Analysis of Dynamic Filtering Properties for Vocal-Tract during Bangla Vowel and Vowel-Consonant-Vowel Sequence Production

by

Sathi Rani Mitra

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Engineering in the Department of Electrical and Electronic Engineering

DEPARTMENT OF ELECTRICAL AND ELECTROINC ENGINEERING

KHULNA UNIVERSITY OF ENGINEERING & TECHNOLOGY

KHULNA-9203, BANGLADESH

DECEMBER, 2016.

Declaration

This is to certify that the thesis work entitled "**Analysis of Dynamic Filtering Properties for Vocal-Tract during Bangla Vowel and Vowel-Consonant-Vowel Sequence Production"** has been carried out by **Sathi Rani Mitra** in the Department of Electrical and Electronic Engineering, Khulna University of Engineering & Technology (KUET), Khulna, Bangladesh. The above thesis work or any part of this work has not been submitted anywhere for the award of any degree or diploma.

-------------------------------	------------------------------
Signature of Supervisor	Signature of Candidate

Approval

This is to certify that the thesis work submitted by **Sathi Rani Mitra** entitled *"Analysis of Dynamic Filtering Properties for Vocal-Tract during Bangla Vowel and Vowel-Consonant-Vowel Sequence Production"* has been approved by the board of examiners for the partial fulfillment of the requirements for the degree of **Masters of Science in Engineering (M. Sc. Eng.)** in the Department of **Electrical and Electronic Engineering**, Khulna University of Engineering & Technology (KUET), Khulna, Bangladesh in December 2016.

BOARD OF EXAMINERS

1. ________________________ **Chairman (Supervisor)**

Dr. Md. Mahbub Hasan
Associate Professor
Department of Electrical and Electronic Engineering
Khulna University of Engineering & Technology (KUET)
Khulna, Bangladesh.

2. ________________________ **Member**

Head of the Department
Department of Electrical and Electronic Engineering
Khulna University of Engineering & Technology (KUET)
Khulna, Bangladesh.

3. ________________________ **Member**

Prof. Dr. Md. Rafiqul Islam
Department of Electrical and Electronic Engineering
Khulna University of Engineering & Technology (KUET)
Khulna, Bangladesh.

4. ________________________ **Member**

Prof. Dr. Mohammad Shaifur Rahman
Department of Electrical and Electronic Engineering
Khulna University of Engineering & Technology (KUET)
Khulna, Bangladesh.

5. ________________________

Prof. Dr. Mohammed Imamul Hassan Bhuiyan **Member (External)**
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering & Technology (BUET)
Dhaka, Bangladesh.

ABSTRACT

This thesis deals with the acoustic analysis of dynamic filtering properties for vocal-tract and the perceptional space of Bangla vowel. Here we capture Bangla vowel, vowel-consonant-vowel speech signal from native Bangla Speakers. Acoustic study of dynamic-shape of vocal-tract for Bangla vowels and vowel-consonant-vowel (VCV) sequences are carried out based on dispersion and cross-correlation of linear predictive coding (LPC) filtering coefficients and short-time Fourier transform (STFT) of formant trajectory. The standard deviation of LPC filtering coefficients and transitional energy of formant trajectories indicate the dynamics of vocal-tract. The consonantal-constriction in VCV sequence accelerates the vocal-tract transitional nature and the transitional nature yields lower-valued cross-correlation with more stable vowels. Fourier transform technique is utilized to determine the cross-correlation of two unequal length LPC trajectories. In the domain of vowel perception, speaker invariant principal components are determined from the multidimensional perceptual space for comparatively stable Bangla vowels. Linguistic content base multidimensional vector space can be formed using the formant frequencies and its dispersion related statistical moments and the well reorganization-able vowels may maintain a separable distance in the multidimensional space. In this work, perceptual vectors and their associated energies are evaluated by determining the principal components of the multidimensional space. Speaker invariance is numerically evaluated using the correlation with principal components of native Bangla speakers. In this research, an algorithm is also proposed for the determination of energy based on the redundancies of the canonical variates for the inter-speaker vowel consistency.

