Chapter 8
Emotion Tracking of Radio Station Broadcasts

8.1 Introduction

The overwhelming number of media outlets is constantly growing. This also applies to radio stations available on the Internet, over satellite and air. On the one hand, the number of opportunities to listen to various radio shows has grown, but on the other, choosing the right station has become more difficult. Music information retrieval helps those people who listen to the radio mainly for the music. This technology is able to make a general detection of the genre, artist, and even emotion.

Listening to music is particularly emotional. People need a variety of emotions, and music is perfectly suited to provide them. Listening to a radio station throughout the day, whether we want it or not, we are affected by the transmitted emotional content. In this paper, we focus on emotional analysis of the music presented by radio stations. During the course of a radio broadcast, these emotions can take on a variety of shades, change several times with varying intensity. This paper presents a method of tracking changing emotions during the course of a radio broadcast. The collected data allowed to determine the dominant emotion in the radio broadcast and construct maps visualizing the distribution of emotions over time.

There are studies focused on facilitating radio station selection from the overwhelming number of radio stations. A method for profiling radio stations was described by Lidy and Rauber [58], who used a technique of Self-organizing Maps to organize the program coverage of radio stations on a two-dimensional map. This approach allows profiling the complete program of a radio station.

A study that combines emotion detection and facilitating radio station selection was presented by Rizk et al. [86], who presented a mobile application that streams music from online radio stations after identifying the user's emotions. The songs from online radio stations were classified into emotion classes based on audio features. The application captured images of the user's face using a smartphone camera and classified them into one of three emotions using a classifier on facial geometric distances and wrinkles.

J. Grekow, *From Content-Based Music Emotion Recognition to Emotion Maps of Musical Pieces*, Studies in Computational Intelligence 747,
https://doi.org/10.1007/978-3-319-70609-2_8

The issue of emotion tracking is not only limited to music. The paper by Mohammad [68] is an interesting extension of the topic; the author investigated the development of emotions in literary texts. Yeh et al. [120] tracked the continuous changes of emotional expressions in Mandarin speech.

8.2 System Construction

The proposed system for tracking emotions in radio station broadcasts is shown in Fig. 8.1. It is composed of collected audio data, a segmentation module, a feature extraction module, classifiers, and a result presentation module.

The recorded radio station broadcasts undergo segmentation, and the obtained fragments are then analyzed in the feature extraction module. The example represented by vectors composed of extracted features then undergo classification by a music/speech classifier. Fragments containing music are additionally classified in terms of emotions. In the last phase, the results are analyzed and visualized in the result presentation module. The music speech classifier and emotion classifier used in this process are trained using features obtained from audio samples.

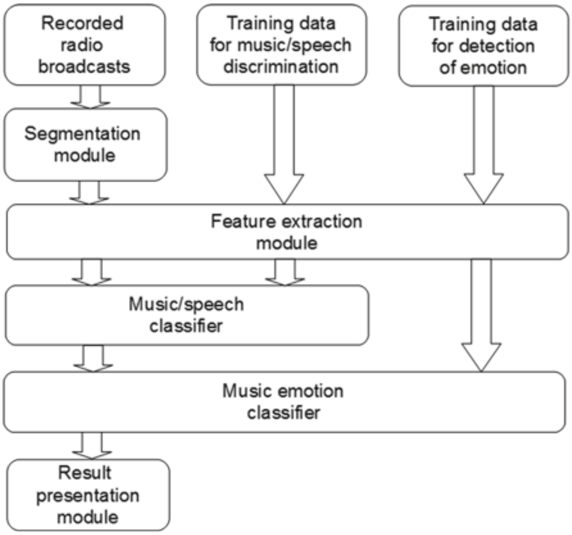
Recorded radio broadcasts
Training data for music/speech discrimination
Training data for detection of emotion
Segmentation module
Feature extraction module
Music/speech classifier
Music emotion classifier
Result presentation module

Fig. 8.1 System construction of emotion tracking in radio station broadcasts

8.3 Music Data

8.3.1 Training Data

To conduct the study of emotion detection of radio stations, we prepared two sets of training data:

1. Data set used for music/speech discrimination;
2. Data set used for the detection of emotion in music.

The set of training data for music/speech discrimination consisted of 128 files, including 64 designated as speech and 64 marked as music. The tracks were all 22050 Hz mono 16-bit audio files in .wav format. The training data were taken from the generally accessible data collection for the purposes of music/speech discrimination from MARSYAS[1] project.

