is performed by checking if there is an acoustic change at each point by acoustically comparing both sides for each candidate using the Bayesian Information Criterion (BIC) Algorithm [9].

On step three the proper selection of advertisements is made. To do so, first is necessary to find out precisely the boundaries of the connecting silences. This is done to eliminate the random amount of silence usually inserted between commercials. Afterwards, the distance between any two start-end marked point is compared with the set of allowed advertisement lengths, which for this study is $10$, $20$ and $30$ seconds with a small error margin allowance.

The resulting segments are considered to be commercials and are sent to the clustering step. In the following sections each of the algorithms cited above are explained in more detail.

### 2.1 Minimum energy points detection

The main acoustic cue used in this work to find commercials is the sudden drop in energy, common before and after commercials, mainly due to the need for connection regions between regular content and advertisements, and between advertisements.

In order to detect such change points, the energy average of the input signal is computed using a very narrow window. The narrowness of the window allows for detection of very low energy points while not triggering on false energy drops. A restrictive threshold is used to determine possible change points. Each energy minimum below the threshold is selected as a change point, and a mask around it is applied in order to avoid multiple triggers for the same advertisement.

| $C_l$ | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 60 | 80 | 90 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # ads | 11 | **127** | 11 | **179** | 12 | **41** | 3 | 2 | 4 | 2 | 1 | 1 |
| %total | 3% | **32%** | 3% | **45%** | 3% | **10%** | 0.8% | 0.5% | 1% | 0.5% | 0.2% | 0.2% |

**Table 1. Advertisements length found in development and test database**

### 2.2 Acoustic change detection

Using energy drop as the queue to hypothesize commercial change points leads to a usually high rate of false alarms. In nowadays with so many digital mixing equipment, apart from finding abrupt energy drops between advertisements, they can be found also in the silences between segments on an edited speech -for example during long interviews, where only a few segments are glued together for transmission - or in long studio silences. We noticed that most of these cases have the same speaker (or at least the same acoustic background conditions) present in both sides of the hypothesized change, which is not usually true in commercial change points as they refer to totally different products.

In order to filter out these false alarms an acoustic change point detection algorithm based on the Bayesian Information Criterion (BIC) [9] is used. For each possible commercial change found in previous stages two hypothesis are modeled and compared. On the one hand $H_0$ considers that both sides of the change point ($\mathcal{X}_a$ and $\mathcal{X}_b$) share the same acoustic environment/belong to the same speaker. On the other hand $H_1$ considers that both sides belong to different acoustic environments/speakers. Each hypothesis is modeled by Gaussian Mixture Models (GMM) following the BIC modification proposed by [1] where $H_1$ was modeled by a GMM per side ($\Theta_{1a}$ and $\Theta_{1b}$), with eight Gauss. each, and $H_0$ was modeled with 16 Gauss. ($\Theta_0$) modeling the acoustic data in either side ($H_1$) or both ($H_0$). Modeled data is composed of MFC coefficients extracted from the acoustics with 26 coefficients and computed every 10ms. BIC distance ($\Delta BIC$) is computed as shown in Eq. 1.

$$\begin{aligned}\Delta BIC(H_0, H_1) = BIC(H_0) - BIC(H_1) = \\ \mathcal{L}(\mathcal{X}_a, \mathcal{X}_b|\Theta_0) - \mathcal{L}(\mathcal{X}_a|\Theta_1 a) - \mathcal{L}(\mathcal{X}_b|\Theta_1 b)\end{aligned} \quad (1)$$

According to the $\Delta BIC$ distance, all hypothesized change points with positive values are not considered anymore as possible commercial changes.

### 2.3 Advertisements filtering and detection

Usual advertisement slots being sold by broadcasting companies fall within a few well defined set of possible lengths $C_l$. Table 1 shows the number of actual advertisements and their % of the total for nearly 15 hours of labeled material (used as development and test in this paper). As mentioned, in this study only the subset $C_l^{'} = \{10, 20, 30\}$ are considered, as they cover over $88\%$ of cases.