Acknowledgements

First of all, I am thanking and expressing my immense gratefulness to the almighty GOD for HIS kindness and grace to us for the successful completion of this M. Sc thesis work. With deep sense gratitude I express my sincere thanks to my esteemed and worthy supervisor, **Dr. Md. Mahbub Hasan**, Associate Professor, Department of Electrical and Electronic Engineering (EEE), Khulna University of Engineering & Technology (KUET), Bangladesh, for his valuable guidance in carrying out this work under his effective supervision, encouragement, enlightenment and cooperation. His feedback and editorial comments were also invaluable throughout the entire progress of this thesis work titled as **"Analysis of Dynamic Filtering Properties for Vocal-Tract during Bangla Vowel and Vowel-Consonant-Vowel Sequence Production"**. The systematic and articulate approach of my supervisor in analyzing a problem, his persistent enthusiasm and overall presence will be always recalled by me in all of my future activities.

Then I would take this opportunity to express my deepest appreciation to **Prof. Dr. Md. Shahjahan**, Head of the department of EEE, KUET, for providing departmental facilities. I am also thankful to all the concerned teachers/alumni of the department for their full cooperation and help.

I wish to complement to all teachers of the Department of Electrical and Electronic Engineering, KUET for their co-operations and relentless encouragements in various ways throughout this work.

This acknowledgement would be incomplete if I do not mention the emotional support and blessings provided by my friends. I had a pleasant enjoyable and fruitful company with them. My greatest thanks are to all who wished me success especially my parents whose support and care makes me stay on earth.

December 2016

Sathi Rani Mitra

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Abbreviations

VV	Vowel-Vowel
VCV	Vowel-Consonant-Vowel
LPC	Linear Predictive Coding
STFT	Short-time Fourier Transform
MFCC	Mel-frequency Cepstral Coefficient

Dedicated to my beloved Parents...

CHAPTER 1

INTRODUCTION

Chapter Outlines

- Introduction
- Literature review
- Objectives of the thesis work
- Layout of the thesis work

1.1 Introduction

Obtaining knowledge by just listening to sounds is a unique feature. It has become a significant objective in the evolvement of species. Most of the animals are not only equipped with the means to capture information from the rich acoustical content of the environment, although they have aptitude to generate sounds to interact with the environment. Human have gone one step more, they have fairly advanced mechanisms that enable communication within the classes by very ideological communication-rules using speech.

Speech consists of the variety pattern of sounds, which is a form of human communication. In today's world, speech processing is becoming more and more popular due to increase in the use of human command dependent devices. Research on speech signals is important in many major research fields, such as speech recognition, linguistics, speech to text transformation, security, mobile communications and many other purposes.

Speech sounds are produced by air pressure vibrations of glottal pulse, which is controlled by the dynamic-shape of vocal-tract and nasal cavities while passing through lips and nasal airways. In modulation perspective, the glottal pulse is modulated by the resonance conditions of the vocal-tract and nasal cavities [1]. Vocal-tract can be approximated as a non-uniform flexible acoustic tube with time-varying cross-sectional areas, which is terminated by vocal-cord at one end and lips and/ or nose at the other [2]-[3]. The vocal-tract shape is changed temporally and spatially according to the speaker's intensions, physical and emotional states [4]-[6]. Very often, the vocal tract is only considered as a filter independent from the glottal source, enhancing the energy of the voice around resonance frequencies of the vocal tract. Like other vowels, vocal-tract shape remains quasi-stationary for Bangla vowel. But in the case of consonantal-constriction, vocal-tract shape is not stationary [7].

Linear predictive coding (LPC) [7]-[8] is an important tool in digital signal processing for predicting the present speech sample as a linear combination of the previous samples. The linear predictive coding filtering coefficients reveal the dynamic-shape of vocal-tract. As sounds are produced by the dynamics of vocal-tract shape, the LPC filtering coefficients will be also time dependent. This analysis will be carried to investigate vocal-tract dynamics for Bangla vowel and VCV (vowel-consonant-vowel) sequences. This research would be important reports as Bangla is the most spoken language in Bangladesh and second most spoken language in India, with about 215 million native and about 233 million total speakers worldwide [9].

This work will also describe speaker invariant perceptual space which of comparatively stationary Bangla vowel by principal component analysis. Speech perception involves the mapping of speech acoustic signals onto linguistic messages (e.g. phonemes, distinctive features, syllabus, words, phrases) [10]. Acoustic cues are captured and stored in sensory memory and mapped onto linguistic information. Perceptual space of vowels is visualized as multidimensional as representation based on the first two formants. If the multi-dimensional acoustical vector consists only the vocal-tract related acoustical features such as formant frequencies, then it will be linguistic content related perception vector [11]. This study also investigates the canonical correlation between the acoustic content of two subjects using canonical correlation [12]. From the vowel consistency it is important to know how much variance or energy of the multidimensional acoustical space of speaker related can be explained by the canonical correlation. This thesis compares the vocal-tract dynamics for Bangla vowel and vowel-consonant-vowel (VCV) sequences and evaluates the consistency of comparatively stable vowel perception spaces consider speaker invariant case.

