www.czasopisma.pan.pl PAN www.journals.pan.pl
INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2015, VOL. 61, NO. 2, PP. 185–190
Manuscript received April 13, 2015; revised June, 2015. DOI: 10.1515/eletel-2015-0024

Change Point Determination in Audio Data Using Auditory Features

Tomasz Maka

***Abstract*—The study is aimed to investigate the properties of auditory-based features for audio change point detection process. In the performed analysis, two popular techniques have been used: a metric-based approach and the △BIC scheme. The efficiency of the change point detection process depends on the type and size of the feature space. Therefore, we have compared two auditory-based feature sets (MFCC and GTEAD) in both change point detection schemes. We have proposed a new technique based on multiscale analysis to determine the content change in the audio data. The comparison of the two typical change point detection techniques with two different feature spaces has been performed on the set of acoustical scenes with single change point. As the results show, the accuracy of the detected positions depends on the feature type, feature space dimensionality, detection technique and the type of audio data. In case of the △BIC approach, the better accuracy has been obtained for MFCC feature space in the most cases. However, the change point detection with this feature results in a lower detection ratio in comparison to the GTEAD feature. Using the same criteria as for △BIC, the proposed multiscale metric-based technique has been executed. In such case, the use of the GTEAD feature space has led to better accuracy. We have shown that the proposed multiscale change point detection scheme is competitive to the △BIC scheme with the MFCC feature space.**

***Keywords*—audio change point detection, auditory features, gammatone filter bank**

I. Introduction

Recently, audio and speech-based services play important role in many human-machine interaction systems. Such services may enhance the process of communication which improves the overall user experience. To achieve satisfactory results at the audio analysis stage, the audio stream has to be decomposed into regions with different acoustical structure. In that way, properties of each audio segment may simplify the description of input data and further processing. The process of audio segmentation uses the variability of one or several attributes of the signal. In order to determine segments within audio stream, the whole time-frequency structure of the signal should be determined. In the real situations, the transitions between audio segments can be smooth or may include acoustical events. Carefully configured audio parametrization stage can improve position accuracy of the change points in audio stream. Therefore, the characteristics of the audio feature space and its dimensionality influences on the efficiency of segmentation process. The popular approaches for segmentation of audio data can be grouped into two main categories: metric-based and model-based. The first group includes methods based on the distance measures between neighbouring frames to evaluate acoustic similarity and to determine boundaries of the segments. The second group includes techniques for data models comparison. The number of classes in the audio data and the type of audio task should affect the choice of the segmentation method. For a specified number of audio classes, an approach using classification process for fixed size frames can be applied to determine the segments in audio data.

In the presented study, the analysis of auditory features and two different approaches for audio segmentation have been investigated. In the section II, a short analysis of existing approaches for audio segmentation is described. The types and properties of auditory features are enumerated in section III. Section IV presents two typical approaches to change point detection. Our proposed approach using multiscale frame-to-frame comparison is introduced in section V. The performed experiments and obtained results are described in section VI. Finally, a summary has been provided in the last part of the paper.

II. Related Works

There are many techniques for segmentation of audio data with different approaches and features. This is due to the fact that such process is an essential part of the audio analysis chain. The typical methods are based on the similarity measures of audio frames [1] and the techniques using the comparison of the signal models [2]. An analysis of the onsets found in audio data is the basis of some approaches [3], [4]. In the [5], a segmentation based on an analysis of the self-similarity matrix by computing the inter-frame spectral similarity is presented. The segments are determined by correlating the diagonal of the similarity matrix with a dedicated template. The changes in the obtained signal are possible candidates for change points. Hanna et. al. [6] presented a new audio feature sets defined for four classes of signals: colored, pseudo-periodic, impulsive and sinusoids within noises. It has been shown that using the proposed feature set increases the discriminant power compared to a usual feature set. Ref. [7] describes a system for auditory segmentation based on onsets and offsets of auditory events. The segments are generated by matching the obtained onsets and offsets. An algorithm for audio scene segmentation is presented in [8]. The presented framework is based on multiple feature models and a simple, causal listener model using multiple time-scales. Recently, an approach for generic audio segmentation by classification has been presented by Castan et. al. in [9]. Such approach based on classifying consecutive audio frames, where the segmentation is performed by an analysis of the sequence of decisions. The proposed system is based on the factor analysis to compensate

The author is with the Faculty of Computer Science and Information Technology, West Pomeranian University of Technology, Zolnierska 49, 71-210 Szczecin, Poland (e-mail: tmaka@wi.zut.edu.pl).

the within-class variability and does not require any dedicated features or hierarchical structure.