The training data set for emotion detection consisted of 324 six-second fragments of different genres of music: classical, jazz, blues, country, disco, hip-hop, metal, pop, reggae, and rock. The tracks were all 22050 Hz mono 16-bit audio files in .wav format. The data set has been described in detail in Chap. 3 Sect. 3.3. Data annotation was done by five music experts with a university music education. The annotation process of music files with emotion classes has been described in Chap. 3 Sect. 3.3.

In this research, we use four emotion classes corresponding to the four quarters of Russell's model: happy, angry, sad, and relaxed. The amount of examples in the training data set for emotion detection labeled by emotions are presented in Table 8.1.

8.3.2 Recorded Radio Broadcasts

To study changes in emotions, we used recorded broadcasts from 4 selected European radio stations:

- Polish Radio Dwojka (Classical/Culture), recorded on 4.01.2014;
- Polish Radio Trojka (Pop/Rock), recorded on 2.01.2014;
- BBC Radio 3 (Classical), recorded on 25.12.2013;
- ORF OE1 (Information/Culture), recorded on 12.01.2014.

For each station, we recorded 10 h beginning at 10 A.M. and converted the recordings into 22050 Hz mono 16-bit audio files in .wav format. The recorded broadcasts were segmented into 6-second fragments, for example, we obtained 6000 segments from one 10 h broadcast.

[1] http://marsyas.info/downloads/datasets.html.

Table 8.1 Amount of examples labeled by emotions

Basic emotion	Emotion abbreviation	Amount of examples
Happy	e1	93
Angry	e2	70
Sad	e3	80
Relaxed	e4	81

8.4 Feature Extraction

For feature extraction for music/speech discrimination, we used the Marsyas framework for audio processing [106], which has been described in detail in Chap. 6 Sect. 6.2.2. For feature extraction for emotion detection, we used the Essentia extractors [8], which have been described in Chap. 6 Sect. 6.2.1.

Essentia has a much richer feature set than Marsyas and is better suited for feature extraction for emotion detection. Marsyas, with its modest feature set, is enough for good music/speech discrimination.

For each 6-second file from the training data, we obtained a representative single feature vector. The obtained vectors were then used for building classifiers and for predicting new instances.

8.5 Construction of Classifiers

8.5.1 *Music/Speech Classifier*

We built two classifiers, one for music/speech discrimination and the second for emotion detection. During the construction of the classifier for music/speech discrimination, we tested the following algorithms: J48, RandomForest, IBk (K-nn), BayesNet, SMO (SVM). The classification results were calculated using a cross validation evaluation CV-10.

The best accuracy (98%) was achieved using SMO algorithm, which is an implementation of support vector machines (SVM) algorithm (Table 8.2). The confusion matrix for the best music/speech classifier obtained for SMO algorithm is presented in Table 8.3.

Table 8.2 Accuracy and F-measure obtained for tested algorithms

	J48	RandomForest	BayesNet	IBk	SMO
Accuracy (%)	89.84	96.09	95.31	96.09	**98.44**
F-measure	0.89	0.96	0.95	0.96	**0.98**

Table 8.3 Confusion matrix for music/speech classifier obtained for SMO algorithm

		Predicted class	
		Music	Speech
Actual class	Music	**63**	1
	Speech	1	**63**

8.5.2 *Classifier for Emotion Detection*

We built classifiers for emotion detection using the following algorithms: J48, RandomForest, BayesNet, IBk (K-nn), SMO (SVM). The classification results were calculated using a cross validation evaluation CV-10.

For emotion detection, we used four binary classifiers dedicated to each emotion. The process of building the binary classifiers for emotion detection has been presented in Chap. 7 Sect. 7.5.2. The best classifier accuracy was obtained for emotion e2 (87.65%), but for e1 and e4 the results were also high (87.04%). The lowest classifier accuracy was obtained for emotion e3 (82.71%).

From the data obtained during classifier construction, we can clearly see that music/speech discrimination in audio recordings is a much easier task (98% accuracy) than emotion detection (accuracy from 82 to 87%). The reason behind this is that the audio feature set that can discriminate music from speech is particularly comprehensive. In the case of emotion detection, the feature set is not yet so ideal.

8.6 Results of Emotion Tracking of Radio Stations

During the analysis of the recorded radio broadcasts, we conducted a two-phase classification. The recorded radio program was divided into 6-second segments. For each segment, we extracted a feature vector, which was first used to detect if the given segment is speech or music. If the current segment was music, then we used a second classifier to predict what type of emotion it contained. For feature extraction, file segmentation, use of classifiers to predict new instances, and visualization of results, we wrote a Java application that connected different software products: Marsyas, Essentia, MATLAB and WEKA package.

