An Efficient Algorithm for Automatic Television Broadcast Monitoring

Arsalaan Ahmed Shaikh, Hammad Qureshi
School of Electrical Engineering and Computer Science
National University of Science & Technology
Islamabad, Pakistan
10msitashaikh@seecs.edu.pk, hammad.qureshi@itu.edu.pk

ABSTRACT: *The last two decades have seen an unprecedented rise in the number of terrestrial cable and satellite TV channels. These channels generate a huge amount of digital content and are the main medium for businesses to advertise their products in a globalized world. Broadcast monitoring is an essential activity which involves evaluating whether the correct content was aired and for the correct length of time and at the time previously agreed upon. As advertisement forms the bulk of the revenue of the television channels, monitoring of advertisements becomes extremely important. Advertisers demand efficacy in the airing of their content and payments are made only when the claims are verified. Lack of efficient and inexpensive monitoring technologies directly impacts growth in advertisement revenues. Monitoring is a tedious task which involves monitoring and analysis of thousands of hours of audio/visual content forming terabytes of data. Hence, efficient automated techniques are required for an industry which relies mostly on manual auditing.*

Heterogeneity in aired advertisements, signal quality, sizable amount of multimedia data, and demand for high accuracy makes it a challenging problem to solve. In this paper, we propose a set of audio features which may be used to perform automatic auditing of broadcast content. A feasibility analysis is conducted based on a proposed Average Dependency & Minimum-Maximum Distance criterion that judges the ability of a feature to differentiate between classes of advertisements. The criteria is combined with sequential floating search method to obtain an optimal or nearoptimal feature subset. The efficiency and effectiveness of audio features is evaluated on a real-world dataset of 216 hours of television broadcasts and 28 classes of advertisements. The results show that Gabor Filter Bank Feature (GBFE), Mel-frequency Cepstral Coefficient (MFCC) and MPEG7 Audio Flatness Mean (having yielded overall recognition rates of 99.33, 98.99 and 97.31 respectively) are the most suitable audio features for an automated television broadcast monitoring system.

Keywords: MC-CDMA (Multi-Carrier Code Division Multiple Access), QAM (Quadrature Amplitude Modulation, BER (Bit Error Rate), SNR (Signal-to-Noise Ratio), LLR (Log Likeli-Hood Ratio)

Received: 27 June 2014, Revised 30 July 2014, Accepted 4 August 2014.

1. Introduction

Advertisement air time is expensive and advertisers are reluctant to pay high rates unless an objective auditing mechanism is in place. The mechanism provides the basis on which broadcasters and advertisers resolve their claims. In traditional broadcast monitoring, an auditing company (on request of clients) provides logs of specific advertisements that aired during transmission. The logs contain the time of day the advert was aired, duration and channel of the advert. The purpose of the logs is to identify deviations from the advert broadcast plan originally agreed upon by the advertisers and broadcasters. Aggressive competition amongst the broadcasters (as the growth in satellite television channels is rapidly increasing) requires the logging of all non-commercial and commercial transmissions. A monitoring company maintain logs of the complete day's transmission for a channel. Manual auditing (i.e. using human effort) is used which is not cost-effective as there are a large number of channels and quantity of multimedia data is huge requiring a large workforce. Moreover, manual auditing is error prone as per the system logs from [1]. An automatic solution is not straightforward as the quality of transmission varies from channel to channel and from time to time. Table I shows difference of audio attributes of an identical advertisement that was aired on two different channels. Similarly differences can also be found in the transmission of the same television channels at different points in the day. Due to the variation in transmission quality it is important to select a robust feature set. Moreover, the huge amount of data in television broadcasts with the added requirement of high recognition rates, robustness and scalability makes the problem a challenging task. In this paper, we propose an algorithm that acquires a robust feature set to audit audio content over many different channels. Such an algorithm has not been proposed in the literature before.

The recorded transmission obtained for auditing contains both audio and video data. Audio data, as compared to video data, is much smaller in size and has less variation across television channels, there by, providing better scalability and accuracy. Moreover, video and audio data are both equally relevant and offer a means for accurate auditing of advertisement as discussed in [2], [3], [4]. To further reduce the size of audio data, a process known as feature extraction is used. Feature extraction transforms the input data into a reduced representation. The features (usually represented in the form of vectors) are used to differentiate one class of object from another. Feature selection is then used to acquire a robust set of features for efficient and high accuracy results. It involves selection of a useful subset of features that represent patterns from a larger set of often redundant and possibly irrelevant features [5]. Classification function then consigns an input object to one of a set of classes [5].

Rest of the paper is organized as follows: related work is given in Section 2, an overview of audio features and methodology used is given in 3 while the results are explained in Section 4. Finally concluding remarks are given in Section 5.

2. Related Work

Broadcast monitoring these days is carried out both manually and using automatic computer-based systems. Manual systems such as the one used by Media Bank Pakistan [1] employs human operators to monitor television broadcast content and creates logs in a computer-based database. Manual auditing is prone to error and susceptible to variations due to the training level of auditors and their skills. Since human labour and effort required is great due to the huge volumes of data such systems become very expensive to operate.