The analysis of auditory features presented in this work has been aimed at showing its properties in the audio segmentation process. We have decided to examine the effectiveness of the segmentation task using two the most popular methods: metric-based and ΔBIC segmentation schemes. In our previous work [10], the features based on the gammatone filter bank (GTEAD) has been proposed in segmentation stage instead of the popular MFCC features. This is because of its higher variability between frames of signals belonging to different acoustical classes. It has been demonstrated that usage of GTEAD features allows to obtain higher efficacy of change point detection using the ΔBIC segmentation technique. For the same reason, we have performed an analysis of segmentation process using a metric-based approach for both features and we have proposed its extension to the multiscale version.

III. Auditory Features

The feature extraction stage plays an important role in the audio segmentation process [2], [6]. Typically, the feature space used in the segmentation schemes includes the Mel-frequency Cepstral Coefficients (MFCC) [11]. Because the segmentation accuracy is connected with changes in a time-frequency structure of a source signal, the MFCC feature gives satisfactory results [12], [2]. However, in many situations such feature set, including its dynamic properties, results in a low detection ratio. Therefore, based on the results presented in [13] we have designed the GTEAD feature (GammaTone/Envelope/Autocorrelation/Distance) [10].

A. Mel-Frequency Cepstral Coefficients (MFCC)

The MFCC feature [14] is widely used in many speech and audio classification tasks. It represents the power spectrum envelope and is calculated by using a set of filter bank mapped onto the Mel-frequency scale which is linear below 1kHz and logarithmic above 1kHz. There are several variants of MFCC filter banks with various numbers of filters and their amplitudes. An example of popular filter bank with 40 filters, introduced in [15], is depicted in Fig. 1.

The MFCC coefficients are calculated in the following steps:

- the signal is split into frames,
- each frame is transformed into power spectrum,
- a set of triangular filters using the Mel-frequency scale is applied,
- for each filter output a logarithm of energy is calculated,
- finally, the MFCC coefficients are obtained by applying the DCT transform:

$$c_n = \sum_{b=1}^{B} \log(Y_b) \cos\left[\frac{\pi \cdot n \cdot (b-0.5)}{B}\right], \qquad (1)$$

where: B is the number of filters, Y_b – energy at the b-th filter output, and n denotes number of the MFCC coefficient ($B \geq n \geq 1$).

B. Inter-Channel Properties of Gammatone Filter Bank (GTEAD)

The GTEAD feature [10] represents the distances between autocorrelation signals of envelopes calculated from the outputs of the gammatone filter bank.

The gammatone filters represents a model for the impulse response of auditory nerve fibres [16]. The n–th order gammatone filter has the impulse response defined as [17]:

$$g_m(t) = t^{n-1} \cdot e^{-b(f_m)\cdot t} \cdot e^{j\cdot 2\pi \cdot f_m \cdot t}, \qquad (2)$$

where f_m is the filter center frequency, $b(f_m)$ denotes filter bandwidth for frequency f_m, $m = 1, 2, \ldots, M$, and M is the number of channels. The bandwidth $b(f_m)$ of gammatone filter is defined according to the equivalent rectangular bandwidth of the human auditory filter [16]:

$$b(f_m) = 1.019 \cdot (24.7 + 0.108 \cdot f_m), \qquad (3)$$

where the order of the gammatone filters is equal to $n = 4$ and the center frequencies are selected in proportion to their bandwidths. The frequency responses of the selected gammatone filters are shown in Fig. 2. From the signal filtered in each channel of a gammatone filter, its envelope is calculated and periodic self-similarities are computed using the autocorrelation function. The algorithm for the GTEAD feature vector extraction is depicted in Algorithm 1.

IV. Audio Change Point Detection

The change point detection process involves the similarity analysis of the selected parts of a signal in order to determine the position where high difference of the content variability is observed. At the first stage, the audio signal is split into

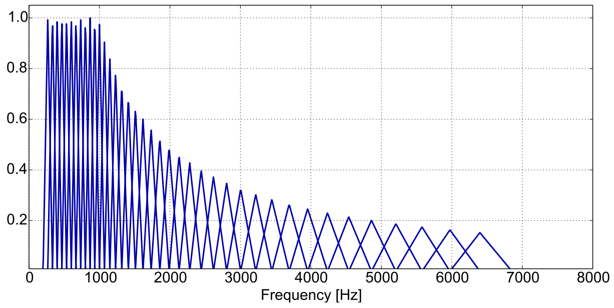


Fig. 1. Filter bank of 40 triangular filters in the Mel-frequency scale [15].

