Accent Identification

T K Harshitha Devang
Dept. of Electronics & Communication
GSSSIETW, Mysuru, Karnataka, India

Sushma M Tilave
Dept. of Electronics & Communication
GSSSIETW, Mysuru, Karnataka, India

Sowmya M.S.
Dept. of Electronics & Communication
GSSSIETW, Mysuru, Karnataka, India

Tejaswini
Dept. of Electronics & Communication
GSSSIETW, Mysuru, Karnataka, India

Abstract**—Accent is the pattern of pronunciation and acoustic in speech which can identify a person’s linguistic origin. Accent is a person’s way of speaking. This paper describes an automatic identification system of English accents for different countries where the accent of the speaker will be identified. We identified the accent for four different countries such as Indian, American, Chinese and Russian. The approach is based on obtaining spectral envelope, glottal excitation waveform and building a reference pattern. The glottal excitation waveform is obtained using spectral tilt. Spectral tilt was calculated as the slope of the line of best fit. The result is obtained by comparing the sample to be identified with the builded reference pattern. The reference pattern is built using Mean and Standard Deviation of the spectral tilt values and fixing ranges to each accent. Accent identification is a crucial factor to the fluency of automatic speech conversion system.**

INTRODUCTION

A. Overview

Speech is the power of expressing or communicating one’s thoughts by speaking. The information that is communicated through speech is intrinsically of discreet nature; hence for the better understanding of speech it is converted to signal form. Speech signals are symbolic representation of information.

Speech production involves three processes. They are

- Respiration
- Phonation
- Articulation

Respiration: Respiration is the body’s cyclic intake and exhalation of air. When we prepare to breathe, the diaphragm drops. This causes the volume with in the lungs to expand and air swoops in the nose or mouth, down through the larynx and into the trachea, bronchi and lungs.

Phonation: We need to create pressure and vibration to produce sound. Phonation means making sound. As air passes through larynx, tissues vibrate to produce sound waves. Faster the vocal folds vibrate, the higher the pitch of voice produced.

Articulation: Articulation is shaping of raw sound into recognisable speech. Without articulation, voice at the level of vocal folds the sound would be buzzing noise. There are 19 points between the vocal folds and the lips allow us to produce speech. Some are: the lips, teeth, pallets, tongue, vocal folds, nasal cavity and jaw.

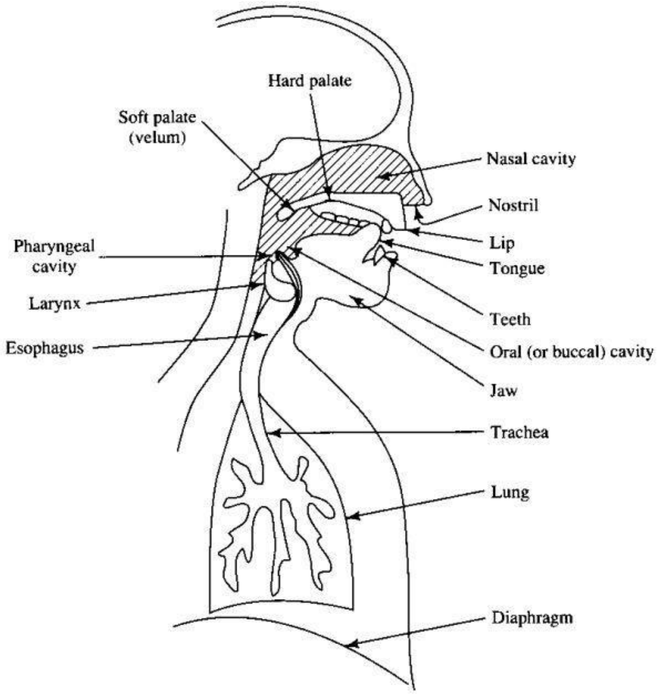


Figure 1: Schematic diagram of human speech production mechanism

Speech varies in a distinctive way of pronouncing a language, especially one associated with the particular area or a social class which is called ACCENT. Accent is the way how people sound when they speak. There are two different kinds of accent. One is a “Foreign accent”: this occurs when a person speaks one language using some of the rules or sounds of another one. For example, if the person has trouble in pronouncing some of the sounds of the second language they are learning, they may substitute similar sounds that occur in their first language. The other kind of accent is a simple way where a group of people speak their native language which is determined by where they live.