A lot of work has already been carried out on the development of electronic broadcast monitoring systems (EBMS). Commercial importance of such systems has lead to multiple patents in the area [9], [10], [11]. EBMS generally fall in two categories; in the first category the systems are designed on pre broadcast information. Welsh et al. [11] proposed a video monitoring system that functions by comparing closed caption text (program content description available in NTSC signal) with stored text to detect commercial and program content. In another paper by Ton et al. [12], a system for multimedia copyright infringement detection has been proposed. The system uses a hidden watermark code embedded inside the aired content. Once broadcast signal is received the system detects the code and identifies the advertisement. Systems described in [11] and [12] require the participation of broadcasters and advertisers and broad acceptance by the industry, hence, it is not readily implementable.

The second category of EBMS, analyze broadcast signal properties to detect advertisement content automatically using computer-based analysis. These systems extract audio and video properties from signals and compare them with a pre extracted feature database. Large amount of data is processed during feature extraction and advertisement segmentation stages. Such systems have been described by Oliviera et al. [2] and Camarena et al. [3]. In [2], a short-term Fourier transform is applied on broadcast audio data. The process results in a series of vectors (i.e. features) that are further processed to extract clusters. The clusters are then processed in an advertisement segmentation phase. Experiments were performed on a small sample set of 41 hours of broadcast transmission for which an accuracy rate of 95% was achieved. In the other paper, Camarena et al. [3] propose a system based on Multi-Band Spectral Entropy fingerprint, the broadcast signals are processed and the features extracted are compared with a library of extracted advertisements. Both systems do not resolve the issue of cross channel (segmentation using same source advertisement for all television channels) advertisement segmentation which limits their scalability and robustness on large number of channels.

3. Methods

The system proposed in this paper is an automatic system which is based upon feature extraction using audio part of the advertisement and then a pattern recognition algorithm is used to match the advertisement. If a match (based upon a predefined criteria) is found, the system creates a log of it in the database. The proposed solution is explained in two sections. Audio feature extraction, assessment and selection in which different audio features are evaluated followed by an advertisement segmentation algorithm that uses a sliding window method for fast processing. Both of these sections are explained below in detail.

3.1 Feature Extraction

We experimented with various features. The features that performed better than the rest are given below:

3.1.1 Mel-frequency Cepstral Coefficient: Mel-frequency Cepstral Coefficient (MFCC) is a feature extraction technique. MFCC is a frequently used technique in applications in which speech or speaker identification is involved such as [13] and [14]. MFCC takes human perception sensitivity with respect to frequencies into considerations [15]. Many different versions of the MFCC extraction algorithms have been proposed by [16], [17] and [18], for this study the implementation proposed by [19] has been used.

3.1.2 MPEG Content Multimedia Description Interface: MPEG (Moving Pictures Experts Group) content multimedia description interface commonly known as MPEG-7 is an ISO/IEC standard. The specification is developed for structuring multimedia content for fast and easy retrieval. The descriptors are classified in low-level and high-level categories. Lowlevel descriptors describe the basic attributes of the content such as its color, size, power, energy, etc. On the other hand high-level descriptors contain a higher semantic hierarchy. High-level descriptors require human intervention such as scene definition [20]. Therefore only low-level descriptors have been considered. This comparative study utilizes the implementation proposed in [21].

3.1.3 Gabor Filter Bank Features: A relatively new audio feature has been proposed by [22] known as Gabor Filter Bank Features (GBFE). GBFE extracts spectro-temporal information from the audio signal. The features have already been used for various audio applications such as automatic speech recognition systems [22] & [23] and speaker recognition [24] etc. GBFE uses a set of 41 filters (Gabor Filters) which have been developed based upon human physiology [22]. During the feature extraction process the log Mel-spectrum of the audio signal is calculated, which is filtered using the Gabor Filters resulting in a 311-dimensional feature vector.

3.2 Audio Feature Assessment

This section assesses audio features based on the average dependency and minimum-maximum distance criteria (explained below). These benchmarks establish the utility of a feature to differentiate between aired content. Furthermore, feature subset selection (based on Average Dependency criteria) is also discussed.

3.1.2 Average Dependency Criterion: Consider two sets *A* & *N* and a dependency function *D*. Set *A* consists of m number of advertisements belonging to the same class (same advertisements aired at different instances) while set *N* consists of $m - 1$ elements belonging to several different classes (program or advertisement content other than that of set *A*). The dependency function *D* determines the correlation, similarity or distance between the members of set *A* such as $D(A_i, A_j)$ where $i \neq j$. The function also calculates the dependency between the members of sets *A* and *N* such as $D(A_i, N_j)$.