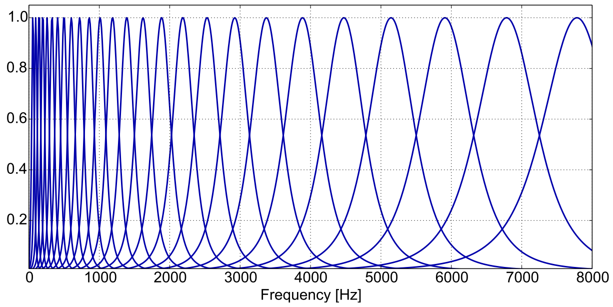


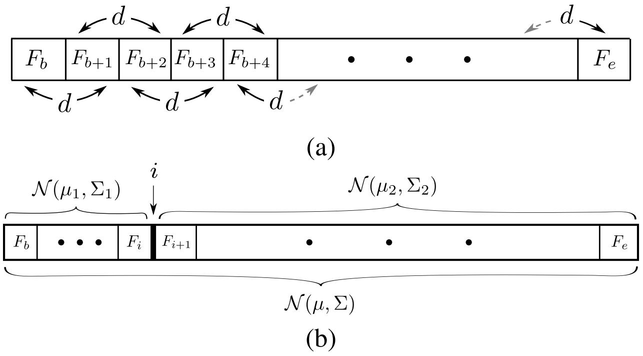
Fig. 2. Frequency responses of selected gammatone filters in 0.1–8kHz band [16].

Algorithm 1: GTEAD feature vector extraction

```
Input: $\mathbf{X} = \{x_i\}_{i=1,\dots,N}$ – input signal, $M$ – number of
       gammatone filters (1–128).
Result: $\mathbf{Z} = \{z_i\}_{i=1,\dots,M-1}$ GTEAD feature vector

for $m \leftarrow 1$ to $M$ do
    • apply $m$-th gammatone filter to $\mathbf{X}$ and generate
      complex output $a_i^{(m)}$,
    • compute envelope $H_i^{(m)}$ of $a_i^{(m)}$:
      $H_i^{(m)} = \sqrt{\Re e^2[a_i^{(m)}] + \Im m^2[a_i^{(m)}]}$,
    • calculate autocorrelation function of $H_i^{(m)}$ for
      $w = 1, \dots, N$:
      $R_w^{(m)} = \sum_{i=1}^{N} H_i^{(m)} \cdot H_{i+w}^{(m)}$.
end
for $i \leftarrow 1$ to $M-1$ do
    $z_i = \sqrt{\sum_{w=1}^{N} \left[R_w^{(i)} - R_w^{(i+1)}\right]^2}$
end
```

frames, then for each frame a D dimensional feature vector F_h is calculated, $h = 1, \dots, H$ where H is the total number of frames. After feature extraction step, a change point detection process is performed. A brief illustration of two typical techniques for such task is presented in Fig. 3.



(a)

(b)

Fig. 3. Audio change point detection techniques: metric-based (a) and ΔBIC (b).

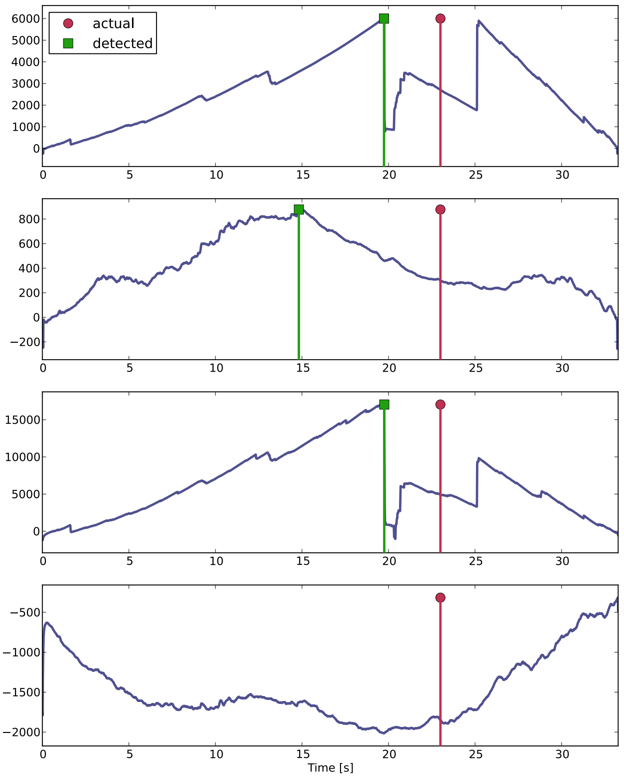


Fig. 4. Examples of ΔBIC trajectories calculated for audio data using (from top to bottom): GTEAD ($D = 4$), MFCC ($D = 4$), GTEAD ($D = 12$) and MFCC ($D = 12$) features.