In order to detect commmercials, for every change point remaining from previous filters, is validated if the distance D to another change point is

$$\left|D - c_l^{'}\right| < \epsilon \quad (2)$$

where $c_l^{'}$ belongs to $C_l^{'}$ and $\epsilon$ is the accepted time error, which corresponds to the imprecision at detection of the beginning and the end of an advertisement. Its value is independent of the actual commercial length and in order to eliminate as many false alarms as possible, it is necessary to reduce its variance.

If more than one possible distance is found, only the smallest is considered (for example there could be three $10$s commercials within a $30$s slot). The simplicity of this algorithm comes at a small cost, for example, on a sequence of $3$ advertisements: $5$, $10$ and $5$ seconds length each one, two advertisements would be selected. The first one, with $20$ seconds, would be a false alarm containing all three advertisements; the second one, with $10$ seconds, would be correct.

## 3 Advertisement repetitions detection

Once the advertisements have been detected within the input video, they are compared with all the commercials of the same length on the database. If no commercial on the database is found to be equal to the new detected advertisement, this advertisement is included as a new one. In order to compare similarity between commercials three different methods were evaluated in this paper: Standard Dynamic Time Warping (DTW) [7], a simplified DTW (DTWmod) algorithm and a Generalized Cross-Correlation (GCC) comparison between signals.

For DTW and the DTWmod the same MFCC parameters computed earlier are used. In order to improve the system performance, the region of possible frame to frame alignments in DTW is restricted by applying a global constraint composed by a Sakoe-Chiba band mask of radius equal to the differences between the length of the advertisements.

Although DTW has extensively been used to find the optimum warp between two signals which are similar, it was seen in this application that in the regions where the two commercials are identical (i.e. everywhere except the initial-ending silence regions) such freedom to choose an appropriate warping was always reduced to the diagonal frame assignment. Therefore in this application a simplification of the DTW was used. On this modified DTW (DTWmod) DTW diagonal that obtains the minimum total error is found. In this implementation insertions/deletions at the beginning/end of the signal pairs during 1s at maximum and their distances are not added to the final distance. In order to retrieve comparable final distances they are normalized by the number of frames in each considered diagonal.

Finally, the third metric (GCC) corresponds to a standard cross-correlation implementation. For each signal pair the maximum of the cross-correlation between them is found, and normalized by their power. The bigger the correlation the more similar both advertisements are. In order to compare this metric with the other two, the inverse of the normalized maximum cross-correlation is taken as the distance measure.

# 4 Experimental Results

## 4.1 Databases

For system optimization and evaluation two databases have been collected and labeled. Table 1 shows the number of actual advertisements and their % of the total nearly 15h of labeled material (used as development and test).

The development database contains 8 video files extracted from 6 different TV channels. They have been aired at different hours, and contain different types of programs: 4h 5min of news, 1h 55min sport transmission, 2h 55min magazines. The total length of the database is of 8h and 55min and it contains 212 advertisements of 10, 20 and 30 seconds, that last 1 hour 3 minutes. Although most of the database is in Spanish, the sport retransmission is in Catalan.

Test database contains three video files extracted from three different TV channels. It contains 1h of news, 50 min of magazine and 2h of prime-time (sketches and reality programs). The total length of the database is of 3h and 50min and it contains 135 advertisements, which last 38 minutes.

## 4.2 Detection results

Different tests were performed on the development database in order to adjust the system parameters. Precision (PRC), recall (RCL) [8] and $F = 2*PRC*RCL/(PRC+RCL)$ were used to check the system. The objective was to determine parameters to get both good PRC and RCL results.

| $det.$ | #ads | #det. | PRC | RCL | F |
|---|---|---|---|---|---|
| dev | 212 | 181 | 85.38% | 85.38% | 85.38% |
| test | 135 | 112 | 81.16% | 82.96% | 82.05% |

**Table 2. Advertisement detection results. First column indicates number of advertisements on databases, second one the number of correctly detected advertisements**

From the results shown in Table 2, the system identifies correctly 82% of the advertisements on the test database. These results are quite good, since the algorithms used are relatively low-time consuming. From the development set analysis, it was observed that optimal detection parameters were different for each TV channel, so instead of using one parameter for all channels as done here, using different parameters for each one of them would further increase the system performance. This assumption seems reasonable since each channel uses a different systems to chain its programs and advertisements.