Using the dependency function *D*, a true dependency matrix *T* and false dependency matrix *F* are calculated. Matrix *T* is formed by calculating dependency between all elements of set *A* while matrix *F* is formed by calculating the dependency between all elements of sets *A* and *N*. The matrices are shown in figure 1. Once the matrices are formed the average *avgT* and *avgF* of all elements of *T* and *F* are calculated respectively. Finally by subtracting *avgT* and *avgF* the Average Dependency (AD) is calculated.

To evaluate a feature, AD is calculated and then averaged across *k* number of advertisement classes. A feature is judged based on its ability to attain a large AD across different classes of advertisements.

Audio Attribute	Channel-A	Channel-B
DC offset	-0.000021	-0.000002
Min level	-0.248169	-0.359497
Max level	0.224548	0.355316
Pk lev dB	-12.11	-8.89
RMS lev dB	-28.5	-24.7
RMS Pk dB	-22.61	-18.06
RMS Tr dB	-61.5	-58.01
Bit-depth	14/16	15/16

Table 1. Audio Difference in Identical Advertisement

$T =$

$D(A_1,A_2)$	$D(A_1,A_3)$	$D(A_1,A_4)$	$D(A_1,A_m)$
$D(A_2,A_1)$	$D(A_2,A_3)$	$D(A_2,A_4)$	$D(A_2,A_m)$
$D(A_3,A_1)$	$D(A_3,A_2)$	$D(A_3,A_4)$	$D(A_3,A_m)$
$D(A_4,A_1)$	$D(A_4,A_2)$	$D(A_4,A_3)$	$D(A_4,A_m)$
$D(A_m,A_1)$	$D(A_m,A_2)$	$D(A_m,A_3)$	$D(A_m,A_{m-1})$

$F =$

$D(A_1,N_1)$	$D(A_1,N_2)$	$D(A_1,N_3)$	$D(A_1,N_{m-1})$
$D(A_2,N_1)$	$D(A_2,N_2)$	$D(A_2,N_3)$	$D(A_2,N_{m-1})$
$D(A_3,N_1)$	$D(A_3,N_2)$	$D(A_3,N_3)$	$D(A_3,N_{m-1})$
$D(A_4,N_1)$	$D(A_4,N_2)$	$D(A_4,N_3)$	$D(A_4,N_{m-1})$
$D(A_m,N_1)$	$D(A_m,N_2)$	$D(A_m,N_3)$	$D(A_m,N_{m-1})$

$$avgT = \frac{\left(\sum_{i=1}^{m} \sum_{j=1}^{m-1} Tij \right)}{m * m - 1}$$

$$avgF = \frac{\left(\sum_{i=1}^{m} \sum_{j=1}^{m-1} Fij \right)}{m * m - 1}$$

Figure 1. True and false dependency matrix

3.2.2 Minimum-Maximum Distance Criterion: The transmission broadcast has variation in signal quality, therefore there lies a possibility that some elements belonging to *T* may have lower values than that of elements in *F* (even though overall AD is high). To restrict this possibility the Minimum- Maximum Distance (MMD) criterion is used. MMD is calculated by subtracting the minimum element of *T* with the maximum element of *F*. A large positive MMD value attained by an audio feature establishes its degree of robustness.

3.2.3 Assessment: To assess the audio features 15 sets of advertisements were prepared. The cardinality *m* of each set of advertisement was 15 (making a total of 225 sam- ples). The samples were cropped from 216 hours of recorded transmission belonging to different television channels. The transmission contained quality variation which is also reflected in the cropped advertisement samples. Selected samples were heterogeneous in nature encompassing sequences of speech, music, speech-music overlay and signal variation. The set *N* was also prepared from the same transmission; however the samples were cropped from sequences other than that of the advertisement sets.

Audio Feature	Dim	AD	MMD
Audio Flatness Variance	16	0.5557	0.3103
Audio Flatness Mean	16	0.5555	0.3717
Audio Fundamental Frequency	1	0.1521	-0.0170
Harmonic Ratio	1	0.3904	0.0346
Upper Limit Harmonicity	1	0.0424	-0.0275
Audio Power Type	1	0.3461	-0.0263
Audio Spectrum Basis	8	0.5069	-0.1376
Audio Spectrum Projection	9	0.0068	-0.0194
Audio Spectrum Centroid	1	0.3784	0.0247
Audio Spectrum Envelope	34	0.2957	0.0871
Audio Spectrum Spread	1	0.3344	-0.0382
Spectrum Basis	8	0.5069	-0.1376
Audio Wave Form	2	0.0998	-0.0067
Spectral Entropy	1	0.4055	-0.1166
Zero-crossing Rate	1	0.4464	-0.1011
GBFE	311	0.0859	0.0192
SMFCC	10	0.0774	-0.5840

Table 2. Ad And Mmd Values For Each Audio Feature

For the dependency function *D* Pearson Correlation Coefficient (PCC) was used. The efficacy of PCC in data analysis and tecting linear relationships has already been established [25]. PCC measures the strength of association or co-dependency of two objects. Its result lies between -1 (strong negative association) and 1 (strong positive association).