1.2 Literature review

As part of a thesis, the literature review enables to demonstrate the knowledge of previous work in this research field and to situate own research in the context of this work. For this thesis work, some associated research studies are given below:

Brad H. Story represented that a vowel sequence as a time-dependent perturbation of the neutral vocal tract shape governed by coefficients of canonical deformation patterns. Consonants were modeled as superposition functions that can force specific portions of the vocal tract shape to be constricted or expanded, over a specific time course [13].

Öhman's spectrographic analysis of vowel-consonant-vowel (VCV) sequence utterances results suggested that vowels and consonants are generated by two parallel events and both events are taken place in vocal-tract [14].Thus, a VCV is considered to be produced as a vowel-vowel (VV) sequence upon which a consonantal vocal-tract gesture is superimposed.

B. Yegnanarayana and R. N. J. Veldhuis described the vocal-tract system, which is characterized by its formant parameters, which are extracted from the analysis segments. Because the segments are always at the same relative position in each pitch period, in voiced speech the extracted formants are consistent across successive pitch periods [15].

Research on Bangla speech analysis, recognition and synthesis is in a preliminary stage of development. A. Hasnat , Md. R. Karim, Md. S. Rahman and Md. Z. Iqbal tried to present a technique of recognition of Bangla spoken letters. This has led to the excitation of spectral parameters for the production model in order to produce different Bangla vowel sounds [16].

S. A. Hossain, M. L. Rahman, and F. Ahmed tried to characterization of Bangla vowels which determines the spectral properties of Bangla vowels for efficient synthesis as well as recognition of Bangla vowels. Acoustic space of Bangla Vowel based on articulatory properties of vocal tract plays a significant role in Bangla speech synthesis and recognition [17]-[18].

A. K. Paul, D. Das, and Md. M. Kamal focused on recognition of vowel and different phonemes based on spectral analysis, autocorrelation, LPC coefficients, and mel-Frequency Cepstral Coefficient [19].

To the best of our knowledge, no research has been done on Bangla vowel and VCV sequence related vocal-tract dynamics and their comparisons, which was our motivation to work on this area.

Over the past 60 years, researchers in speech perception have focused on the mapping between the acoustic properties (i.e. formant frequencies) with perception cues, which is based on first and second formant [20]-[22].

T. M. Nearey work reviewed the problem of perceptual invariance in vowel perception. Here acoustic space is clustered in regions for different vowels. Although a considerable variability in mapping between the clustered regions in acoustic space to vowels have been revealed, clustering is considered as one of the best vowel perception basis [10], [23].

Variability may be the results of characteristics of the speaker, speaking rates, speaking styles, and noise, which are not considered in perception space. Despite this variability, vowels retain their perceived identity in perception space [24]. The listeners perceived linguistic information by identifying the clustered regions in perception space and this model is known as categorical perception [10].

Formation of vowel perception vector is not uniquely defined and a number of reports have been published regarding the vowel perception vector [25]-[27]. In 1973, Matsumoto et. al. urged that fundamental frequency (f0) and first three formant frequencies (F1; F2; F3) were sufficient to account for most of the energy and cues for vowel perception [26]. Fundamental frequency (f0), word duration, age, and voice qualities based four-dimensional perceptual space is modeled by Walden and his group in 1978 [28]. In addition to fundamental (f0) and formant frequencies (F), more acoustical features such as jitter and shimmer were included in perception space by Kreiman et.al. [29].

Speaker-invariant or speaker-normalized perception space and its factors are very much important for automatic speech reorganization and cochlear implants for deaf persons [30]-[31]. Although investigators are trying to explore the relationship between linguistic content of speech and listener's perception, but it is not fully understood yet. Most of the studies regarding to vowel perception space related studies are mainly based on English language. Speaker-invariant Bangla vowel perceptual space and its factor analysis will be an important report to automatic Bangla vowel reorganization and the treatment of hearing-impairment Bengali people. Speaker-independent Bangla vowel perceptual space and its factor analysis and acoustical attributes between two subjects are not reported yet. In this work, invariance of Bangla vowel perceptual space has been discussed with principal components considering only the linguistic content as well as vocal-tract dynamics.