Accents usually come from the articulation habits of the speaker in his/her own native language. It has been shown that speaker’s first language affects production and perception of English as a second language and that these effects are experienced dependent.

B. *Objectives*

The main aim of this project is to identify the accent of the speaker. To achieve this, firstly different English accent such as Indian, American, Chinese and Russian are recorded. Then reference pattern for the above accents are constructed. Finally, the sample to be identified is compared with the builded reference pattern and accent is identified.

II LITERATURE SURVEY

“Accent identification by clustering and scoring formants” by Denjan Stantic and Jun Jo[1]- In this paper they addressed a problem of identifying the English accented speech of speakers from different background based on the pronunciation of a short phrase. The Q factor has been introduced which is defined by sum of relationships between frequencies of the formants and showed that this factor can be used to identify certain English accents. The scoring method and proposed concept indicates, accent can be identified by their formants.

“Automatically identifying characteristic features of non-native English accents” by Jelke Bloem, Martijin Wieling and John Nerbonne[6]- This paper statistically measures the accented English speech. They used combination of representativeness and distinctiveness method. A high representativeness indicates that the differences between pronunciations within the dialect area are small. A high distinctiveness indicates that the differences between pronunciation within and outside the dialect area are large.

Accent identification is a challenging problem related to field of multilinguality area such as dialect identification and language identification [7]. ”Accent Identification” by Carlos Teixeira, Isabel Trancoso and Antonio Serralherio- In this paper tested accent identification in isolation and integrated in a speech recognition system.

III METHODOLOGY

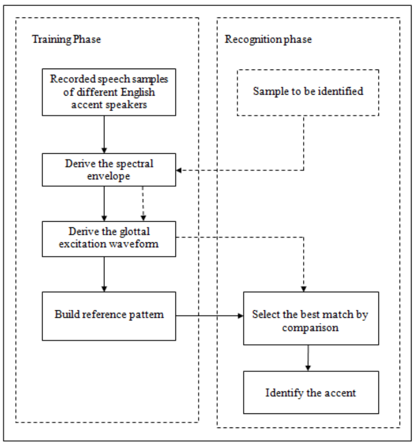

Training Phase
Recorded speech samples of different English accent speakers
Derive the spectral envelope
Derive the glottal excitation waveform
Build reference pattern
Recognition phase
Sample to be identified
Select the best match by comparison
Identify the accent

Figure 2: flow chart of process involved in the accent identification

The Figure 2 represents the procedure for Accent Identification. It has two phases as follows:

A. Training Phase

Step 1: Recorded speech samples of different English accents.
Step 2: Obtaining its spectral envelope.
Step 3: Derive the glottal excitation waveform from spectral envelope.
Step 4: Build the reference pattern.

B. Building reference pattern

Step 1: Slope values of all the samples are obtained.
Step 2: Obtained the mean and standard deviation.
Step 3: Threshold has been fixed to 0.006
Step 4: Range has been fixed to the accents.
Step 5: The above steps is repeated for all the accents.

C. Recognition Phase

Step1: Sample which is to be identified is given as input, to derive the spectral envelope.
Step 2: Derive the glottal excitation waveform from spectral envelope.
Step3: Select the best match on comparison between the derived accent and reference pattern.
Step 4: Identify the accent.

The Harvard sentences are used for recording. It is recorded using head phone with microphone by the speakers. After the natural voice obtained by the microphone headset, it was then formatted as analog signal whereas the digital form of voice signal can be analysed easier. The designed system firstly requires a voice pre-processing technique, which deals with the natural human speech from analog formation into the form of machine understandable digital signal.