In the metric-based approach, a distance or divergence function $d(F_p, F_{p+1})$ between adjacent frames is calculated as shown in Fig. 3a. The peaks in the resulting trajectory may represent possible changes in the audio data. The ΔBIC method [2] is based on the comparison of two models – the first where data is modelled by two Gaussians – $\mathcal{N}(\mu_1, \Sigma_1)$ and $\mathcal{N}(\mu_2, \Sigma_2)$, and the second where data is modelled as a single Gaussian – $\mathcal{N}(\mu, \Sigma)$ (see Fig. 3b). The obtained trajectory is computed as the difference between BIC values of these two models (where i is the point in the data $\{F_b, \dots, F_i, \dots, F_e\}$, $b < i < e$):

$$\Delta BIC_i = N_1^{(i)} \log\left|\Sigma_1^{(i)}\right| - N_2^{(i)} \log\left|\Sigma_2^{(i)}\right| - N \log|\Sigma| - \frac{1}{4}\left[(D^2 + 3D) \cdot \log(N)\right], \tag{4}$$

where: N – is the total length of analysed data window $\{F_b, \dots, F_e\}$, $N_1^{(i)}$ – the size of left-side window $\{F_b, \dots, F_i\}$; $N_2^{(i)}$ – the size of right-side window $\{F_{i+1}, \dots, F_e\}$, $i \in [b, e]$; $\left|\Sigma_1^{(i)}\right|$, $\left|\Sigma_2^{(i)}\right|$ and $|\Sigma|$ are the determinants of the covariance matrices for the left-side / right-side / whole window and D is the dimension of the feature space. The change in the audio stream at position i ($\arg\max_i(\Delta\mathrm{BIC}_i)$) occurs when $\max_i(\Delta\mathrm{BIC}_i) > 0$.

The MFCC and GTEAD features have been compared using several audio signals with a single change point. Some examples of ΔBIC trajectories are depicted in Fig. 4. From this figure it follows that the obtained change points have been detected at different positions. In case of MFCC for $D = 12$ the change point has not been detected (Fig. 4, bottom panel). More results are presented in section VI.

V. Multiscale Metric-based Change Point Detection

Due to the low detection ratio of the MFCC feature space and the lower accuracy of GTEAD (see Tab. II), we have decided to design a new technique using a multiscale metric-based approach. In such scheme, a signal is decomposed in the

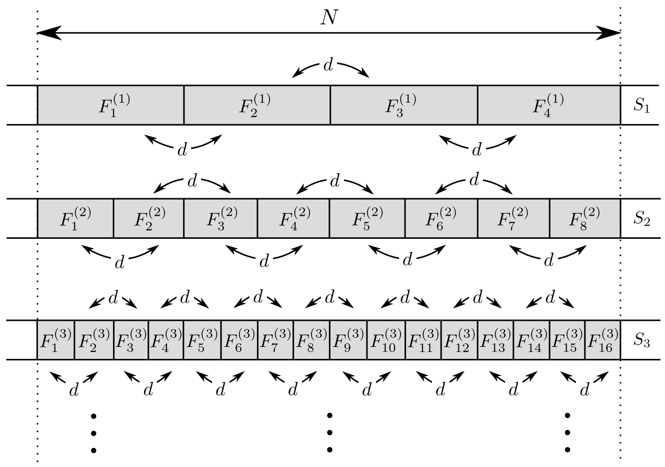

N
d
$F_1^{(1)}$ $F_2^{(1)}$ $F_3^{(1)}$ $F_4^{(1)}$ S_1
$F_1^{(2)}$ $F_2^{(2)}$ $F_3^{(2)}$ $F_4^{(2)}$ $F_5^{(2)}$ $F_6^{(2)}$ $F_7^{(2)}$ $F_8^{(2)}$ S_2
$F_1^{(3)}$ $F_2^{(3)}$ $F_3^{(3)}$ $F_4^{(3)}$ $F_5^{(3)}$ $F_6^{(3)}$ $F_7^{(3)}$ $F_8^{(3)}$ $F_9^{(3)}$ $F_{10}^{(3)}$ $F_{11}^{(3)}$ $F_{12}^{(3)}$ $F_{13}^{(3)}$ $F_{14}^{(3)}$ $F_{15}^{(3)}$ $F_{16}^{(3)}$ S_3

Fig. 5. Illustration of multiscale signal decomposition for change point trajectory generation.

same way as in the metric-based approach. At the next stage, the frame size is decreased and the process is repeated until the number of defined levels (M) is reached. The scheme is illustrated in the Fig. 5.