## 4.3 Repetition detection results

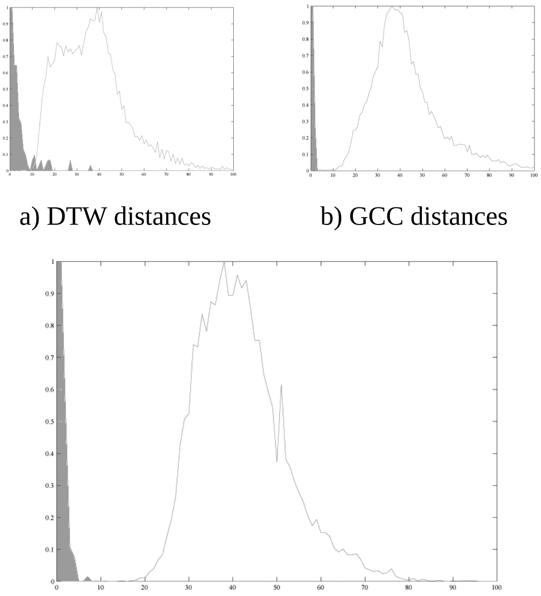
On the advertisement repetitions detection study both databases are used: development and test. This approach is followed for two reasons: there is not enough material on the test database, there are only 16 repeated advertisements; and no threshold or parameter is applied to detect repetitions.

| $Clus$ | GCC | DTWod |
|---|---|---|
| PRC | 97,37% | **99,12%** |

**Table 3. PRC for 100% RCL threshold**

For each of the advertisements on the database the difference with respect to all other advertisements is computed. Three different algorithms are applied: DTW, DTWmod and GCC. On Figure 2 are depicted the normalized histograms of the distances computed with the different algorithms. Distances between repetitions of the same advertisement are shadowed, and the distances between different advertisements are blank filled.

As it can be observed on Figure 2, for DTWmod and GCC there is a considerable distance between the repeated and different advertisements. Table 4 shows the ratio between means of both distances. As it can be observed, DTWmod achieves the best performance $(40, 78)$.



a) DTW distances

b) GCC distances

c) DTWmod distances

**Figure 2. Histograms for repeated (shadowed plot of each figure) and different (no-filled) advertisements**

Advertisements classified as different with lower distances were manually checked. Some of them belong to the same advertisement but in a different language (they belonged to different channels in different languages). Others belong to advertisements of the same commercial campaign, sharing part of the content. For instance two different $10$ second advertisements were detected that shared almost $7$ seconds of their video/audio signal.

For the best two metrics, GCC and DTWmod, Table 3 also shows the precision that it would be obtained if a threshold was set for each metric to achieve $100\%$ recall. Precision for DTWmod would be $99.12\%$ and for GCC would be $97,37\%$. It would not be possible to select such threshold for DTW, because precision would drop dramatically.

Even more, for DTWmod and GCC, if different thresholds are applied for different $C_l^{'}$ no precision errors would be found. Future studies could check if it would be possible to recognize advertisements of the same campaign, meaning advertisements with small modifications, using a combination of GCC and DTW_mod.

| $Clus$ | DTW | GCC | DTWod |
|---|---|---|---|
| $\mu_{dif}/\mu_{rep}$ | 8,69 | 27,86 | **40,78** |

**Table 4. Ratio between means of repeated and different advertisement distances.**

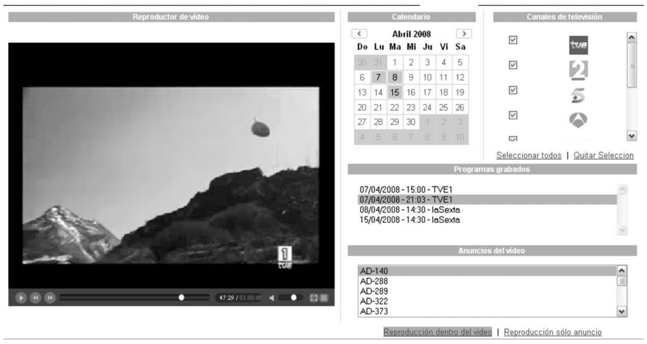
## 5 Visualization

In order to visualize the results obtained by the described algorithms a web application has been developed. It extracts the information from the recorded database and from the advertisements description files and presents it to the user in a visual way, allowing for an intuitive visualization of the algorithms results.