1.3 Scope & objectives of the thesis work

The goal of this research is to analyze the vocal-tract dynamics related to Bangla vowel and vowel-consonant-vowel (VCV) sequence. In this study, speaker invariant properties of vowel perceptual space are shown mathematically with the principal components. Next deals with experimentally evaluation of consistency of Bangla vowel perceptual space. Finally, an algorithm is proposed to measure of acoustic content of vowel perceptual spaces between the two subjects utilizing the redundancy of the canonical variates which refers the above mentioned explainable variance. The specific objectives of this thesis are stated below-

- To measure the dispersion of LPC filtering coefficients and formant transitional energy
- To determine the correlation coefficients using two unequal length LPC trajectories
- To study the speaker invariance of comparatively stable Bangla vowel perceptual space with principal components considering only the linguistic content.
- To determine Bangla vowel consistency or linguistic content similarity using the explainable variance of vowel perception spaces using canonical correlation.

1.4 Layout of the thesis work

In **Chapter 2**, fundamental concept of speech sound, speech production, speech perception and mathematical formulation of acoustical properties are described.

In **Chapter 3**, the result and discussion are obtained from the analysis of the acoustical properties. Firstly, the vocal-tract dynamics for Bangla vowels with and without consonantal constriction are studied using Short-time Fourier transform (STFT). Then speaker invariant principal component for Bangla vowel perceptual space is analyzed. Finally canonical correlation based vowel consistency of acoustical spaces is discussed.

In **Chapter 4,** the concluding remarks and recommendation for future work are provided.

CHAPTER 2

Theoretical Background and Mathematical Formulation

Chapter Outlines

- Introduction
- Speech production
- Speech perception
- Mathematical formulation
  - Acoustic property of vocal-tract
  - Formant
  - Vocal-tract transitional energy
  - Effect of vocal-tract dynamics on acoustical correlation
  - Speaker invariant principal component for vowel perceptual space consistency
  - Canonical correlation redundancy index for vowel consistency

2.1 Introduction

Speech processing contains lots of special concepts and terminology. To realize different speech synthesis, analysis methods and physics of speech production is important. The basic theory of these topics and mathematical formulation will be briefly discussed here.

2.2 Speech production

The purpose of speech is communication. Speech signal carries temporal information from speaker to listeners. Speech related information is the key features for voice activity detection, speech recognition, speech-to-text conversion and many other purposes [2]. Vocal organs produce human speech and these are presented in Figure 2.1.

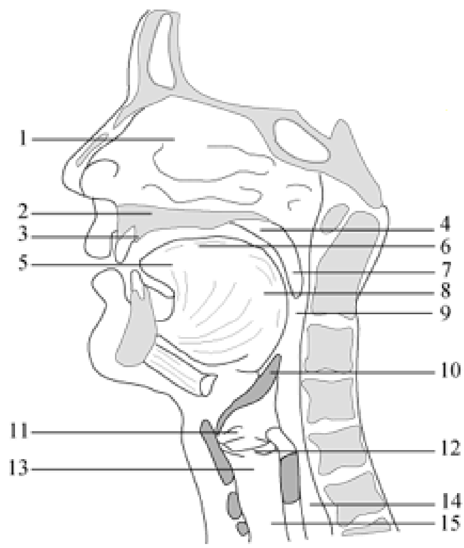


Figure 2.1: Human vocal organ [32]

(1) Nasal cavity, (2) Hard palate, (3) Alveolar ridge, (4) Soft palate (Velum), (5) Tip of the tongue (Apex), (6) Dorsum, (7) Uvula, (8) Radix, (9) Pharynx, (10) Epiglottis, (11) False vocal cords, (12) Vocal cords, (13) Larynx, (14) Esophagus, and (15) Trachea.

The oral cavity is one of the most important parts of the vocal-tract. The oral cavity consists of the palate, the tongue, the lips, the cheeks and the teeth. It has almost fixed dimensions, but its length may be changed slightly by raising or lowering the larynx at one end and soft palate at other end. The movements of its parts change the shape and size of oral cavity. The most flexible part of oral cavity is tongue, which can move forward, up and down and play main role in changing the shape of oral cavity. The lips control mouth opening and speech sound is radiated. Unlike the oral cavity, the nasal cavity has fixed dimensions and shape. Its length is about 12cm and volume $60cm^3$. The air stream to the nasal cavity is controlled by the soft palate. From the technical point of view, the vocal system may be treated mathematically as a single acoustic tube closed at one end (the glottis) and open at other (the lips) for the purposes of calculating the resonances of the vocal tract.

Speech sounds result from a combination of a source of sound energy (the larynx) modulated by a time-varying transfer function filter (vocal articulators) determined by the shape and size of the vocal tract. This results in a shape spectrum with broadband energy peaks. This model is known as the source-filter model of speech production that is shown in Figure 2.2.