Voice pre-processing technique involves:

- ➢ Analog to digital conversion
- ➢ Frame blocking and windowing
- ➢ Pre-emphasis technique
- ➢ End point detection

A. *Analog to digital conversion*

ADC converts a continuous time and continuous amplitude analog signal to a discreet time and discreet amplitude digital signal. Using sampling and quantization techniques, analog signal is converted into digital signal.

B. *Sampling and quantization*

Sampling technique breaks up the sound wave into intervals along the time axis to produce a sequence of signals [10]. As a result, the values in time axis are converted from continuous to discreet values with corresponding magnitudes. The signal has discreet values in X axis only which makes signal half continuous and cannot be correctly represented digitally, hence quantization is carried out. Quantization is the process of mapping a larger set of values to smaller set. The amplitude values are either rounded up or down to the nearest predetermined value.

C. Frame blocking and windowing

A signal is considered to be stationary if its frequency or spectral components do not change over time. As human speech is built from a dictionary of phonemes, while for most of the phonemes the properties of speech remain invariant for a short period of time (~5-100ms). The signal behaves stationary for those time frames. In order to obtain frames speech signal is multiplied with windowing function. The important step of pre-processing and feature extraction is a spectral analysis of each frame.

D. Pre-emphasis technique

In pre-processing technique voice signal will be enhanced. As input signal will be having lower energy by increasing the amplitude of high frequency bands and decreasing the amplitude of lower frequency bands, it increases the energy of signal to higher rate.

E. End point detection

End point detection detects the starting point and ending point of a given speech signal, so that the pure voice signal and noise signal can be distinguished. It eliminates the interference to improve the performance of voice recognition system; the amount of data collection is reduced in the signal so that processing time is reduced.

F. Deriving spectral envelope

In this project, spectral envelope is derived using periodogram and interpolation techniques. Periodogram is mainly used for spectral analysis. The spectral analysis measures the magnitude of an input signal versus frequency, as it measures the power of spectrum of known and unknown signals.

Interpolation constructs new data points within the range of discreet set of known data points. Cubic interpolation method gives an interpolating polynomial that is smoother and has smaller error than some other interpolating polynomials such Lagrange polynomial and Newton polynomial.

G. Deriving glottal excitation waveform

The glottal excitation is generated by vibrating vocal folds. Spectral tilt originates from the glottal excitation. The spectral tilt is used to describe slope of the power spectral density. Smooth closing of the vocal folds results in a large spectral tilt and more abrupt closing gives a glottal flow with a small spectral tilt.

IV REQUIREMENTS AND SPECIFICATION

A. Hardware specifications:

The hardware requirements are as follows:

- USB Headphone + MIC

B. Software specifications:

The software requirements are as follows:

- Operating system : Windows 8.1
- Coding language : MATLAB
- Tool : MATLAB R2017a, Praat

C. Implementation

This chapter gives information about the implementation of the proposed system.

Software tool:

MATLAB R2017a the software tool used for the implementation of the proposed system.

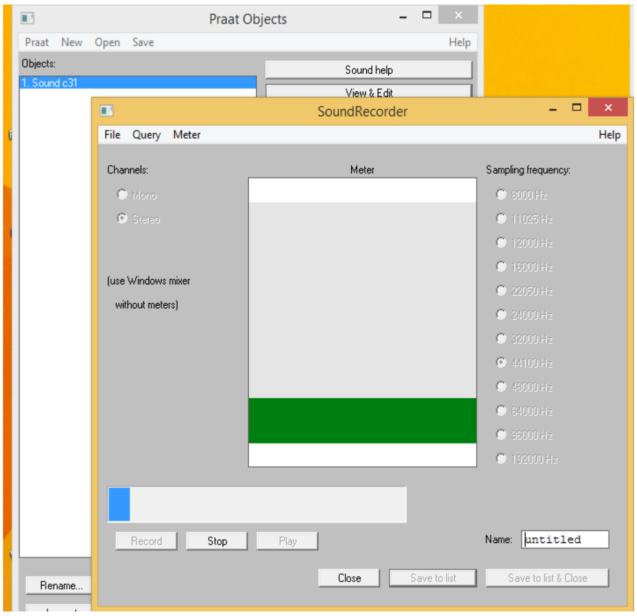
MALTAB R2017a:

MATLAB (Matrix Laboratory) is multi-paradigm numerical computing environment. A proprietary programming language developed by Math works, MATLAB allows matrix manipulations plotting of functions and data, implementation of algorithm, creation of user interface, interfacing with programs written in other languages, including C, C++, C#, JAVA, Fortran and Python.

Results and Discussions

This chapter gives the overall results and conclusion of the proposed system.

Results



Channels: ● Stereo

(use Windows mixer without meters)

Sampling frequency: ● 44100 Hz

Name: untitled

Figure 3: Shows recorded input speech signal

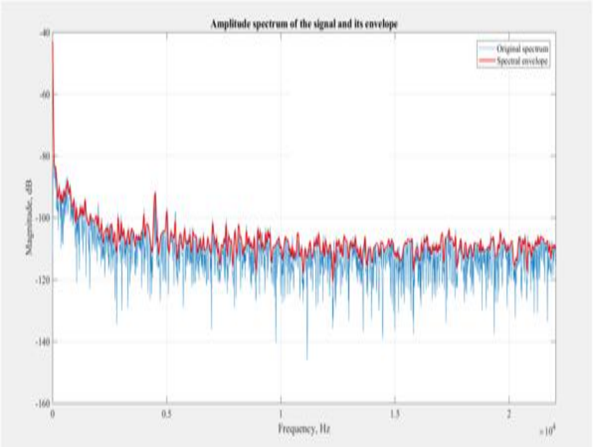
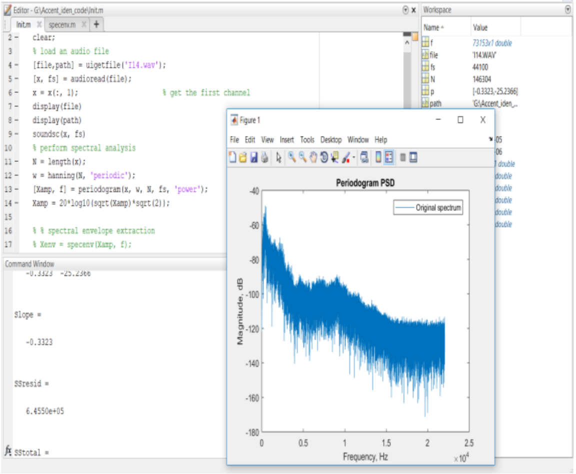


Figure 4: Shows the spectral envelope of the input signal

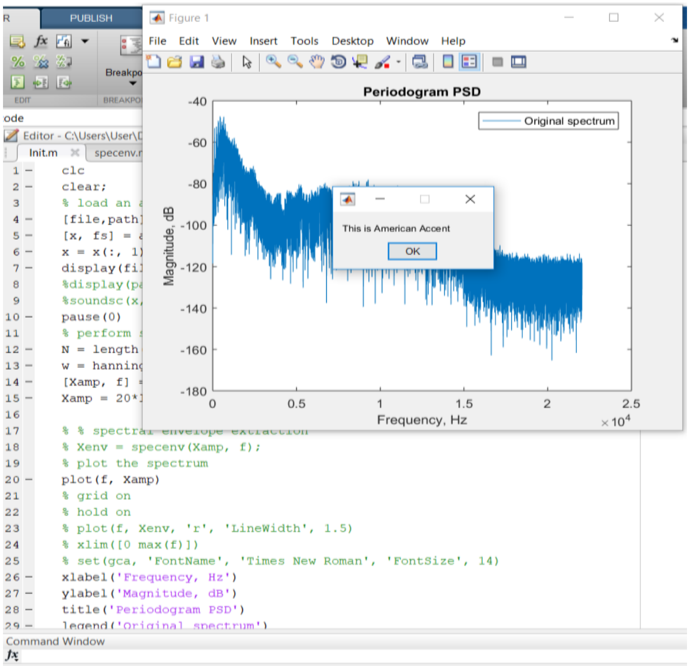


```
clear;
% load an audio file
[file,path] = uigetfile('I14.wav');
[x, fs] = audioread(file);
x = x(:, 1);                    % get the first channel
display(file)
display(path)
soundsc(x, fs)
% perform spectral analysis
N = length(x);
w = hanning(N, 'periodic');
[Xamp, f] = periodogram(x, w, N, fs, 'power');
Xamp = 20*log10(sqrt(Xamp)*sqrt(2));

% % spectral envelope extraction
% Xenv = specenv(Xamp, f);
```