The accuracy of such decomposition depends on the number of levels (M) and the size of input signal (N). For example, the signals calculated for consecutive levels of audio data with length $N = 10$s and decomposition levels $M = 6$ are presented in Fig. 6 (the actual change point occurred for offset equal to about 50%). The bottom panel shows the signal being a sum of the signals from all levels which is used as a trajectory for change point detection. In this way, applying various fusion schemes (peaks tracking, weighted sum, etc.) between signals of all scales, a spurious peaks in the final trajectory can be reduced. The algorithm for multiscale metric-based trajectory generation is depicted in Algorithm 2, where Euclidean distance has been exploited as a metric.

VI. Experiments

To illustrate the properties of both change point detection methods and feature spaces we have performed several tests using database of audio scene recordings. All signals have a single change point and have been recorded in real conditions. The database contains 14 mono signals recorded at 22.05kHz sampling rate as shown in Tab. I. The feature vectors used in the parametrization stage for ΔBIC scheme have been calculated with 30ms frame size and 50% frame-to-frame overlapping. In the first experiment, an analysis of feature spaces in the ΔBIC change point detection has been performed. During the experiment, each trajectory has been generated with an increasing size of the feature space dimensionality $D = 1, \dots, 12$. As a quality factor we have used the absolute difference $\Phi = |t_d - t_a|$, where t_d – denotes the offset of the detected change point and t_a is a position of the actual change point. The results of the change point detection are shown in Tab. II. As it can be noted, for all test signals a better accuracy, has been obtained for the MFCC feature in most cases. Despite the lower accuracy, all change points have been detected using the GTEAD feature space.