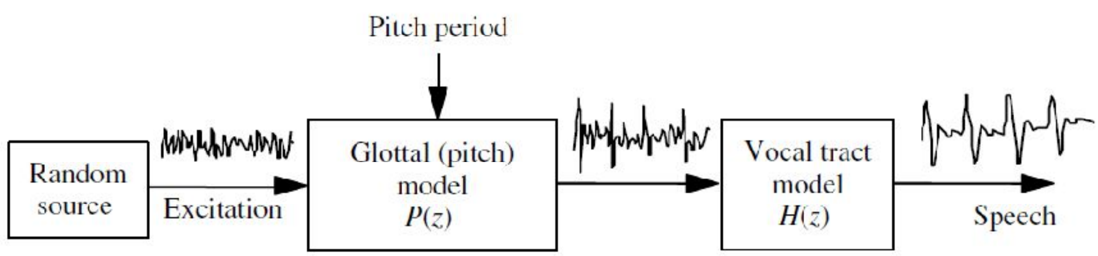


Figure 2.2: A source-filter model of speech production [2]

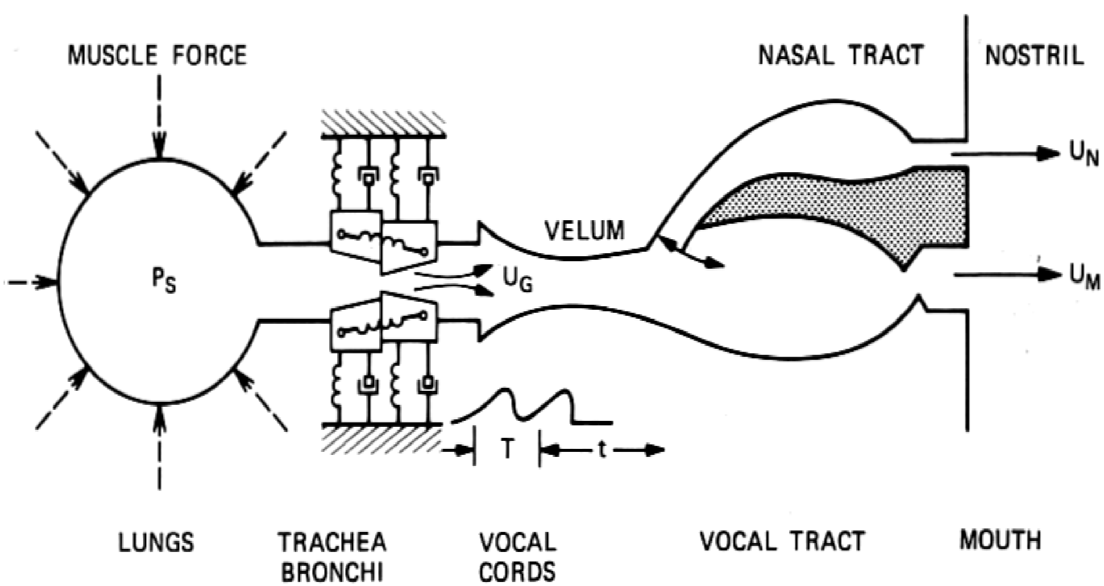


Figure 2.3: Schematic diagram of Speech production [33]

The source of the non-uniform flexible acoustic tube is airflow pressure U_G Vocal-tract which is shown in Fig. 2.3 mechanically. The lung volume (random noise) is charged to subglottal lung pressure P_s, which is applied via the trachea-bronchi network, to the glottal (vocal-cord opening) impedance Z_g. This nonlinear glottal impedance depends upon the glottal flow and area A_g, which in turn depend upon the self-oscillating properties of the vocal cord. The output of the process is glottal pulse U_G. The vocal cord remains open during vowels and stops consonants production and let the glottal pulse into vocal-tract. In case of unvoice consonant, the vocal-cord remains closed.

The vowel sound is the shaped form of glottal pulse by the resonance conditions of quasi-stationary vocal-tract [34]. Vocal-tract resonance states move to transient states at imposing and releasing of consonantal constrictions and the duration is short comparing to quasi-stationary period. The acoustic phenomena, which sustains for short duration, represent the consonantal sound to the listeners [15], [35]. Vowels and consonants are produced by two parallel events and both events occurred in vocal-tract. Thus, a VCV sequence is formed when a consonantal vocal-tract structure is superimposed upon a vowel-vowel (VV) sequence.