The percentages of speech, music, and emotion in music obtained during the segment classification of 10-hour broadcasts of four radio stations are presented in Table 8.4. On the basis of these results, radio stations can be compared in two ways: the first is to compare the amount of music and speech in the radio broadcasts, and the second is to compare the occurrence of individual emotions.

Table 8.4 Percentage of speech, music, and emotion in music in 10-hour broadcasts of four radio stations

	PR Dwojka (%)	PR Trojka (%)	BBC Radio 3 (%)	ORF OE1 (%)
Speech	59.37	73.35	32.25	69.10
Music	40.63	26.65	67.75	30.90
e1	4.78	4.35	2.43	2.48
e2	5.35	14.43	1.00	0.92
e3	20.27	6.02	56.19	22.53
e4	10.23	1.85	8.13	4.97
e1 in music	11.76	16.32	3.58	8.02
e2 in music	13.16	54.14	1.47	2.98
e3 in music	49.89	22.59	82.93	72.91
e4 in music	25.17	6.94	12.00	16.08

8.6.1 Comparison of Radio Stations

The dominant station in the amount of music presented was BBC Radio 3 (67.75%). We noted a similar ratio of speech to music in the broadcasts of PR Trojka and ORF OE1, in both of which speech dominated (73.35% and 69.10%, respectively). A more balanced amount of speech and music was noted on PR Dwojka (59.37% and 40.63%, respectively).

Comparing the content of emotions, we can see that PR Trojka clearly differs from the other radio stations, because the dominant emotion is e2 energetic-negative (54.14%) and e4 calm-positive occurs the least often (6.94%).

We noted a clear similarity between BBC Radio 3 and ORF OE1, where the dominant emotion was e3 calm-negative (82.93% and 72.91%, respectively). Also, the proportions of the other emotions (e1, e2, e4) were similar for these stations. We could say that emotionally these stations are similar, except that considering the speech to music ratio, BBC Radio 3 had much more music.

The dominant emotion for PR Dwojka was e3, which is somewhat similar to BBC Radio 3 and ORF OE1. Compared to the other stations, PR Dwojka had the most (25.17%) e4 calm-positive music.

8.6.2 Emotion Maps of Radio Station Broadcasts

The figures (Figs. 8.2, 8.3, 8.4 and 8.5) present speech and emotion maps for each radio broadcast. Each point on the map is the value obtained from the classification of a 6-second segment. These show which emotions occurred at given hours of the broadcasts.