Table 2 shows the dimensions and the AD & MMD values calculated over the audio features. It can be observed that all features have yielded positive AD values, however majority of them are close to 0 which means that the features cannot distinguish between true and false samples. Moreover, features with strong AD values are shown to have a weak MMD.

Figures 2 and 3 show the AD values (only 5 advertisement classes for clear illustration) of Audio Flatness Mean and Zerocrossing Rate. Although both features have similar overall AD values, Zero-Crossing Rate has a low MMD. Considering uniformity in AD values, Audio Flatness Mean is a more suitable feature.

An automatic broadcast monitoring system based on features that have low MMD or AD value would yield a high false positive and false negative result. It can be stated that none of the features (using current implementations and dimension) can be used for a highly accurate, robust and scalable broadcast monitoring system. Table 2 indicates that Audio Flatness Variance and Audio Flatness Mean are the only relatively strong candidates, however much larger AD and MMD values will be required for a concrete automatic broadcast monitoring system.

3.2.4 Feature Subset Selection: The impact of large feature dimensions on accuracy and processing time has been stated in [26], [5], [27], [28]. Features such as Audio Flatness Mean, GBFE, Audio Spectrum Envelope, SMFCC , etc. have large dimensions (as seen in Table 2), which may have impacted their AD and MMD values.

Different feature subset selection algorithms can be found in literature such as SWR [29], [30] and Branch & Bound algorithm [31]. However, According to [32] sequential Forward Floating Selection method (SFFS) is one of the most effective algorithms. In short SFFS starts with a null feature set and for each step, the algorithm adds a dimension to the current set if it satisfies a criterion function. The process continues until each of the feature dimensions are accepted or rejected by the criterion function.

In this paper the criterion function used is a minimum Average Dependency (AD) value. On each step of SFFS, AD is calculated on the next dimension of the audio feature along with the dimensions that have already satisfied the criterion in previous iterations. If the resulting AD is greater than the minimum Average Dependency (AD) value the dimension is included into the selected feature dimension set and minimum Average Dependency (AD) is set to the calculated AD. Initially minimum Average Dependency (AD) of the audio feature is set to its overall AD shown in Table 2. The effect of SFFS can be seen in figures 4 and 5 which shows AD values of GBFE calculated before and after the feature subset selection respectively.

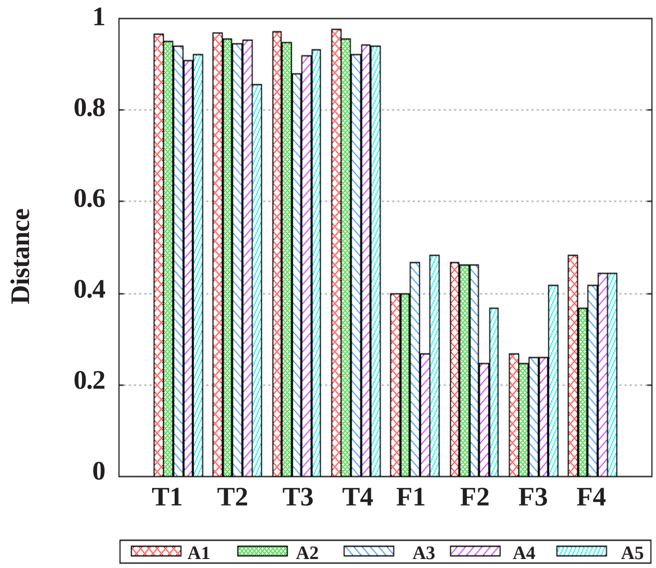

Figure 2. Dependency matrix values for Audio Flatness Mean

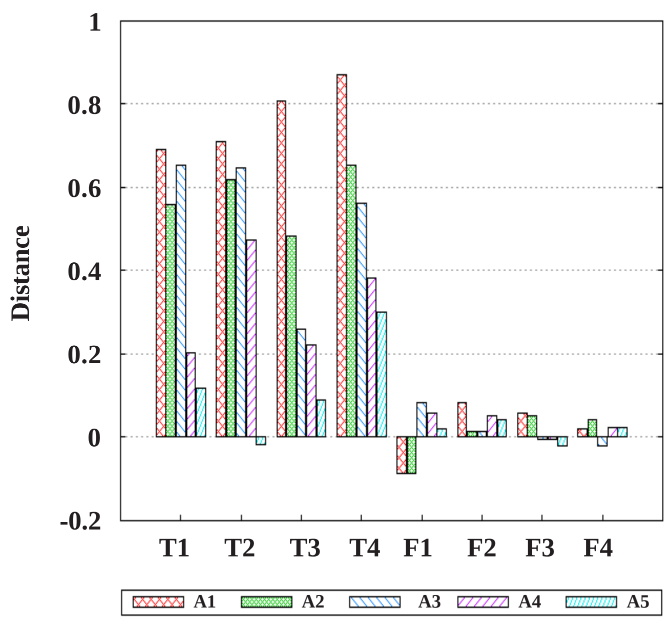

Figure 3. Dependency matrix values for Zero-crossing Rate

Table 3 shows the reduced dimension size and AD & MMD values calculated on the optimum dimension subset of each audio feature. It can be observed that applying feature reduction process has significantly increased the AD of all audio features. The increase in AD has also reflected in substantial rise in MMD values for some of the features. Moreover, the process has reduced the dimension size to such an extent that processing time will be significantly less than if the original dimensions were used.