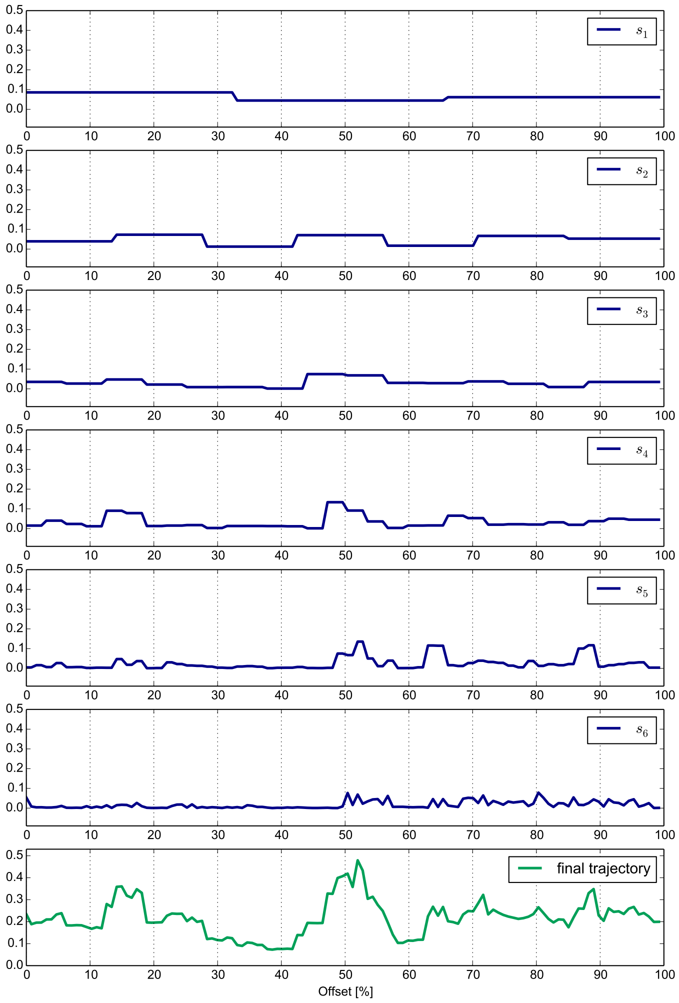
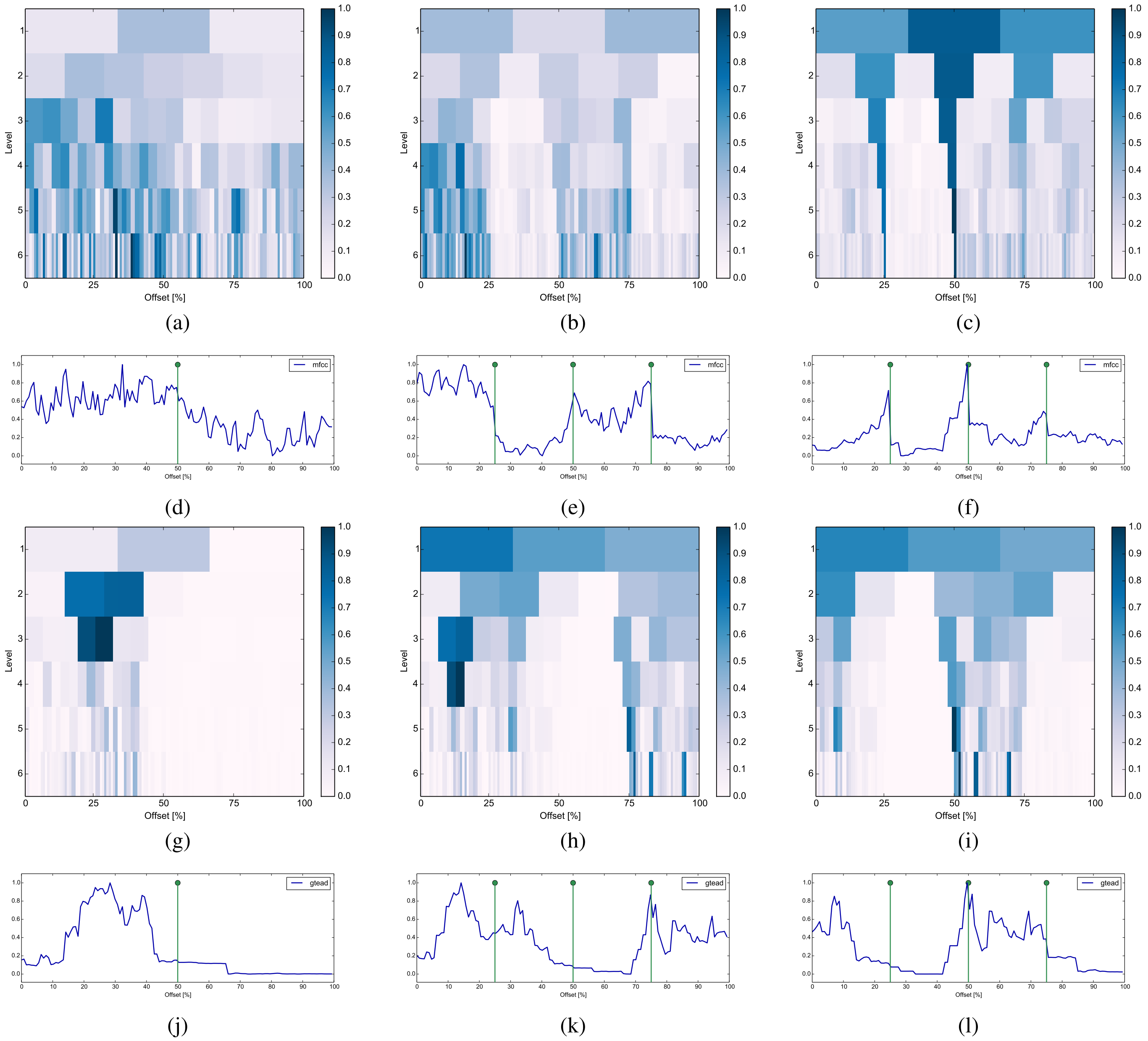


Fig. 6. Example signals obtained for subsequent six scales (calculated for 6th dimensional GTEAD feature space) and the final trajectory calculated as the sum of all six components (bottom).

The second experiment involves the proposed multiscale metric-based change point detection scheme. We have used the same criterion as in case of the ΔBIC method. This is possible since each signal includes a single change point. In real conditions the metric-based approach requires the thresholding stage to detect the peaks in the trajectory which can be candidates for the change points. In Tab. III the results are depicted. In most cases a better accuracy has been obtained for the GTEAD feature space. The performed analysis shows that both features have a discrimination power for the audio change point detection.

VII. Summary

An analysis of auditory features for the change point detection in audio data has been presented. Using two types of features, we have performed change point detection tests for a unique set of audio scenes, where each recording contained a single change point. In the change point detection process we have employed the popular approach called ΔBIC, but due



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Fig. 7. Examples of change point trajectories for three manually prepared signals: male speech / female speech (a,d,g,j); male speech / music / female speech / music (b,e,h,k); music 1 / background sound 1 / music 2 / background sound 2 (c,f,i,l). The multiscale representations have been generated using 6th dimensional MFCC (a,b,c) and GTEAD (g,h,i) feature spaces.

to the computational cost of this technique, we have proposed an approach which is based on frame-to-frame comparison. In the multiscale metric-based technique, the discrimination trajectory is calculated by summing up the feature contours obtained for different time scales.