2.3 Speech perception

This work also deals with the perceptual invariance of comparatively stable acoustic event vowel perception. Speech perception (SP) most commonly refers to the perceptual mapping from the highly variable acoustic speech signal to a linguistic representation, whether it phonemes, diaphones, syllables, or words [10]. Here, speech perception is based on the classification of the acoustic properties of the presented speech signal according to principal component and canonical variates. Human voice is a very important acoustic event as it contains the features of glottal pulse and vocal-tract dynamics [34], [36]. Personal identity and linguistic information are related to glottal pulse and the vocal-tract dynamics respectively. Vocal-tract is modeled by time-varying filter as its shape is dynamically changed according to the linguistic information. The dynamic filtering is usually represented by linear predictive coding (LPC) filtering coefficient or formant frequencies. Formant frequencies are the peak spectral response of the time-varying vocal-tract filter. Considering these facts, the formant trajectories are considered as the representative of linguistic information. Formant related properties such as mean formant frequency, formant dispersion and bandwidth are utilized as phonetic judgments cues for vowel and consonantal constriction [37]-[38]. The formant trajectories are also called as "acoustic features" of speech.

People have acquired or have been born with slightly different structured vocal-tract that incorporates perturbation in sound production. These perturbations make the speech perception difficult as the perception strongly depends on the regularities of acoustic content. Regularities of multidimensional acoustic content can be evaluated by the inter-subject consistency of the multidimensional acoustical space. The canonical correlation redundancy index (explainable variance) is a measurement of regularity of acoustical content between the two subjects.

2.4 Mathematical formulation

This section describes the mathematical model of vocal-tract dynamics, speaker invariant properties of vowel perceptual space and vowel acoustic consistency based on canonical correlation. In order to analyze of vocal-tract dynamics, the dispersion of LPC filtering coefficients and formant transitional energy for Bangla vowels with and without consonantal constriction is focused here. Short-time Fourier transform (STFT) is used to evaluate the spectral components. The correlation coefficients are determined using Fourier coefficient of the two unequal length LPC trajectories. Speaker invariant properties of vowel perceptual space have been shown mathematically with the help of principal components. In case of vowel acoustical consistency, a tool has also been proposed for determination how much energy of one canonical variate can be predicted by its counterpart partner based on the redundancy index. In this section, the expression of canonical correlation redundancy index has also been derived.

2.4.1 Acoustic property of vocal-tract

Linear predictive analysis is widely used in speech research due to its ability to produce accurate estimates of the spectral characteristics of speech and its small computational cost. The spectral qualities of sounds can be represented efficiently by linear prediction, using only a low number of parameters. In this work, linear predictive analysis is used to extract formant frequencies from the impulse responses obtained by the vocal tract model, as well as the recorded speech signals. The idea of linear prediction lies in the source-filter model of the speech production and is described well in the reference [9].

In speech production system, vocal-tract acts as a time-dependent acoustical filter which is excited by glottal pulse and output is speech. The time-dependent filtering characteristics can be expressed by the LPC filtering coefficients. Mathematically, vocal-tract characteristics or transfer function can be expressed by

$$H(z)=\frac{S(z)}{U(z)}=\frac{G}{1-\sum_{k=1}^{p}\alpha_k z^{-k}} \qquad (2.1)$$

where, $S(z)$ is the speech in z domain, $U(z)$ is the glottal pulse in z domain, G is the gain factor, p is the linear prediction order and α_k is the LPC filtering coefficient. The speech samples $s(n)$ can be written by

$$s(n)=\sum_{k=1}^{p}\alpha_k s(n-k)+Gu(n) \qquad (2.2)$$

The linear prediction with coefficients α_k and order p is a system whose output is

$$\bar{s}(n)=\sum_{k=1}^{p}\alpha_k s(n-k) \qquad (2.3)$$

The system function of this linear predictor in z domain is

$$P(z)=\sum_{k=1}^{p}\alpha_k z^{-k} \qquad (2.4)$$

The prediction error is defined as

$$\begin{aligned} e(n) &= s(n)-\bar{s}(n) \\ &= s(n)-\sum_{k=1}^{p}\alpha_k s(n-k) \end{aligned} \qquad (2.5)$$

The error term consists of impulsive glottal pulse with other noises. The coefficients α_k are evaluated by minimizing the predication error $e(n)$. The coefficients α_k are conventionally named as LPC1, LPC2,.....,LPCn [9].