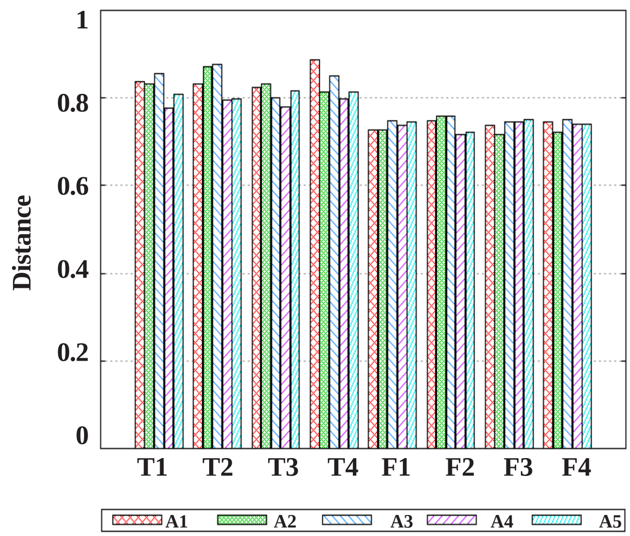

Figure 4. Dependency matrix values for GBFE (Before SFSM)

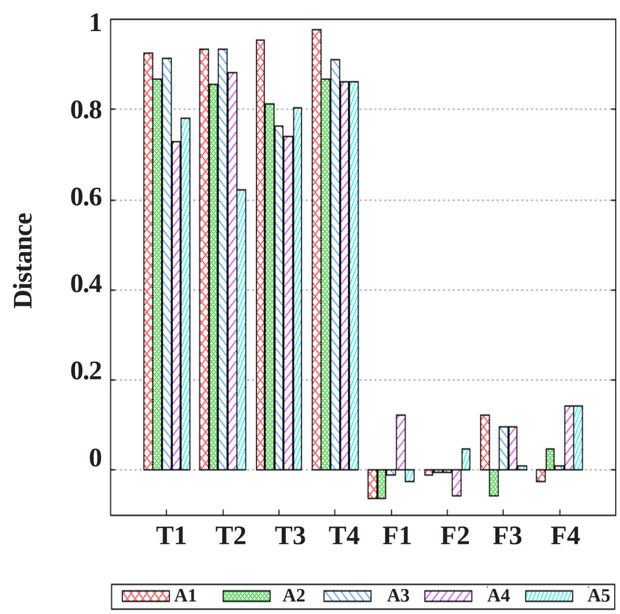

Figure 5. Dependency matrix values for GBFE (After SFSM)

The feature assessment based on Average Dependency and Minimum-Maximum Distance criteria has yielded three potentially strong feature candidates Audio Flatness Mean, GBFE and SMFCC. The advertisement sequence classification experiments that were performed on the three audio features are discussed in the next section.

3.2.5 Feature Vector Subtraction: One of problems that occur due to transmission quality is highlighted in Figure 6 which shows a difference in amplitude in identical advertisements across channels. To overcome this problem coefficient value of an audio feature vector *feat*(n, m) is subtracted by the coefficient value of the corresponding vector *feat*(n —1, m). Where n is the vector number and m is the dimension number. The effects of subtraction can be seen in Figure 7 which shows the waveform of the same advertisements depicted in Figure 6 after subtraction. The advertisements are now close to identical.

Audio Feature	Dim	AD	MMD
Audio Flatness Variance	1	0.7533	-0.0039
Audio Flatness Mean	3	0.7536	0.4102
Audio Spectrum Basis	2	0.8185	0.0153
Audio Spectrum Projection	1	0.5276	0.0467
Audio Spectrum Envelope	1	0.3679	-0.1155
Spectrum Basis	2	0.8184	0.0142
Audio Wave Form	1	0.4101	-0.0121
GBFE	8	0.8236	0.4777
SMFCC	1	0.7403	0.4766

Table 3. AD and MMD Values For Each Audio Feature (After SFSM)

3.3 Advertisement Sequence Classification

In this section, two stage algorithm for advertisement sequence classification is proposed. The algorithm pseudocode is provided in figure 8.

- '*AD*' a sequence of vectors $[V_1, V_2, ...V_n]$ containing features of the advert that must be detected. Each vector represents a single frame which varies in size depending on the feature.
- '*TRANS*' a sequence of vectors $[V_1, V_2, ...V_m]$ containing features of a transmission, in which advert will be detected. Each vector represents a single frame which varies in size depending on the feature.
- '*windowsize*' Cardinality of the array of start frames used in phase 1.
- '*RESULT*' an array containing start position of matched advertisements.