Using two types of auditory features and set of signals with single change point, we have performed experiments to compare both techniques. In the result, a better accuracy has been obtained for the MFCC feature space in the most cases using ΔBIC approach. However, in the case of multiscale metric-based change point detection, the GTEAD feature outperforms the MFCC. The important fact to note is that in ΔBIC all change points have been detected for GTEAD feature. The obtained detection ratio for MFCC has been equal to about 64%. These results suggest that both techniques and features should be used together to achieve better accuracy and detection ratio. As the future work, we plan to investigate properties of different audio classes and mixed sets of auditory features. Such analysis will be used to find a configuration of the segmentation stage for a specific audio analysis task.

REFERENCES

[1] T. Kemp and M. Schmidt and M. Westphal and A. Waibel, *Strategies for automatic segmentation of audio data*, In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP '00, 5-9 June, Istanbul, 2000, DOI: 10.1109/ICASSP.2000.861862.

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

```
Algorithm 2: Metric-based, multiscale change point trajectory generation

Input: $\mathbf{X} = \{x_i\}_{i=1,\dots,N}$ – input signal, $K$ – number of
        levels ($K \geq 2$), $D$ – feature space size ($D \geq 1$)
Result: $\mathbf{R} = \{r_i\}_{i=1,\dots,2^K}$ – final trajectory

$\mathbf{R} = \{r_i = 0\}_{i=1,\dots,2^K}$
for $j \leftarrow 2$ to $K$ do
    $H = N \ / \ 2^j$
    for $n \leftarrow 1$ to $2^j - 1$ do
        • $c_1 = (N \cdot n - N) \ / \ 2^j$
        • $\mathbf{F}^{(j)}_{c_1} = \{x_i\}_{i=c_1,\dots,c_1+H}$
        • $c_2 = [N \cdot (n+1) - N] \ / \ 2^j$
        • $\mathbf{F}^{(j)}_{c_2} = \{x_i\}_{i=c_2,\dots,c_2+H}$
        • calculate feature vectors $\mathbf{A}_D$, $\mathbf{B}_D$
          of $\mathbf{F}^{(j)}_{c_1}$ and $\mathbf{F}^{(j)}_{c_2}$
        • update $\mathbf{R}$ vector by adding Euclidean distance
          $d(\mathbf{A}_D, \mathbf{B}_D)$ between feature vectors:
          for $p \leftarrow 1$ to $2^{K-j}$ do
              $\alpha = r_{(n-1) \cdot 2^{K-j}+p}$
              $r_{(n-1) \cdot 2^{K-j}+p} = \alpha + d(\mathbf{A}_D, \mathbf{B}_D)$
          end
    end
end
```

TABLE I
AUDIO DATA CHARACTERISTICS USED IN EXPERIMENTS

Signal	Length [s]	Change point position [s]	Change point offset [%]
1	33.37	17.2377	51.66
2	32.7802	17.0504	52.01
3	30.4599	11.0441	36.26
4	18.7131	6.7179	35.9
5	53.7314	31.3799	58.4
6	21.6805	9.8064	45.23
7	35.2395	21.9247	62.22
8	38.1776	21.4294	56.13
9	46.876	29.1244	62.13
10	29.0067	15.9186	54.88
11	33.3098	15.1876	45.59
12	50.9766	31.0212	60.85
13	28.9865	15.155	50.54
14	26.9167	15.4508	57.40

TABLE II
CHANGE POINT DETECTION ACCURACY FOR MFCC AND GTEAD FEATURES USED IN THE ΔBIC APPROACH

Signal	MFCC Detected points	MFCC Best accuracy D	MFCC Best accuracy Φ [s]	GTEAD Detected points	GTEAD Best accuracy D	GTEAD Best accuracy Φ [s]
1	7 / 12	2	0.1827	12 / 12	1	1.1727
2	6 / 12	1	0.2654	12 / 12	1	6.5054
3	6 / 12	1	1.6841	12 / 12	1	0.0791
4	6 / 12	1	0.0921	12 / 12	8	3.5121
5	12 / 12	4	3.6149	12 / 12	2	9.3901
6	6 / 12	5	0.4086	12 / 12	1	2.4486
7	6 / 12	1	0.4447	12 / 12	1	15.9997
8	12 / 12	1	1.5844	12 / 12	8	0.7294
9	6 / 12	1	12.4144	12 / 12	10	0.3544
10	10 / 12	1	0.0186	12 / 12	12	2.2236
11	6 / 12	2	0.2926	12 / 12	12	4.4474
12	8 / 12	2	0.1788	12 / 12	1	5.5488
13	12 / 12	11	0.11	12 / 12	4	0.97
14	6 / 12	2	2.0408	12 / 12	1	1.2442

