TELKOMNIKA, Vol.13, No.2, June 2015, pp. 571 ~ 577
ISSN: 1693-6930, accredited A by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v13i2.1124

Integrated System Design for Broadcast Program Infringement Detection

Sukmawati Nur Endah*[1], Satriyo Adhy[2], Sutikno[3]
Informatics Department, Faculty of Science and Mathematics, Universitas Diponegoro
*Corresponding author, e-mail: sukma_ne@undip.ac.id[1]; satriyo@undip.ac.id[2]; tik@undip.ac.id[3]

Abstract

Supervision of television and radio broadcast programs by the Komisi Penyiaran Indonesia Jawa Tengah is still being carried out manually i.e. direct supervision by humans. The process certainly has some weaknesses related to human error, such as tiredness and weary eyes. Therefore, we need intelligent software that could automatically detect broadcast infringement. Up to now, research in this area has not been carried out. This research was aimed at designing an integrated system to detect broadcast infringement, including architecture design and main module interface design. Two main stages in this system are Indonesian-language speech recognition and detection of infringements of the broadcast program. With the method of Mel Frequency Cepstral Coefficients (MFCC) and Hidden Markov Model (HMM), a speech recognition application that uses 1050 sample data produces about 70% accuracy rate. This research will continue to implement the design created using speech recognition applications that have been developed.

Keywords: *Broadcast Infringement, Infringement Detection, Komisi Penyiaran Indonesia (KPI), Speech Recognition*

1. Introduction

The public's great interest in watching television or listening to the radio prompted the government to pass the Broadcasting Act No 32 of 2002 in order that the broadcast media could perform the role of a healthy public information service. One of its provisions was to set up the formation of the Komisi Penyiaran Indonesia (KPI) with authority to oversee the implementation of the rules and code of conduct as well as the broadcasting standards of broadcast programs, and also to impose penalties on those who violate them. In performing its duties, the KPI issued Regulation of the Komisi Penyiaran Indonesia No. 02/P/KPI/12/2009 about the Code of Broadcasting Conduct [1] and KPI Regulation No 03/P/KPI/12/2009 about the Broadcasting Program Standards [2]. In accordance with the legislation, the codes of broadcasting conduct and broadcasting program standards, the infringements can be an audio/speech-based infringement, a visual infringement or a content-based infringement. Some examples of audio-based infringements in a broadcasting program can include harsh words, cursing, harassing, insulting or degrading minorities and marginalised society groups such as a group with a specific job (domestic workers, security guards), a group that has an aberration (transsexuals), a group with unordinary size and physical form (buck teeth, overweight, midget, strabismus), a group with physical disabilities (deaf, blind, mute), a group who have a mental retardation or disability (autism, idiotism), and a group of people with specific diseases (HIV/AIDS, leprosy, epilepsy, Alzheimer's, latah). Examples of visual infringements are the exploitation of body parts which are commonly considered to be able to generate sexual desire, such as thighs, buttocks, breasts and/or genitals. The last one, an example of content-based infringement, is for example an announcement that promotes and encourages prostitution to be accepted by religion and society.

The KPI consists of KPI Pusat and KPI Daerah (KPID) at the provincial level. One KPID that has already been formed is KPID in Central Java. In carrying out surveillance of both television and radio broadcasting, KPID Central Java still does this manually (direct supervision by people). Several televisions are turned on to be seen and monitored by officers to find out whether the delivery of broadcast content is against the rules. This has certainly had some weaknesses related to human error, such as tiredness and weary eyes. Inconsistencies in assessment can also occur because of the subjectivity of those supervising the broadcast

Received January 3, 2015; Revised March 17, 2015; Accepted April 3, 2015

programs. Therefore, we need intelligent software that can automatically detect the infringements of broadcast content, which has never been studied up to now.

This research aims to design a system that can automatically identify infringements committed against the law, the codes of broadcasting conduct and the broadcast program standards, with the result that it can help the KPI's performance, especially KPID Central Java, in controlling broadcast content. The system has two main stages. They are Indonesian-language speech recognition and detection of infringements in the broadcast program. Audio-based infringement can be detected automatically by the existence of a speech recognition application. As regards programs broadcast in the Indonesian language, they definitely require a speech recognition system in the Indonesian language. The research of sound in Indonesia is relatively limited, such as the identification of gamelan instruments [3] and an Indonesian-language speech recognition system. An Indonesian-language speech recognition system is still limited to writing a simple message on a mobile device; it is far behind English-speech recognition systems, of which several have already been applied to related fields [4]-[6]. For that, Indonesian-language speech recognition needs to be further developed, so it can be used in various fields, especially in the automatic detection of infringements in audio broadcast programs.