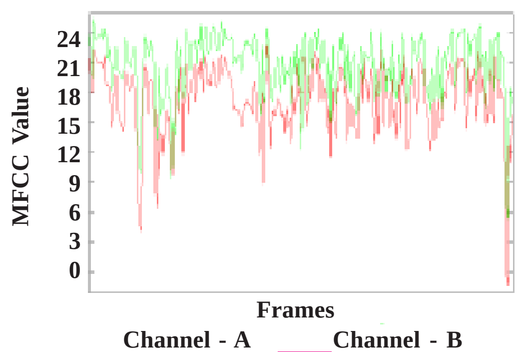

Figure 6. MFCC representation of identical advertisements-

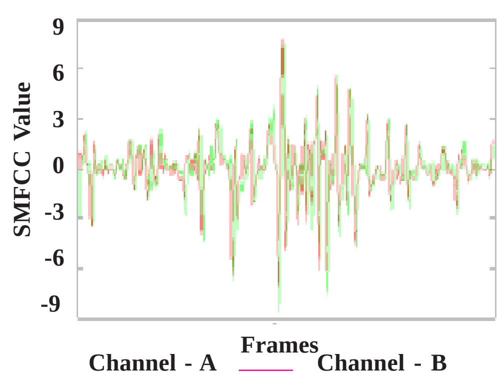

Figure 7. Representation of identical advertisements after subtraction

```
SLIDING-WINDOW (AD, TRANS, windowsize)

ADStart ← copy[AD(1 to windowsize)]
for j ← 1  to length[TRANS] – windowsize
        TransW ← copy[TRANS(j to j+windowsize)]
        res ← Calculate-Frame-PCC(ADStart, TransW)
                if res = TRUE
                        TransC ← copy[TRANS(j to j+Length[AD])]
                        matched ← Calculate-Frame-PCC(AD, TransC)
                        if matched = TRUE
                                add j in RESULT

Calculate-Frame-PCC (AD, TRANS)

pcc ← Pearson-Correlation-Coefficient(AD, TRANS)
        if pcc >= th
                return TRUE
        else
                return FALSE
```

Figure 8. Pseudocode of sliding window algorithm

Figure 9 graphically explains the sequence classification process. In this example source advertisement contains five frames A [0..4] and the transmission contains N frames T[0..N —1]. An occurrence of the source advertisement starts at frame 4 in the transmission. At i = 0, using windowsize = 2, the starting two frames of the source advertisement are matched with the starting two frames of the transmission. As the occurrence starts at frame 4, *Calculate — Frame — PCC*( ) returns *false*, the window then slides on to the next frame in the transmission and the process starts again. Similarly at i = 1 and i = 2, *Calculate — Frame — PCC* () returns *false*. But at i = 3, the function returns true as occurrence of the source advertisement has started. Phase two is initiated and all five frames of source advertisement are matched with five transmission frames, *Calculate — Frame — PCC* () returns *true* and the timings are logged to a database.

4. Results and Discussion

Experiments were conducted on 216 hours of captured transmission from multiple television channels. Recordings were provided by a local monitoring company in a lossy format wma@128kbps distributed over 1296 files of 10 minutes each. A total of 28 advertisements were cropped from the transmissions files which served as the source for advertisement detection. Manual logs of aired advertisements were also provided by the local monitoring company which acted as the ground truth. SMFCC, Audio Flatness Mean and GBFE features of each of the 1296 files were extracted to a database along with the features of cropped source advertisements.

The accuracy rates attained on each advertisement across the three channels by SMFCC, Audio Flatness Mean and GBFE can be seen in Table V. Overall the audio features achieved a recognition rate of 98.99, 97.31 and 99.33 percent respectively. In terms of false positives the features produced overall rates of 0.004, 0.021 and 0.001 respectively. The results show significant improvement in terms of accuracy and false postive rates as can be seen in IV. The accuracies improved for all features except for GBFE where they were comparable but before feature reduction the GBFE feature takes many hours to process the data. After feature reduction the time for processing is reduced to a few minutes.