Name	Value
f	73153x1 double
file	'114.WAV'
fs	44100
N	146304
p	[-0.3323,-25.2366]
path	'G:\Accent_iden_...

```
  -0.3323  -25.2366


Slope =

  -0.3323


SSresid =

  6.4550e+05


SStotal =
```

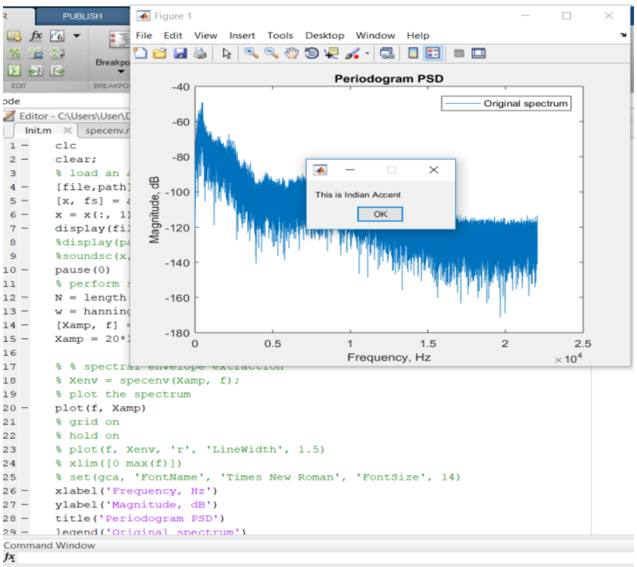
Figure 5: Shows the slope of the spectral envelope



This is American Accent

```
clc
clear;
% load an a
[file,path]
[x, fs] = a
x = x(:, 1)
display(fi
%display(pa
%soundsc(x,
pause(0)
% perform s
N = length
w = hanning
[Xamp, f]
Xamp = 20*l

% % spectral envelope extraction
% Xenv = specenv(Xamp, f);
% plot the spectrum
plot(f, Xamp)
% grid on
% hold on
% plot(f, Xenv, 'r', 'LineWidth', 1.5)
% xlim([0 max(f)])
% set(gca, 'FontName', 'Times New Roman', 'FontSize', 14)
xlabel('Frequency, Hz')
ylabel('Magnitude, dB')
title('Periodogram PSD')
legend('Original spectrum')
```

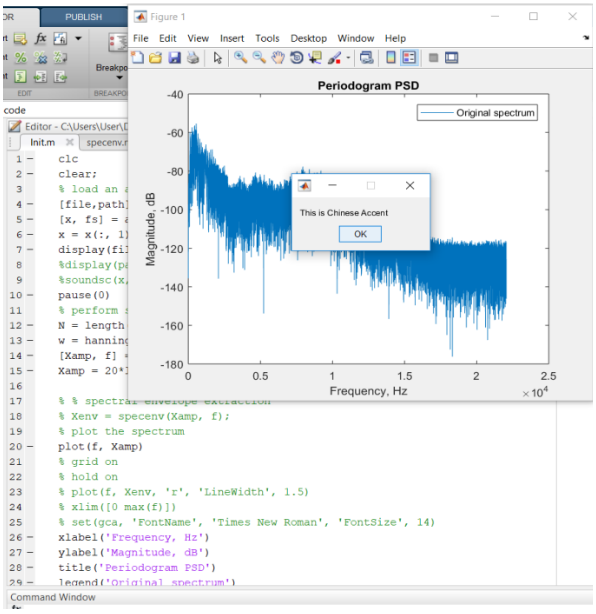
Figure 6: Shows the accent identified as American accent