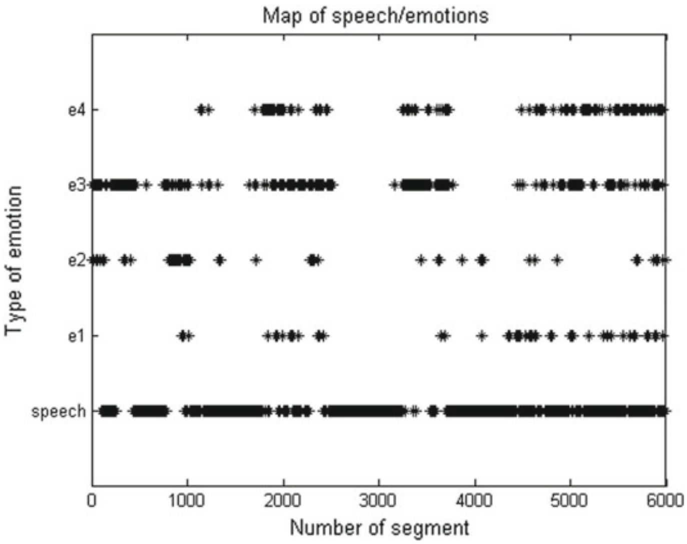

Fig. 8.2 Map of speech and music emotion in PR Dwojka 10 h broadcast

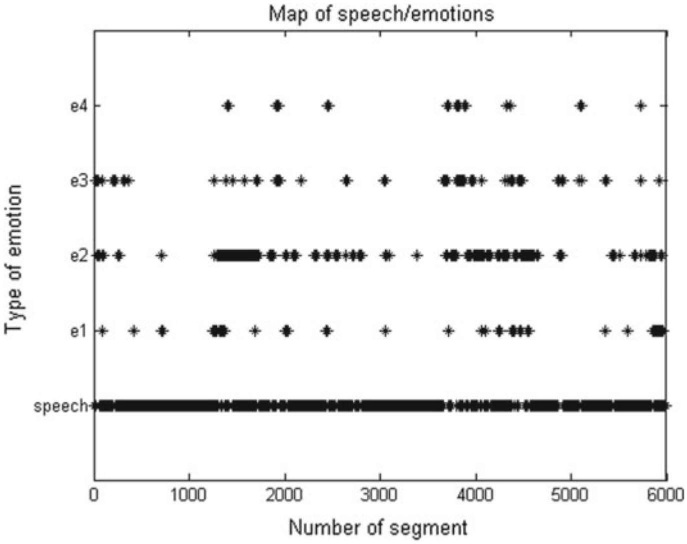

Fig. 8.3 Map of speech and music emotion in PR Trojka 10 h broadcast

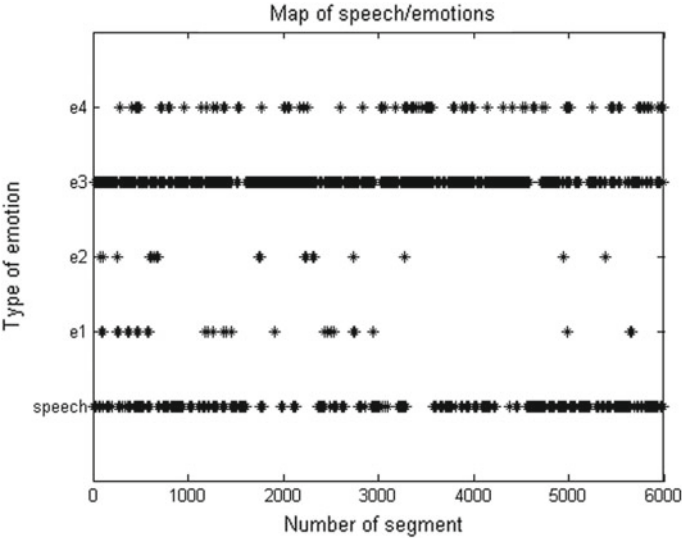

Fig. 8.4 Map of speech and music emotion in BBC Radio 3 10 h broadcast

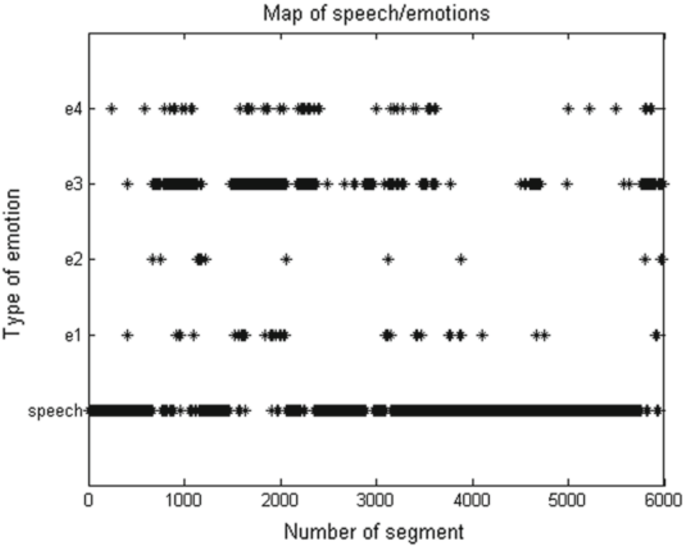

Fig. 8.5 Map of speech and music emotion in ORF OE1 10 h broadcast

For PR Dwojka (Fig. 8.2), there are clear musical segments (1500–2500, 2300–3900) during which e3 dominated. At the end of the day (4500–6000), emotion e2 occurs sporadically. It is interesting that e1 and e4 (from right half of Russell's model) did not occur in the morning. For PR Trojka (Fig. 8.3), emotion e4 did not occur in the morning, and e2 and e3 dominated (segments 1200–2800 and 3700–6000). For BBC Radio 3 (Fig. 8.4), we observed almost a complete lack of energetic emotions (e1 and e2) in the afternoon (segments after 3200). For ORF OE1 (Fig. 8.5), e3 dominated up to segment 3600, and then broadcasts without music dominated. The presented analyses of maps of emotions could be developed by examining the quantity of changes of emotions or the distribution of daily emotions.

8.7 Conclusions

This chapter presented an example of a system for the analysis of emotions contained within radio broadcasts. The collected data allowed to determine the dominant emotion in the radio broadcast and present the amount of speech and music. The obtained results provide a new interesting view of the emotional content of radio stations.

A system for the analysis of emotions contained within radio broadcasts could be a helpful tool for people planning radio programs enabling them to consciously plan the emotional distribution in the broadcast music. Another example of applying this system could be an additional tool for radio station searching.