The dispersion (standard deviation μ) of time-dependent LPC filtering coefficients can be determined by

$$\mu = \sqrt{\frac{\sum_{n=1}^{N}(LPC_n - \overline{LPC})^2}{N}} \tag{2.6}$$

here, N is the total number of samples, $\overline{LPC}$ the mean value of samples and LPC_n the nth value. The dynamic-nature of vocal-tract also changes its resonance frequencies. The resonance frequencies can be estimated from the peak spectral response obtained by using (2.1). The resonance frequencies are also called formant frequencies.

2.4.2 Formant

Vowels are produced with a relatively open vocal tract and the airstream is not severely impeded. The resulting acoustic signal is therefore relatively loud. In addition, vowels are usually produced with vocal fold vibration. The primary acoustic characteristic of vowels is the location of the formant frequencies. With vowels, the frequencies of the formants determine which vowel you hear and, in general, are responsible for the differences in quality among different periodic sounds. At any one point in time (as with spectra) there may be any number of formants, but for speech the most informative are the first three, appropriately referred to as F1, F2, and F3. For a given speaker or for a group of speakers with the same vocal-tract length, each vowel is associated with a distinct acoustic formant frequency pattern. As vowel quality changes, the frequency of the third formant does not change nearly as much as that of F1 and F2, with the possible exception of the vowels [ই, ঈ], for which F3 is quite high.

2.4.3 Vocal-tract transitional energy

The changing tendency of formant positions also indicates the vocal-tract dynamics during uttering period. In this work, the vocal-tract transitional energy is formulated by local spectral energies of formant trajectories. As the formant trajectories are non-stationary type, short-time Fourier transform (STFT) is used to obtain transitional energy. The discrete STFT of the formant trajectory can be expressed as:

$$A_v[n,\Omega] = \sum_{m=1}^{N} F_v[n+m]w[m]e^{-j\Omega m} \tag{2.7}$$

where, $A_v[n,\Omega]$ is the STFT coefficient and F_v the v^{th} order formant trajectory. $w[m]$ the window function. We are interested in transitional energy and $\Omega = 0$ frequency component is neglected here. The total transitional energy of the v^{th} formant trajectory becomes:

$$E_v = \sum_{n=1}^{N} \sum_{\Omega} A_v{}^2[n,\Omega] \tag{2.8}$$

Considering the all formant trajectories, the average transitional energy can be expressed as:

$$E = \frac{1}{N} \sum_{v} E_v \tag{2.9}$$

2.4.4 Effect of vocal-tract dynamics on acoustical correlation

LPC filtering coefficients refer to acoustic property of vocal-tract at the time of sound production and the correlations among the LPC trajectories indicate the similarity among sounds or phonemes. Transitional nature of LPC trajectories are the results of vocal-tract dynamics constriction. The consonantal-transition of vocal-tract will yield lower valued correlation coefficient with vowel. The cross-correlation coefficients between two sounds have been modeled by the summation of zero-lag cross-correlation of LPC trajectory of the two sounds. Mathematically,

$$P_{XY} = \frac{1}{P} \sum_{i}^{P} X_C\left(LPC_{Xi}, LPC_{Yi}, 0\right) \tag{2.10}$$

Here, P_{XY} is cross-correlation coefficient between X and Y sounds, X_C is the zero-lag cross-correlation coefficient. Generally, the length of $LPC_{X,}$ and LPC_Y are different, then the zero-lag cross-correlation coefficient has been evaluated using Fourier coefficient and given in (2.11).

$$X_C = \frac{\sum F_x(w) F_y{}^*(w)}{\sqrt{\sum F_x(w)^2}\sqrt{\sum F_y(w)^2}} \tag{2.11}$$

here, $F_x(w)$ and $F_y(w)$ are the Fourier coefficients of LPC_X and LPC_Y respectively. The term $\sum F_x(w) F_y{}^*(w)$ is the spectral representation of zero-lag cross-correlation between LPC_X and LPC_Y as

$$XCorr\ (0) = IFT\left[F_x(w)F_y{}^*(w)\right] \\ = \sum_w F_x(w)F_y{}^*(w) \tag{2.12}$$

By using (2.11), the zero-lag cross correlation between two different length sequence can be computed.