TABLE III
CHANGE POINT DETECTION ACCURACY FOR MFCC AND GTEAD FEATURES USED IN MULTISCALE, METRIC-BASED APPROACH

Signal	Best accuracy (MFCC) D	Best accuracy (MFCC) Φ [s]	Best accuracy (GTEAD) D	Best accuracy (GTEAD) Φ [s]
1	5	4.8882	3	2.469
2	7	9.0489	1	1.2755
3	1	8.8855	1	4.0659
4	10	0.6767	1	3.4490
5	1	19.1106	2	2.8897
6	2	0.2656	1	0.0758
7	1	20.2598	3	8.0428
8	1	6.2268	1	0.9878
9	1	0.7035	1	21.7424
10	1	5.4123	3	4.2703
11	1	6.0077	3	2.3358
12	1	7.5123	3	8.9447
13	1	9.7232	12	1.7348
14	3	1.4626	1	1.5046

[3] K. West, and S. Cox, *Finding an Optimal Segmentation for Audio Genre Classification*, in Proceedings of 6th International Conference on Music Information Retrieval ISMIR'2005, 11-15 September, London, UK, 2005.

[4] G. Hu and D. Wang, *Auditory Segmentation Based on Onset and Offset Analysis*, IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, no. 2, pp. 396-405, February, 2007, DOI: 10.1109/TASL.2006.881700.

[5] J. Foote, *Automatic audio segmentation using a measure of audio novelty*, Multimedia and Expo –ICME 2000, IEEE International Conference, New York, NY, USA, 2000, DOI: 10.1109/ICME.2000.869637.

[6] P. Hanna and N. Louis and M. Desainte-Catherine, and J. Benois-Pineau, *Audio features for noisy sound segmentation*, International Society for Music Information Retrieval Conference – ISMIR'2004, Barcelona, Spain, October 10–14 2004, vol. 1, pp. 120–124.

[7] G. Hu and D. Wang, *Auditory segmentation based on event detection*, Workshop on Statistical and Perceptual Audio Processing – SAPA'2004, Jeju, Korea, October 2004.

[8] H. Sundaram and S. Chang, *Audio scene segmentation using multiple features, models and time scales*, IEEE International Conference on Acoustics, Speech, and Signal Processing – ICASSP '2000, June 2000, vol. 6, pp. 2441–2444, DOI: 10.1109/ICASSP.2000.859335.

[9] D. Castan, A. Ortega, A. Miguel and E. Lleida, *Audio segmentation-by-classification approach based on factor analysis in broadcast news domain*, EURASIP Journal on Audio, Speech, and Music Processing, vol. 34, pp. 1–13, 2014, DOI: 10.1186/s13636-014-0034-5.

[10] T. Maka, *An Auditory-Based Scene Change Detection in Audio Data*, International Conference on Signals and Electronic Systems (ICSES), 11-13 September 2014, Poznan, Poland, 2014, DOI: 10.1109/IC-SES.2014.6948723.

[11] L. Rabiner and W. Schafer, *Theory and Applications of Digital Speech Processing*, Prentice-Hall, 1st edition, 2010.

[12] T. Nwe, M. Dong, S. Khine, and H. Li, *Multi-Speaker Meeting Audio Segmentation*, in Proceedings of INTERSPEECH'2008, 22-26 September, Brisbane, Australia, 2008.

[13] T. Maka, *Auditory Features Analysis for BIC-based Audio Segmentation*, SIGMAP 2014 – 11th International Conference on Signal Processing and Multimedia Applications, August 27-30, Vienna, Austria, 2014.

[14] S. Davis and P. Mermelstein, *Comparison of parametric representation for monosyllabic word recognition in continuously spoken sentences*, IEEE Transactions on ASSP, August, 1980.

[15] M. Slaney, *Auditory Toolbox*, Apple Technical Report #45, 1998.

[16] D. Wang and G. Brown, *Computational Auditory Scene Analysis*, John Wiley & Sons, Inc., 2006.

[17] M. Cooke, *Modelling Auditory Processing and Organisation*, Cambridge University Press, 2005.