This is Indian Accent

```
clc
clear;
% load an a
[file,path]
[x, fs] = a
x = x(:, 1)
display(fi
%display(pa
%soundsc(x,
pause(0)
% perform s
N = length
w = hanning
[Xamp, f]
Xamp = 20*l

% % spectral envelope extraction
% Xenv = specenv(Xamp, f);
% plot the spectrum
plot(f, Xamp)
% grid on
% hold on
% plot(f, Xenv, 'r', 'LineWidth', 1.5)
% xlim([0 max(f)])
% set(gca, 'FontName', 'Times New Roman', 'FontSize', 14)
xlabel('Frequency, Hz')
ylabel('Magnitude, dB')
title('Periodogram PSD')
legend('Original spectrum')
```

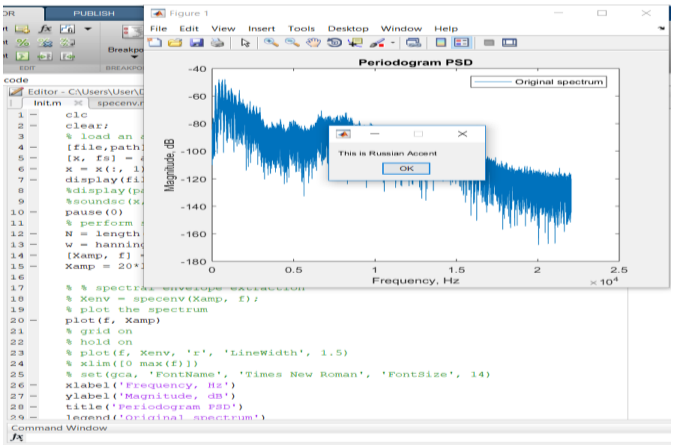
Figure 7: Shows the accent identified as Indian accent



This is Chinese Accent

```
clc
clear;
% load an a
[file,path]
[x, fs] = a
x = x(:, 1)
display(fi
%display(pa
%soundsc(x,
pause(0)
% perform s
N = length
w = hanning
[Xamp, f]
Xamp = 20*l

% % spectral envelope extraction
% Xenv = specenv(Xamp, f);
% plot the spectrum
plot(f, Xamp)
% grid on
% hold on
% plot(f, Xenv, 'r', 'LineWidth', 1.5)
% xlim([0 max(f)])
% set(gca, 'FontName', 'Times New Roman', 'FontSize', 14)
xlabel('Frequency, Hz')
ylabel('Magnitude, dB')
title('Periodogram PSD')
legend('Original spectrum')
```

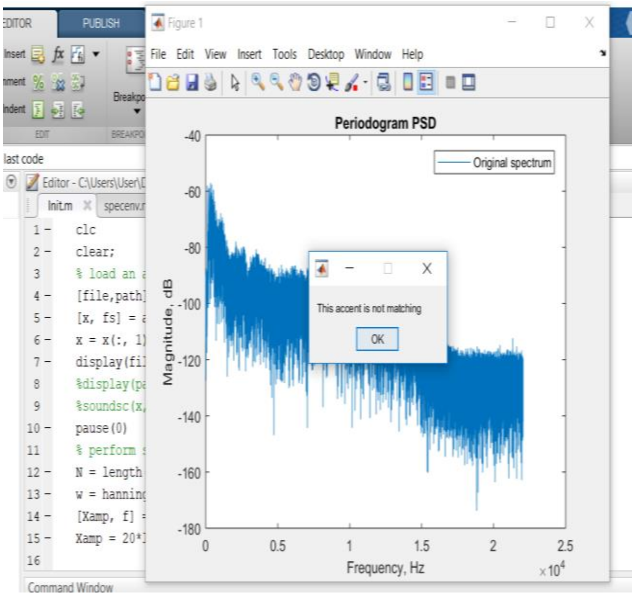
Figure 8: Shows the accent identified as Chinese accent



