

Chapter 5

Filter Selection for Speaker Diarization Using Homomorphism: Speaker Diarization

K. Jairam Naik

 <https://orcid.org/0000-0002-6332-418X>
National Institute of Technology, Raipur, India

Awani Mishra

 <https://orcid.org/0000-0002-6396-1517>
National Institute of Technology, Raipur, India

ABSTRACT

In the current scenario, audio classification and recognition of a particular source of voice is a major challenge when several speakers are speaking at a time. The process of identifying each speaker in an audio segment is called speaker diarization. The major steps involved in speaker diarization are speech detection, speaker change, and speaker merges. Finding and suggesting the best filter is one of the most important tasks involved in every step of this process. No current researches at present have yet focused on the impact of filter with optimized approaches. In this chapter, a simple yet effective method using homomorphism has been implemented to recommend the best filter for any audio classification task for this purpose. After having performed the classification task for a number of commonly used filters, their accuracies have been compared at every step of speaker diarization. In this work, a module using Bayesian information criterion, convolutional neural networks is implemented, and diarization algorithm that performs the task of speaker diarization is proposed.

DOI: 10.4018/978-1-7998-3238-6.ch005

1. INTRODUCTION

Given an audio or a video clip, it is desirable to have an algorithm that segments the clip into portions where the input audio stream can be attributed to energy sources like a person speaking or music or noise, and then automatically partition the input audio stream into homogeneous segments assigning each segment to a particular speaker. Doing this has benefits that include speaker diarization during long meetings in corporate environment. To this end there is already a survey conducted by NIST where they evaluated telephonic conversations and broadcast news [NIST, 2003].

In order to do this, the most important criteria that we use is the frequency of the audio input at any point of time and based on its value, we may be very well able to classify the audio as being sound, music or the voice of a particular person after we have analyzed a sufficiently long input stream to learn a pattern with frequency being the covariate. Simply speaking, speaker diarization is our main challenge and is stated as: Given an Audio Recording where N numbers of people are speaking in different intervals, our task is to label and cluster all intervals where a particular speaker is speaking. Feature vector selection for audio classification is a well-researched area in audio processing.

The proposed method is an intuitive approach to solve a common audio processing problem. The idea for this problem came when I was finalizing my intern Report noticed how much informative spectrogram can be for the purpose of detecting presence of music in an audio clip. Upon further research, by generating more spectrograms, tuning parameters, it became apparent that for most of audio classification tasks, perfect filters do exist.

Most of the methods used for filter selection in the project are formalized by this group without much external reference and are not previously implemented in a similar fashion, best to our knowledge. The algorithm used for computing homomorphism can be sped up by using greedy algorithm by assigning each cluster the label which occurs most frequently in it.

There has been an excellent survey on the state-of-the-art methods on speaker diarization [Anguera et al, 2012]. The algorithm that we used for speaker diarization is inspired by the works of [Qin Jin et al, 2004].

We present the objectives, motivation and novelty of our work along with literature review in Section 1 and 2 respectively. Section 3 talks about filter selection and Speaker Diarization using the proposed methodology. The software tools used and result analysis are discussed in Section 4. Section 5 proposes the conclusion of the work.

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/filter-selection-for-speaker-diarization-using-homomorphism/269583

Related Content

40-GHz Inductor Less VCO

Abhishek Kumar (2020). *AI Techniques for Reliability Prediction for Electronic Components* (pp. 288-298).

www.irma-international.org/chapter/40-ghz-inductor-less-vco/240503

Literature Survey for Applications of Artificial Neural Networks

Pooja Deepakbhai Pancholiand Sonal Jayantilal Patel (2022). *Research Anthology on Artificial Neural Network Applications* (pp. 669-682).

www.irma-international.org/chapter/literature-survey-for-applications-of-artificial-neural-networks/288981

Neural Networks in Manufacturing Operations

Eldon Gunnand Corinne MacDonald (2006). *Artificial Neural Networks in Finance and Manufacturing* (pp. 165-181).

www.irma-international.org/chapter/neural-networks-manufacturing-operations/5355

A Method for Identifying Fatigue State of Driver's Face Based on Improved AAM Algorithm

Zuojin Li, Jun Peng, Liukui Chen, Ying Wuand Jinliang Shi (2020). *Deep Learning and Neural Networks: Concepts, Methodologies, Tools, and Applications* (pp. 1263-1278).

www.irma-international.org/chapter/a-method-for-identifying-fatigue-state-of-drivers-face-based-on-improved-aam-algorithm/237933

A Neural-Network-Assisted Optimization Framework and Its Use for Optimum-Parameter Identification

Tapabrata Ray (2006). *Artificial Neural Networks in Finance and Manufacturing* (pp. 221-235).

www.irma-international.org/chapter/neural-network-assisted-optimization-framework/5358