Research on speech recognition started in the 1950s [7]. Speech recognition can be defined as the process of converting voice signals into the ranks of the word, by applying a specific algorithm that is implemented in a computer program. It is also often called Speech to Text. There are two main processes in speech recognition: feature extraction and recognition. Various methods have been developed to produce a high level of accuracy. Feature extraction techniques that have been developed include Linear Predictive Analysis (LPC), Cepstral Analysis [8], Mel Frequency Cepstral Coefficients (MFCC) [9], Wavelet Cepstral Coefficients (WCC) [10] and retrieval based prosodic features [11]. Basically there are three approaches to speech recognition, namely [12] the acoustic-phonetic approach, pattern recognition approach and artificial intelligence approach. The speech recognition technique that is included in pattern recognition is the Hidden Markov Model (HMM) and Support Vector Machine (SVM) [8]. Research conducted by a group of researchers from the Lab Riset Sistem Cerdas Jurusan Ilmu Komputer/Informatika has been able to prove that the Support Vector Machine (SVM) method can be used to classify harsh words and non-harsh words. Swear words constitute a violation of Law No 32 on broadcasting. The results showed that spoken words can be classified with a high enough degree of accuracy (72.5% to 92.5%) [13]. The sound features used are the value of pitch, because at the beginning of the study, it was analysed as being useful for distinguishing the pronunciation of the word in the form of insults and non-insults [14].

Singh et al. (2012) and O'Shaughnessy (2008) [12],[15] mentioned that the technique for speech recognition which consists of hundreds of thousands of words and that is still accepted up to now is the Hidden Markov Model (HMM), which appeared in 1975. According to Rabiner (1989), HMM is a stochastic process that occurs twice, with one of them being not a direct observation. A hidden stochastic process can be observed only through another set of stochastic processes that can produce the sequence of observation symbols. This is the reason that causes HMM to perform better than other methods [9]. In addition, the HMM technique is generally accepted in current speech recognition systems (state-of-the-art) in modern times because of two reasons, namely, its ability to model the non-linear dependence of each unit of the sound on the unit in question, and because it is a set of powerful analytical approaches that are available to estimate the model parameters [12]. This study uses HMM as a method in speech recognition of broadcasting programs.

2. Research Method

A general description of the system that is built can be seen in Figure 1 below. The Integrated System of Infringement Broadcasting Program Detection (*Sistem Terintegrasi Deteksi Pelanggaran Program Siaran,* abbreviated as SINDEPROSI) is constructed in accordance with the general picture in Figure 1 below.

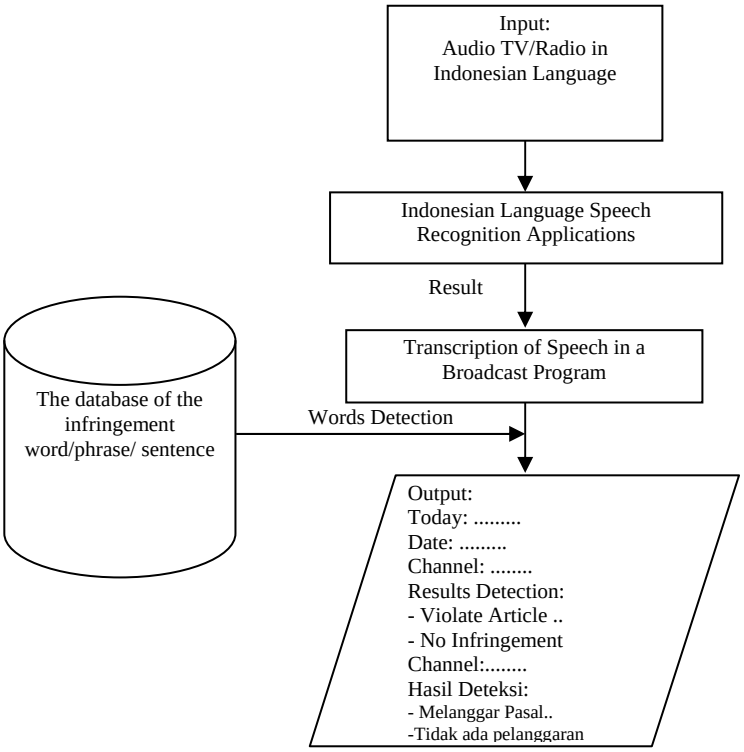

Figure1. System Description

The following steps of the process in SINDEPROSI are:

a. Audio Input
   The input can be the audio from either television or radio programs that use the Indonesian language.
b. Speech Recognition
   Spoken audio broadcasting programs will be recognised and will be written in text form. To design these speech recognition applications, the steps of the process can be seen in Figure 2.