This is Russian Accent

```
clc
clear;
% load an a
[file,path]
[x, fs] = a
x = x(:, 1)
display(fi
%display(pa
%soundsc(x,
pause(0)
% perform s
N = length
w = hanning
[Xamp, f]
Xamp = 20*l

% % spectral envelope extraction
% Xenv = specenv(Xamp, f);
% plot the spectrum
plot(f, Xamp)
% grid on
% hold on
% plot(f, Xenv, 'r', 'LineWidth', 1.5)
% xlim([0 max(f)])
% set(gca, 'FontName', 'Times New Roman', 'FontSize', 14)
xlabel('Frequency, Hz')
ylabel('Magnitude, dB')
title('Periodogram PSD')
legend('Original spectrum')
```

Figure 9: Shows the accent identified as Russian accent



```
clc
clear;
% load an a
[file,path]
[x, fs] = a
x = x(:, 1)
display(fil
%display(pa
%soundsc(x,
pause(0)
% perform s
N = length
w = hanning
[Xamp, f] =
Xamp = 20*l

```

This accent is not matching

Figure 10: Shows that the accent is not matched

CONCLUSION

Accent identification is the first step in accent conversion. The accuracy of the conversion system is determined by the peruse identification. Hence the outcome of this project will lead building a robust accent conversion system.

ACKNOWLEDGEMENT

Authors would like to thank one and all for giving the proper guidance and also special thanks to Dr. Ajish K Abraham, professor Dept. of Electronics AIISH, Mysuru gor sharing their knowledge and guiding us throughout the work. My kind gratitude for Mrs. Sunitha R for her suggestions.

REFERENCE

1. Dejan Stantic and Jun Jo, “Accent Identification by Clustering and Scoring Formants,” in World Academy of Science, Engineering and Technology, International Journal of Computer and System Engineering Vol:6, No:3, 2012.
2. T. Chen, C. Huang, E. Chang and J. Wang, “Automatic accent identification using Gaussian mixture models”. *In proceedings of the IEEE Workshop on Automatic speech recognition, pages* 343-346, 2001.
3. Wieling, Martijin, Jelke Bloem, Kaitlin Mignella, Mona Timmermeister and John Nerbonne. 2014. Measuring foreign accent strength in English. Validating Levenshtein distance as a measure. *Language Dynamics and Change* 4(2). 253-263.
4. T. Schultz and A. Waibel, “Language Independent And Language Adaptive Acoustic Modeling For Speech Recognition”, *Speech Communication,* Volume 35, Issue 1-2, pp 31-51, August 2001
5. Emma jokinen and paavo alku, “Estimating the spectral tilt of the glottal source from telephone speech using a deep neural network”, the journal of the acoustical society of America 141, EL327 (2017).
6. “Automatically identifying characteristic features of non-native English accents” Jelke Bloem, Martijin Wieling and John Nerbonne, *The future of dialects,* 155-173, 2016.
7. Carlos Teixeira, Isabel Trancoso and Antonio Serralherio, “Accent Identification” in *Proc.* ICSLP’96, vol.3, pp. 1784-1787,1996.
8. Shri Lekha and Ashish Chopra, “Comparative study of speech parameterization technique”, volume 6, Number2, pp 133-135, July-December 2013.
9. Michel lutter, “Speech production”, January 2015.
10. Julian heuser, “Preprocessing”, December 2014.