The detection and clustering results for all analyzed advertisements are stored on two different XML files. One, associated directly to the multimedia file, and containing its acoustical description (description file). Another one, more application-related, associated to the database (clustering file), containing information of the clustered advertisements.

The use of XML files not only allows easy access to the indexed multimedia content, but also provides a system to incrementally refine, complete, increase, enhance the description of the file, applying different algorithms, based on different technologies and it can all be done off-line.

The advertisement visualization application has three different views:


Reproductor de video
Calendario
Canales de televisión
Seleccionar todos | Quitar Seleccion
Programas grabados
Anuncios del video
Reproducción dentro del video | Reproducción sólo anuncio

Abril 2008

| Do | Lu | Ma | Mi | Ju | Vi | Sa |
|---|---|---|---|---|---|---|
| 30 | 31 | 1 | 2 | 3 | 4 | 5 |
| 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| 20 | 21 | 22 | 23 | 24 | 25 | 26 |
| 27 | 28 | 29 | 30 | 1 | 2 | 3 |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 |

☑
☑
☑
☑
☑

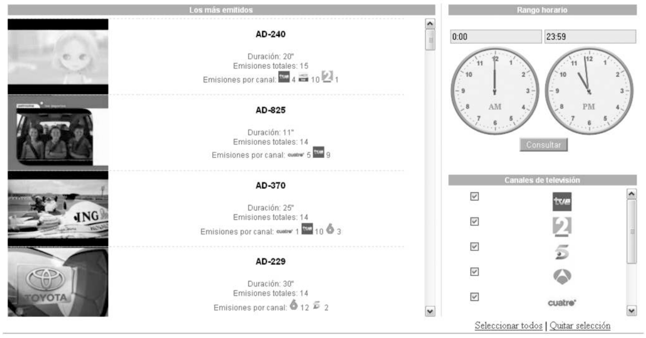
07/04/2008 - 15:00 - TVE1
07/04/2008 - 21:03 - TVE1
08/04/2008 - 14:30 - laSexta
15/04/2008 - 14:30 - laSexta

AD-140
AD-288
AD-289
AD-322
AD-373

**Figure 3. Multimedia description screenshot.**

1) Multimedia description (Figure 3): It allows to search a recorded file, and it shows the advertisements detected on it. On the embedded player it is possible to manually check all content of the file - *i.e.* not only the advertisements detected, but also if an advertisement has not been detected.

2) Clustering (Figure 4): It lists the most repeated advertisements detected on the database, and presents to the user the number of times each advertisement has been repeated



AD-240
Duración: 20"
Emisiones totales: 15
Emisiones por canal: 4 10 1

AD-825
Duración: 11"
Emisiones totales: 14
Emisiones por canal: 5 9

AD-370
Duración: 25"
Emisiones totales: 14
Emisiones por canal: 1 10 3

AD-229
Duración: 30"
Emisiones totales: 14
Emisiones por canal: 12 2

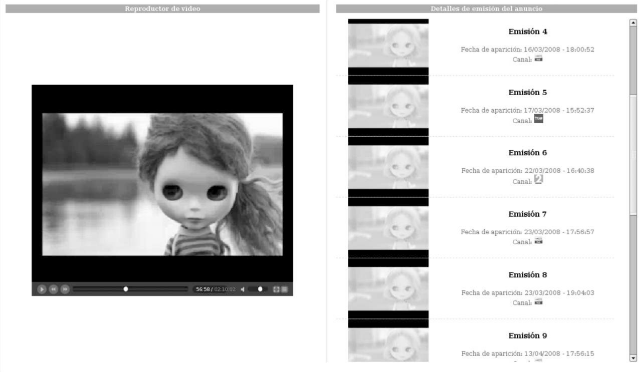
0:00 23:59

☑ ☑ ☑ ☑ ☑

**Figure 4. Clustering screenshot.**

for each of the analyzed channels. It is possible to filter the results by tv channel and the time it was aired.

