|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Mini Project and Seminar** | | | | | | | | |
| **Class & Batch** | | **:** | **TE-5(N5)** | | **Group No.** | | **:** | **2** |
| **Synopsis**  **On** | | | | | | | | |
| **“SINGER RECOGNITION”** | | | | | | | | |
|  | | | | | | | | |
| **Submitted By** | | | | | | | | |
|  | | | | | | | | |
| **Roll No.** | **Name of the Student** | | | **e-mail id** | | **Mobile No.** | | |
| **3559** | **AADITI SAOJI** | | | **aaditisaoji96@gmail.com** | | **9545942310** | | |
| **3560** | **YASH SHAH** | | | **96yaash@gmail.com** | | **7219122923** | | |
| **3568** | **SUSHIL KHAIRNAR** | | | **sushilkhairnar84@gmail.com** | | **9767362011** | | |
|  | | | | | | | | |
| **Project Guide**  **Prof. H.S.Thakar** | | | | | | | | |
|  | | | | | | | | |
|  | | | | | | | | |
|  | | | | | | | | |
|  | | | | | | | | |
| **DEPARTMENT OF ELECTRONICS AND TELECOMMUNICATIONS**  **PUNE INSTITUTE OF COMPUTER TECHNOLOGY** | | | | | | | | |
| **Academic Year : 2016 – 2017** | | | | | | | | |

1. **Project Title :SINGER RECOGNITION**
2. **Introduction/ Motivation :**

The singer's information is essential in organizing, browsing and retrieving music collections. We wish to develop a system for automatic singer identification which recognizes the singer of a song by analysing the music signal.

Hitherto, there has been no solution proposed for solving the singer identification problem. Nevertheless, human beings can recognize the voice of a familiar singer, and distinguish similar singing voices by listening to only a small portion of the song. Therefore, we believe that by extracting and analysing audio features properly, an automatic system should be able to achieve certain degree of singer identification as well.

This technique can be used to easily retrieve all songs performed by a particular singer in a distributed music database. Furthermore, it can be used to cluster songs of similar voices of singers in a music collection, or search for songs which are similar to a query song in terms of the singer’s voice. The technique of automatic singer identification will add significant functionalities to a digital music management system.

1. **Literature Survey / Prior work :**

In **Singer Recognition** we didn’t find any organization or an individual who has worked explicitly on this topic. But a lot of work has been done on **speaker** recognition in the industry, technological labs and educational universities. Speaker recognition is a [pattern recognition](https://en.wikipedia.org/wiki/Pattern_recognition) problem+ which we call as biometric system The various technologies used to process and store speech signal include [frequency estimation](https://en.wikipedia.org/wiki/Frequency_estimation), [hidden Markov models](https://en.wikipedia.org/wiki/Hidden_Markov_model), [Gaussian mixture models](https://en.wikipedia.org/wiki/Gaussian_mixture_model), [pattern matching](https://en.wikipedia.org/wiki/Pattern_matching) algorithms, [neural networks](https://en.wikipedia.org/wiki/Neural_networks), [matrix representation](https://en.wikipedia.org/wiki/Matrix_representation), Vector Quantization and [decision trees](https://en.wikipedia.org/wiki/Decision_tree_learning). Some systems also use "anti-speaker" techniques, such as [cohort models](https://en.wikipedia.org/wiki/Cohort_model), and world models. Spectral features are predominantly used in representing speaker characteristics.In1974, AT & T (American telephone and telegraph) had designed a text dependent system, in which cepstrum features were taken. In that system only 2% of verification and recognition error had been detected.

The first international patent was filed in 1983, coming from the telecommunication research in [CSELT](https://en.wikipedia.org/wiki/CSELT) (Italy) by Michele Cavazza and [Alberto Ciaramella](https://en.wikipedia.org/wiki/Alberto_Ciaramella) as a basis for both future telco services to final customers and to improve the noise-reduction techniques across the network.

Many organizations like Massachusetts Institute of Technology Lincoln Labs, National TsingHua University (Taiwan), Nagoya University (Japan), Nippon Telegraph and Telephone (Japan), Rensselaer Polytechnic Institute, Rutgers University, and Texas Instruments (TI) have developed much accurate speaker recognition system using different features.

In August 2014 GoVivace Inc. deployed a speaker identification system that allowed its telecom industry client to positively search for an individual among millions of speakers by using just a single example recording of their voice.Speaker recognition may also be used in criminal investigations, such as those of the 2014 executions of, amongst others, [James Foley](https://en.wikipedia.org/wiki/James_Foley_%28journalist%29) and [Steven Sotloff](https://en.wikipedia.org/wiki/Steven_Sotloff).In February 2016 UK high-street bank [HSBC](https://en.wikipedia.org/wiki/HSBC) and its internet-based retail bank [First Direct](https://en.wikipedia.org/wiki/First_Direct) announced that it would offer 15 million customers its biometric banking software to access online and phone accounts using their fingerprint or voice.

1. **Problem Definition and Objectives :**

The main objective of the proposed scheme is to automatically identify the singer of a song by preprocessing the signal, extracting features of the singer and then matching it with features provided in the database. It follows the framework of common speaker identification systems, but special efforts are made to distinguish the singing voice from instrumental sounds in a song.

The proposed scheme for automatic singer identification contains two phases: Training phase, and Working phase.

In the training phase, training audio samples are selected from one or more typical songs of each singer, and audio features of these samples are computed. These audio features form a vector-sequence which is used to build a statistical model of the singer’s voice.