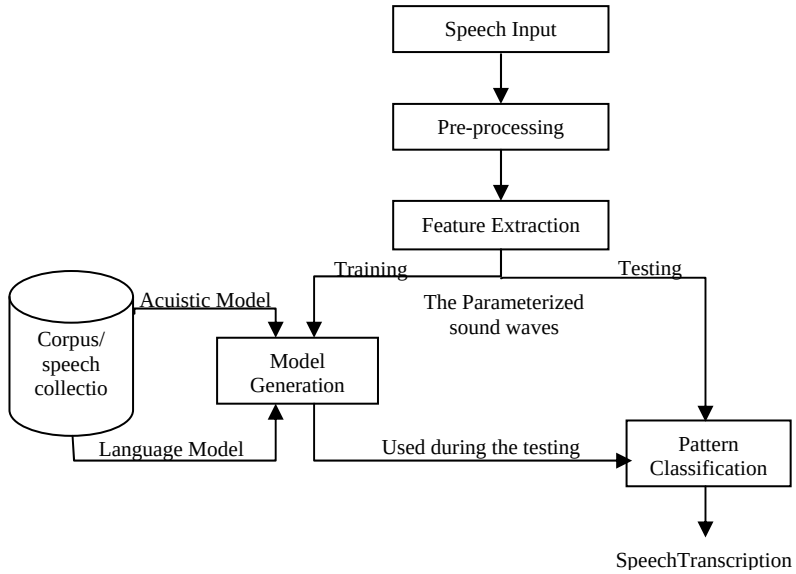

Figure 2. Process of Speech Recognition Application

Here is the explanation of each process.

1) Speech input
   Speech input can be a video recording or taking a sound file with .wav extension.
2) Pre-processing
   Pre-processing on MFCC includes DC-removal, Pre-emphasise, Frame Blocking and Windowing. DC-removal is used to obtain the normalised value of the input data, by calculating the average of the sample utterance data. Pre-emphasise is used to reduce the noise ratio in the signal and balance the spectrum of the voice. The Frame Blocking process is used to cut a voice signal of long duration and shorten it, in order to obtain the characteristic signal periodically. The Windowing process aims to reduce spectral leakage or aliasing, which is the effect of Frame Blocking that causes the signal to become discontinued.
3) Feature extraction
   The method used for this is the Mel Frequency Cepstral Coefficients (MFCC). This stage has several processes, which are:
   i. FFT (*Fast Fourier Transform*)
      FFT is a transformation method to get a signal in the frequency domain of the available discrete signals.
   ii. *Mel-Frequency Wrapping*
      Filterbank is carried out in order to determine the energy in the sound signal. The frequency of a signal is measured using the Mel scale.
   iii. *Cepstrum*
      Mel-Frequency Cepstrum is obtained from the DCT process to get the signal back in the time domain. The result is called the Mel-Frequency Cepstral Coefficient (MFCC).
   iv. *Cepstral Filtering*
      The MFCC results have several weaknesses, namely a very sensitive low-order to the spectral slope and a very sensitive high-order to noise. Therefore, the cepstral filtering becomes one of the methods to minimise such sensitivity.
4) Model Generation
   This process models the patterns of speech into text form based on the training data that are used by employing the Hidden Markov Model (HMM).
5) Pattern Classification
   The final stage takes form a pattern classification that searches for equivalency between the formed generation models and the tested data.

c. Speech Transcription
   Transcription of speech is the result of the speech recognition process for broadcasting programs that have been included.
d. Infringement Detection
   From the transcription's results, the available text or word will be checked in the infringement regulation words database. The results of these checks can be printed as a report for the broadcasting program targeted.

3. Results and Discussion

This system has two users, each of whom acts as Admin and Operations Officer. Admin is the person who is in charge of managing user data, performing the training process in the creation of Indonesian-language speech recognition applications, managing the list of word infringements and managing the relevant legislation data related to Broadcasting Law No 32 of 2002, the Komisi Penyiaran Indonesia Regulation No 02/P/KPI/12/2009 about The Code of Broadcasting Conduct and the Komisi Penyiaran Indonesia Regulation No 03/P/KPI/12/2009 about the Broadcasting Program Standards. The Operations Officer is an officer from KPI who runs this system by entering the broadcasting program data and audio input both from television and radio programs.

The resulting system design includes architecture design and main module interface design.

a. Architecture Design
   Architecture design is shown in Figure 3.