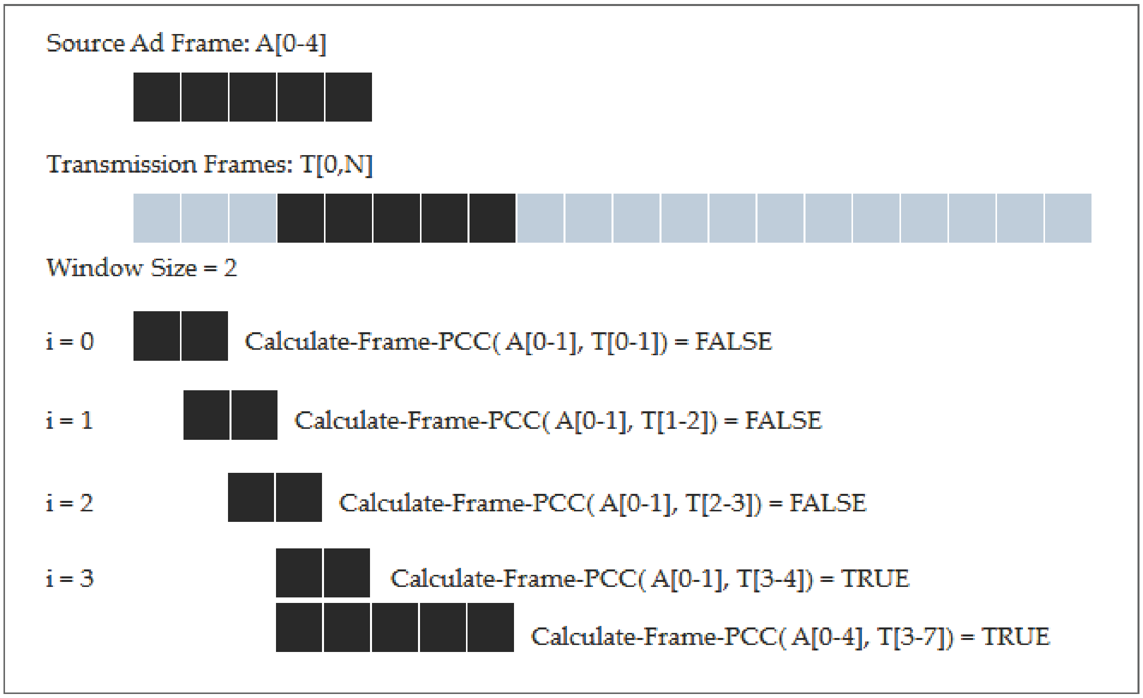

Figure 9. Sliding window algorithm explained

It is observed that the three audio features satisfy the high accuracy and robustness requirements for an automatic broadcast monitoring system. However, since GBFE has a large dimension it requires a large amount of processing time. Hence, the results before feature reduction are not presented. They would be added to an extended version of the paper which would also include a comparison of the processing times for the various features.

It is important to note that after the reduction of features through our SFFS-based algorithm a robust set of features are acquired that provide high classification accuracies over multiple audio channels. This is not the case before feature reduction is carried out. Hence, it is proved that our algorithm provides a robust and efficient set of features for audio advertisement matching and broadcast channel auditing.

	CH1 Accuracy (%)		CH2 Accuracy (%)		CH3 Accuracy (%)	
Audio Feature	True +	False +	True +	False +	True +	False +
SMFCC	65.74	20.4	72.92	25.2	74.3	16.64
Audio Flatness Mean	83.21	15.19	86	13.54	92	10.81

Table 4. Recognition Rates Attained Before Feature Optimization

	CH1 Accuracy (%)		CH2 Accuracy (%)		CH3 Accuracy (%)	
Audio Feature	True +	False +	True +	False +	True +	False +
SMFCC	98.43	0.004	98.54	0.004	100	0
Audio Flatness Mean	95.81	0.015	96.11	0.021	100	0.052
GBFE	99.19	0.002	98.8	0	100	0

Table 5. Recognition Rates Attained After Feature Optimization

5. Conclusion

This paper proposes a method to assess the feasibility of audio features using Average Dependency (AD) and Minimum-Maximum Distance (MMD) criteria. Furthermore, features selection is also carried out to reduce the dimensions thereby increasing robustness and efficiency. Using a proposed algorithm for advertisement sequence classification experiments were conducted, which show that Gabor Filter Bank features (GBFE) provide the highest accuracy rate. However, Melfrequency Cepstral Coefficient (MFCC) and MPEG7 Audio Flatness Mean fair better considering overall aspects of scalability, robustness and accuracy combined.

Currently a full scale monitoring service is being implemented using the technique presented in this paper and results are being generated for more data for various advertisements over various channels. For optimization purposes, possible areas for parallel computation are being identified for low processing times.

References

[1] Media Bank Pakistan, Media monitoring system,.

[2] Bruno, Oliveira., Alexandre, Crivellaro., Roberto, C´esar. M. (2005). Jr, Audiobased radio and tv broadcast monitoring, *In*: Proceedings of the 11th Brazilian Symposium on Multimedia and the web, New York, 2005, ACM.

[3] Antonio Camarena-Ibarrola., Edgar Chvez., Eric Sadit Tellez. (2009).Robust radio broadcast monitoring using a multi-band spectral entropy signature, *In*: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Springer Berlin Heidelberg.

[4] Gauch, J. M., Shivadas, A. (2005). Identification of new commercials using repeated video sequence detection, in Image *In*:Processing,IEEE International Conference on.

[5] Sivagaminathan., Rahul Karthik., Sreeram Ramakrishnan., (2007). A hybrid approach for feature subset selection using neural networks and ant colony optimization, *Expert Syst. Appl.*, 33 (1) 49–60, July.

[6] Jurafsky, D., Martin, J. H. (2000).An introduction to natural language processing, computational linguistics, and speech recognition.

[7] Li, Tao., Ogihara, Mitsunori., Li, Qi (2003). A comparative study on contentbased music genre classification, New York, ACM.