Then, in the working phase, for a song to be identified, the starting point of the singing voice in the song is first detected, and a fixed length of testing data is taken from that point. The testing data are divided into testing samples, and audio features are computed of these samples. The audio features are matched with existing singers’ models in the database.

1. **Block Diagram :**

**PRE-PROCESSING**

Energy

**FILTERING**

(Separation of vocals and music in a song)

**FEATURE EXTRACTION** (MFCC and PITCH) **ANDSTATISTICAL COMPARISON**

Frame blocking

**DECISION**

**FEATURE MATCHING**

**VERIFICATION/ RESULT**

Windowing

Memory

STFT

**HARDWARE**

**BLOCK DIAGRAM**

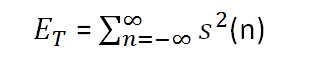
Processing of audio signals in applications such as data compression, noise reduction, or singer recognition usually requires realization of several specific stages. Such a multistage processing for the singer recognition is composed of: a parameterization of the input signal (pre- processing e.g. with calculation of energy ,frame blocking(i.e dividing the audio signal into small frames of 10-20 msec), multiplication by a window function, calculating STFT(Short Term Fourier Transform), determining the linear prediction coefficients or mel-cepstral coefficients with the use of the FFT), a modeling phase (such as the vector quantization) and then comparing with the base pattern. A simplified block diagram of singer recognition is shown in the figure above.

**Filtering:**

In this stage,the song input is taken and processed in audacity software, wherein we separate the music and vocal from a song.The filtering is done in such a way that only the vocal part i.e lyrics are filtered out and the background music is eliminated.

**Energy calculation:**the energy associated with speech is time varying in nature. Hence the interest for any automatic processing of speech is to know how the energy is varying with time and to be more specific, energy associated with short term region of speech. By the nature of production, the speech signal consist of voiced, unvoiced and silence regions. Further the energy associated with voiced region is large compared to unvoiced region and silence region will not have least or  negligible energy. Thus short term energy can be used for voiced, unvoiced and silence classification of speech.

The relation for finding the short term energy can be derived from the total energy relation defined in signal processing.The total energy of an energy signal is given by



In case of short term energy computation we consider speech in terms of 10-30 msec . Let the samples in a frame of  speech are given by ***"n=0 to n=N-1"***, where ***" N "*** is the length of frame (samples), then for energy computation the  speech will be zero outside the frame length. Then for energy computation amplitude of the speech samples will be zero outside the  frame.

**Frame Blocking:**

In this step, the continuous speech signal is blocked into frames of *N* samples, with

adjacent frames being separated by *M* (*M < N*). The first frame consists of the first *N*

samples. The second frame begins *M* samples after the first frame, and overlaps it by *N - M*

samples. Similarly, the third frame begins 2*M* samples after the first frame (or *M* samples

after the second frame) and overlaps it by *N* - 2*M* samples. This process continues until all

the speech is accounted for within one or more frames.

**Windowing:**

After frame blocking we go for windowing. In windowing we multiply

each frame by the window function so as to minimize the signal discontinuities

at the beginning and end of each frame. The idea is to set the signal value to

zero in the beginning and end of each frame.

We used a Hamming window for minimizing the interference from side

lobes.The Hamming window has the formula = 0.54-cos(2\*pi\*n/N-1) for 0<=n<=N-1.

**STFT – Short Time Fourier Transform:**

Then we compute the fast Fourier transform which maps each frame

onto the frequency domain.The FFT actually reduces the number of

calculations required to compute the DFT of a given set of samples.

Here the X(K) values are complex numbers and we consider only their

absolute values.

Positive frequencies 0≤F≤Fs/2 correspond to 0 ≤n≤N/2-1

Negative frequencies -Fs/2≤F<0 correspond to N/2≤n≤N-1

Where Fs is the sampling frequency.

**Mel Frequency Wrapping:**

Psychological studies have found out that our perception of frequency

contents for speech signals doesn’t follow a linear scale. Due to this,for each

tone with actual frequency f(in Hz), a subjective pitch is found out and is

measured on a scale known as **mel scale.** The mel scale is a special scale.It

has a linear frequency spacing for frequencies upto 1 kHz and above that it

has a log scale.

To create a virtual subjective spectrum we use a mel filter bank.For

computational simplicity we use a triangular band pass filter whose bandwidth

is determined by a mel frequency interval which is typically a constant. We

chose the number of melcepstrum coefficients to be say,20.We have shown

below a triangular mel band pass filter.We can visualize this filter as a

histogram bin in frequency domain.

**Feature Matching**

Speaker recognition problem comes under the category of pattern recognition.

The main aim of pattern recognition is to divide a number of objects into

classes.The objects to be classified are called patterns and the process of

classifying them is called pattern classification.The most well known algorithms for pattern recognition used in speaker

recognition algorithms are Dynamic Time Warping(DTW),Hidden Markov

Modeling(HMM), and Vector Quantization(VQ)

1. **Hardware and Software Requirements:**

* TMS320-6713 DSP PROCESSOR
* Laptop
* MATLAB - Simulink
* CCS – Simulink plugin

1. **References:**

* Campbell ,J.P., Jr. , “Speaker recognition a tutorial ” Proceedings of the IEEE Volume 85,Issues 9, Sept. 1997 Pages:1437 – 1462
* iitg.vlab.co.in(IIT Guwahati,speech processing virtual laboratory)
* Seddik, H.; Rahmouni, A.; Samadhi, M.; “Text independent speaker recognition using the Mel frequency cepstral coefficients and a neural network classifier” First International Symposium on Control, Communications and Signal Processing, Proceedings of IEEE 2004 Page(s):631 – 634.
* <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>