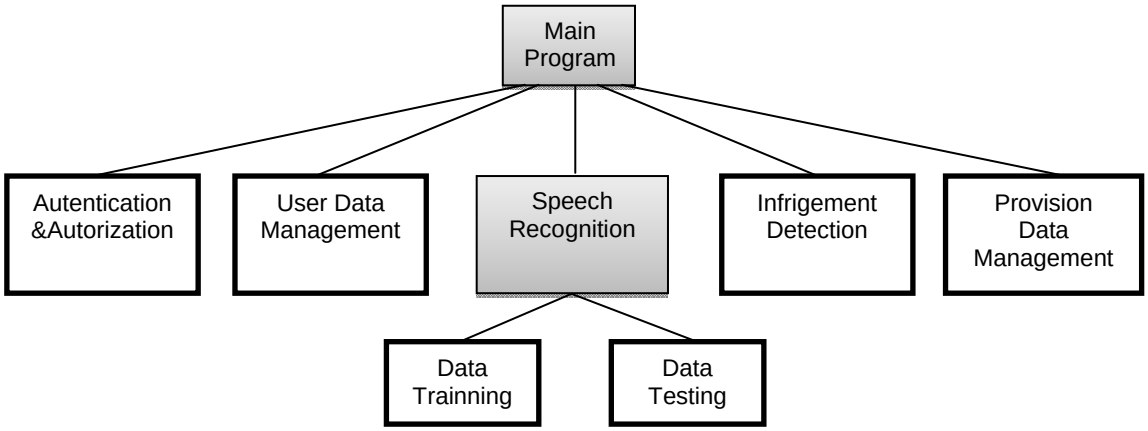

Figure 3. Architecture design

Here is an explanation of each main program.

1) Authentication and authorisation
   This menu is used as an identifier for users of the system, by displaying the log-in menu for the username and password entered by the user of the system.
2) User Data Management
   This menu is used to manage user data.
3) Speech Recognition
   Speech recognition is a major program in the form of Indonesian-language speech recognition applications.
4) Infringement Detection
   This menu is also the main program, which is used to detect violations of the broadcast program.
5) Provision Data Management
   Provision Data Management is used to manage the data legislation.

b. Main Module Interface Design
Main module interface design is divided into three procedures as follows:

Procedure Data Training (speech_trainning_data: voice, vartrainning_recognition_text: char)
Initial State : System waiting for the data input of voice training
Final State : System issued data modelling with HMM
Algorithm :
1) The system accepts data input of voice training
2) The system preprocessing voice training data
3) The system extracting feature extraction
4) The system produces data modelling with HMM

Procedure Data Testing (speech_test_data: voice, broadcasting_program_data: voice, varrecognition_result_text: char)
Initial State : System waiting for data input of voice test or broadcast program data
Final State : System issues text recognition results
Algorithm :
1) The system accepts data input of voice test or broadcast program data
2) The system preprocessing voice test data or broadcast program data
3) The system extracting the feature
4) The system recognising speech pattern with user database as the input
5) System issues text recognition results

Procedure Infringement Detection (recognition_result_text: char, provision_data: char, varinfrigment_report: char)
Initial State : System awaits input text recognition results
Final State : System issues infringement report
Algorithm :
1) The system accepts input text recognition results
2) The system checks recognition results text with database of infringements data as the input
3) The system determines whether the recognition results text is included as an offence or not
4) The system stores the recognition results text as an infringement if the recognition results text is included in the infringement data
5) System issues an infringement report

In creating speech recognition, a speech database is used, involving 50 people with variations of the following characteristics:

a. Young Group: 15-25 years old
   Number of male data samples : 15 people
   Number of female data samples : 15 people
b. Adult Group: 25-45 years old
   Number of male data samples : 10 people
   Number of female data samples : 10 people

Selection of the above age groups is considered based on the fact that the broadcast program contains infringements mostly spoken by the age group.

Of the 50 data samples, some come from different tribes and areas in Indonesia, because each tribe or region has its own accent in uttering a syllable, word or phrase in the Indonesian language. Meanwhile, the ethnicity and regions of data samples in the data retrieval words are Javanese, Sundanese, Betawi, Sumatra, Medan, Sumatra Padang, Sumatra Batak, Jambi, Riau, Bali and Sulawesi.

Each participant in collecting data pronounced 1572 syllables and 1658 words; the words uttered represent all existing syllables, where syllabification refers the Great Dictionary of the Indonesian Language (KBBI). Spoken syllable has 11 species consisting of Singing (V) and consonants (K). The types and their 11 number of syllables are V (5), VK (61), KV (97), KVK (919), KKV (56), KKVK (250), VKK (10), KKVKK (44), KKKV (5), KKKVK (11), and KVKK (114).