[8] Spevak, C., Polfreman, R. (2001).Sound spotting–a frame-based approach, in International Symposium on Music Information Retrieval, Bloomington, IN, USA.

[9] Pitman, M. C., Fitch, B. G., Abrams, S., Germain, R.S., et al., (2003). System for monitoring broadcast audio content, June 3 , US Patent 6,574,594.

[10] Logan, J. D., Goldhor, R., Goessling, D. et al., (2006). Apparatus and methods for broadcast monitoring, May 30, US Patent 7, 055, 166.

[11] Welsh, R. J. (1989). Video monitoring system, August. 15 , US Patent 4,857,999.

[12] Kalker, Ton., Depovere, Geert., Haitsma, Jaap., Maes, Maurice J.(1999). Video watermarking system for broadcast monitoring.

[13] Hasan, M. R., Jamil, M., Rahman, M. G. R. M. S. (2004). Speaker identification using mel frequency cepstral coefficients, variations.

[14] Milner, B., Shao, X. (2002).Speech reconstruction from mel-frequency cepstral coefficients using a source-filter model, in International Conference on Spoken Language *In*: Processing (ICSLP). Citeseer.

[15] Jang, R. (2006).bAudio signal processing and recognition.

[16] Wei Han, Cheong-Fat Chan, Chiu-Sing Choy, Kong-Pang Pun, R. (2006). An efficient mfcc extraction method in speech recognition, in Circuits and Systems, ISCAS 2006. *In*: Proceedings. IEEE International Symposium on, p. 4 p.–.

[17] Hossan, M. A., Memon, S., Gregory, M. A. (2010). A novel approach for mfcc feature extraction, *In*: Signal Processing and Communication Systems (ICSPCS), 2010 4th International Conference on, p. 1–5.

[18] Zhao, Huan., Zhao, Kai He Liu, Fei Yu, (2012).Improved mfcc feature extraction combining symmetric ica algorithm for robust speech recognition, *Journal of multimedia*, 7 (1) 74–81.

[19] Zbigniew Siciarz, Aquila open source and cross-platform dsp (digital signal processing) library.

[20] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION ISO/IEC JTC1/SC29/WG11, (2002).Coding of moving pictures and audio, mpeg-7 overview.

[21] Crysand, Holger., Tummarello, Giovanni., Piazza, Francesco. (2004). Mpeg- 7 encoding and processing: Mpeg7audioenc+ mpeg7audiodb, in 3*rd* Musicnetwork Open Workshop, Munich.

[22] Ren´e Sch¨adler, Marc., Meyer, Bernd T., Kollmeier, Birger. (2012).Spectrotemporal modulation subspace-spanning filter bank features for robust automatic speech recognition, *The Journal of the Acoustical Society of America*,131, p. 4134,

[23] Ren´e Sch¨adler, Marc., Kollmeier, Birger (2012). Normalization of spectrotemporal gabor filter bank features for improved robust automatic speech recognition systems, *In*: INTERSPEECH.

[24] Lei, H., Meyer, B.T., Mirghafori, N. (2012). Spectro-temporal gabor features for speaker recognition, *In* : Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, p. 4241–4244.

[25] Mueen, Abdullah., Nath, Suman., Liu, Jie (2010).Fast approximate correlation for massive time-series data, *In*: Proceedings of the ACM SIGMOD International Conference on Management of Data, New York, NY, USA, SIGMOD '10, p. 171–182, ACM.

[26] Pourhabibi, Tahereh., Imani, Maryam Bahojb., Haratizadeh, Saman (2011). Feature selection on persian fonts: A comparative analysis on gaa, gesa and ga, Procedia Computer Science, 3, p. 1249–1255.

[27] Mitra, Pabitra., Murthy, CA., Pal, Sankar K. (2002).Unsupervised feature selection using feature similarity, *IEEE Transactions on Pattern analysis and Machine Intelligence*, 24 (3) 301–312, 2002.

[28] Robnik-ikonja, Marko., Kononenko, Igor (2003).Theoretical and empirical analysis of relieff and rrelieff, Machine Learning, 53 (1-2) 23–69.

[29] Colak, S., Isik, C.(2003). Feature subset selection for blood pressure classification using orthogonal forward selection, in Bioengineering Conference, IEEE 29*th* Annual, *In*: Proceedings of, p. 122– 123.

[30] Yang, Jihoon., Honavar, Vasant(1998). Feature subset selection using a genetic algorithm, in Feature Extraction, Construction and Selection, Huan Liu and Hiroshi Motoda, Eds., 453 of The Springer International Series in Engineering and Computer Science, p. 117–136. Springer US.

[31] Pudil, P., Novovicova, J., Kittle R, J. (1994). Floating search methods in feature selection, *Pattern Recognition Letters*, 15 (11) 1119 – 1125.

[32] Jain, A., Zongker, D. (1997).Feature selection: evaluation, application, and small sample performance, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 (2) 153–158.