3) Advertisement details (Figure 5): It lists all the repetitions of a selected advertisement. The embedded player allows for playing of each individual advertisement.



Emisión 4
Fecha de aparición: 16/03/2008 - 18:00:52
Canal:

Emisión 5
Fecha de aparición: 17/03/2008 - 15:52:37
Canal:

Emisión 6
Fecha de aparición: 22/03/2008 - 16:40:38
Canal:

Emisión 7
Fecha de aparición: 23/03/2008 - 17:56:57
Canal:

Emisión 8
Fecha de aparición: 23/03/2008 - 19:04:03
Canal:

Emisión 9
Fecha de aparición: 13/04/2008 - 17:56:15

**Figure 5. Advertisement detail screenshot.**

## 6 Conclusions and Future Work

Automatic detection and clustering of commercial advertisements plays an important role in multimedia indexing as well as in the creation of personalized user content. Its applications range from the analysis of broadcasted advertisements to the suppression of undesired content or its interchange for targeted audiences. In this paper a video advertisement detector has been introduced based only on acoustic features. This system not only allows to detect advertisements, but it also allows classification and determination of how many times each advertisement has been repeated. The performances achieve $82\%$ and almost $100\%$ precision/recall for advertisements detection and repetitions detection.

Extending these algorithms to longer databases will allow for new developments to improve the precision, for instance: if a one week database is obtained, all non repeated advertisements on that database should be considered as false alarms, because advertisements always repeat themselves. It could also be interesting to implement these algorithms for real time advertisement detection and identification.

Next steps could also include some simple, low-time consuming image processing techniques to improve the precision and recall of the overall system, and to incorporate a new function: detect and substitute commercials targeted to the user's personal preferences.

## References

[1] J. Ajmera, I. McCowan, H. Bourlard. Robust speaker change detection. Tech. report, IDIAP, 2003.

[2] M. Covell, S. Baluja, M. Fink. Advertisement detection and replacement using acoustic and visual repetition. In *Proc. IEEE 8th Workshop on Multimedia Signal Processing*, pp. 461-466, Oct. 2006.

[3] P. Duygulu, M. yu Chen, A. Hauptmann. Comparison and combination of two novel commercial detection methods. In *Proc. ICME*, Taiwan, 2004.

[4] A. G. Hauptmann, M. J. Witbrock. Story segmentation and detection of commercials in broadcast news video. In *Proceedings ADL'98*, Santa Barbara, USA, 1998.

[5] X.-S. Hua, L. Lu, H.-J. Zha. Robust learning-based tv commercial detection. In *Proc. ICME*, 2005.

[6] R. Lienhart, C. Kuhmnch, W. Effelsberg. On the detection and recognition of television commercials. In *Proc of IEEE Conference on Multimedia Computing and Systems*, pages 509–516, Otawa, Canada, 1997.

[7] C. Myers, L. Rabiner. A comparitive study of several dynamic time-warping algorithms for connected word recognition. *The Bell System Technical Journal*, 60(7):pp. 1389–1409, 1981.

[8] D. A. Sadlier, S. Marlow, N. O'Connor, N. Murphy. Automatic tv advertisement detection from mpeg bitstream. *Journal of the Pattern Recognition Society*, 35(12):2–15, 2002.

[9] S. Shaobing Chen, P. Gopalakrishnan. Speaker, environment and channel change detection and clustering via the bayesian information criterion. In *Proceedings DARPA Broadcast News Transcription and Understanding Workshop*, Virginia, USA, 1998.

[10] J. Snchez, X. Binefa. Audicom: a video analysis system for auditing commercial broadcasts. In *Proc. of ICMCS'99*, Firenze, Italy, 1999.