Testing for speech recognition is performed on 21 existing words; the sample data used are 1050 data, and the data are divided into 60% training data and 40% testing data. This test uses the MFCC coefficient 8, and 14 in the HMM state. Experiments are carried out by changing the voice data used in the training process. Table 1 shows the experimental results.

Table 1. Experiments Result

Training Data	Experiments	Degree of Accuracy (%)	
		Man Speech	Woman Speech
Man Speech	1	80.95	23.8
	2	85.71	19.05
	Average	83.33	21.42
Woman Speech	1	23.8	80.95
	2	19.05	71.42
	Average	21.45	76.18
Man and Woman Speech	1	76.19	61.90
	2	76.19	76.19
	Average	76.19	69.04

Table 1 shows that the gender used as training data have a great influence on the results of the accuracy of the system. The tests using training data containing both male and female utterances can produce a level of accuracy that is relatively comparable between testing with male voice data and female voice data. Thus in the process of detection of broadcast program infringements, training data that contain the voices of both men and women will be used.

4. Conclusion

The design results of the broadcast program infringement detection system have been built in the form of architectural design and the main module interface design. This design has been tested for the speech recognition testing phase. Implementation of the system will be developed using training data that contain men and women from young and adult groups.

References

[1] Sendjaja SD (Chief of Komisi Penyiaran Indonesia Pusat). Regulation of Komisi Penyiaran Indonesia No 02/P/KPI/12/2009 about The Code of Broadcasting Conduct. Jakarta. 2009.
[2] Sendjaja SD (Chief of Komisi Penyiaran Indonesia Pusat). Regulation of Komisi Penyiaran Indonesia No 03/P/KPI/12/2009 about The Broadcasting Program Standard. Jakarta. 2009.
[3] Tjahyanto A, Suprapto YK, Wulandari DP. Spectral-based Features Ranking for Gamelan Instruments Identification using Filter Techniques. *TELKOMNIKA Telecommunication Computing Electronics and Control*. 2013; 11(1): 95–106.
[4] Jadhav A, Patil A. A Smart Texting System for Android Mobile Users. *International Journal of Engineering Research and Applications*. 2012; 2(2): 1126-1128.
[5] Jadhav A, Patil A. Android Speech to Text Converter for SMS Application. *IOSR Journal of Engineering*. 2012; 2(3): 420-423.
[6] Sharma FR, Wasson SG. Speech Recognition and Synthetis Tool: Assistive Technology for Physically Disabled Persons. *International Journal of Computer Science and Telecommunications*. 2012; 3(4): 86-91.
[7] Patel I, Rao YS. Speech Recognition using HMM with MFCC-AN Analysis using Frequency Spectral Decomposion Technique*. Signal & Image Processing*: *An International Journal.* 2010; 1(2): 101-110.
[8] Anusuya MA, Katti SK. Speech Recognition by Machine: A Review. *International Journal of Computer Science and Information Security (IJCSIS)*. 2009; 6(3): 181–205.
[9] Kumar K, Aggarwal RK. Hindi Speech Recognition System Using HTK. *International Journal of Computing and Business Research*. 2011; 2(2): 2229-6166.
[10] Adam TB, Salam MS, Gunawan TS, Wavelet Cepstral Coefficients for Isolated Speech. *TELKOMNIKA Telecommunication Computing Electronics and Control*. 2013; 11(5): 2731–2738.
[11] Wang Y, Yang X, Zou J. Research of Emotion Recognition Based on Speech and Facial Expression. *TELKOMNIKA Telecommunication Computing Electronics and Control*. 2013; 11(1): 83–90.
[12] Singh B, Kapur N, Kaur P. Speech Recognition with Hidden Markov Model: A Review. *International Journal of Advanced Research in Computer Science and Software Engineering.* 2012; 2(3): 400-403.
[13] Endah SN, Nugraheni DMK. Klasifikasi Ucapan Kata dengan Support Vector Machine. *Jurnal Masyarakat Informatika*. 2012; 3(6): 7-14.
[14] Endah SN, Nugraheni DMK. *Analisis Pitch dan Formant Sinyal Ucapan Kata*. Prosiding Seminar Nasional Ilmu Komputer Universitas Diponegoro 2012. Semarang. 2012; 101–107.
[15] O'Shaughnessy D. Invited paper: Automatic Speech Recognition: History, Methods and Challenges. *Pattern Recognition Science Direct.* 2008; 41: 2965-2979.