2.4.5 Speaker invariant principal component for vowel perceptual space consistency

From each Bangla vowel sound, nine-dimensional acoustical vector,

$$[V = Fi(1 \le i \le 5), F51 = disp(F5 - F1), F43 = disp(F4 - F3), F53 = disp(F5 - F3), F54 = disp(F5 - F4)]$$

has been formed using mean value of the whole utterance duration. Here, the glottal source properties: jitter, shimmer, f0 have not been considered, as we are interested in only linguistic content. The acoustical vectors of the vowels have been tabulated and the table represents nine-dimensional Bangla vowel perceptual space. Perception vector is the mathematical form of excitation to the listener to recognize the vowel. Portraying the vowel-vectors in lower dimension is convenient for clustering. The measured or observed variables of the perceptual space represent physical phenomena of vowel production, and dimensional reduction technique considering variance or information can be utilized on the multi-dimensional space [39]. The principal components are explained with the eigen-values and eigen-vectors. The eigen-vectors are considered as the coefficients or as the contributions for each observed variable in principal components. The principal components play prominent role in perceptual space clustering according to vowels. The energies and directions of the principal components can be evaluated by determining the eigen-values and eigen-vectors of the column-wise covariance matrix defined in equation (2.13), given by

$$R = E\left[V^T V\right] \tag{2.13}$$

where T and E refers transpose and expected operators. Mathematically, the eigen-value and eigen-vectors can be written by the following expression:

$$Ru = \lambda u \tag{2.14}$$

here u and λ are eigen-vector and eigen-value respectively. First principal component contains the major information about the perceptual space and the correlations among the first principal components of different perceptual spaces indicate inter-speaker consistency.

For the determination of correlation, first eigen-vector matrix, U considering different perceptual spaces has been formed by the following way:

$$U = \begin{vmatrix} u(1,F1) & u(2,F1)\cdots\cdots & u(N,F1) \\ u(1,F2) & u(2,F2)\cdots\cdots & u(N,F2) \\ u(1,F3) & u(2,F3)\cdots\cdots & u(N,F3) \\ u(1,F4) & u(2,F4)\cdots\cdots & u(N,F4) \\ u(1,F5) & u(2,F5)\cdots\cdots & u(N,F5) \\ u(1,F51) & u(2,F51)\cdots\cdots & u(N,F51) \\ u(1,F43) & u(2,F43)\cdots\cdots & u(N,F43) \\ u(1,F53) & u(2,F53)\cdots\cdots & u(N,F53) \\ u(1,F54) & u(2,F54)\cdots\cdots & u(N,F54) \end{vmatrix} \tag{2.15}$$

where, $u_{(i,m)}$ is the contribution of m-th $(m = F1, F2, F3, F4, F5, F51, F43, F53, F54)$ perception vector member for i-th speaker or perceptual space, and N the number of perceptual spaces. Inter-speaker invariance of first principal component can be computed by

$$c_1 = E\left[U^T U\right] \tag{2.16}$$

The average value of the correlation matrix refers the inter-speaker consistency of the Bangla perceptual space according to first principal component. Consistencies of other principal components can also be calculated similarly. As the perceptual vector is not well established and the speech collection environment is not unified, there is a strong possibility of making weak correlation with the corresponding principal components.

2.4.6 Canonical correlation redundancy index

For the redundancy analysis between two speakers the multidimensional acoustic vector, V is reduced to five dimension from nine dimension $([V = F_i(1 \le i \le 5)])$. The reasons of eliminating formant dispersion components is making the covariance matrix, $\left(C = E\left[V1^T V2\right]\right)$ well-conditioned as ill and well condition depends upon the ratio of number of members of the vector and the number of observation [40]. The major variational energies and directions can be determined from the matrices, $\left(C1 = E\left[V1^T V1\right]\right)$ and $\left(C2 = E\left[V2^T V2\right]\right)$ by evaluating larger eigen-values and associated eigen-vectors. In general physical system, the energy distributions matrix $\left([diag(\lambda_1)][u_1]^T\right)$ and $\left([diag(\lambda_2)][u_2]^T\right)$ are the important factors of the two vowel spaces. The terms $[u_i]$ and λ indicates the eigen-vector and eigen-value of the matrices. But, the inter-subject consistency does not depend upon the individual energies; it depends upon the shared or explained variance between the two vowel spaces. Considering these facts, the inter-speaker consistency can be evaluated by determining the shared or explained variance. The shared or explained variance can be determined using canonical correlation technique. The canonical correlation between the acoustic spaces searches the vectors $a_1 \in AS^5$ and $a_2 \in AS^5$ that maximize the following correlation:

$$\rho = \max_{a_1, a_2} corr(V_1 a_1, V_2 a_2) \tag{2.17}$$

where, AS is multidimensional acoustical space, ρ canonical correlation coefficient, (a_1, a_2) the maximal correlated vector pair. These are also called canonical variates for the corresponding spaces. Both speakers related acoustic spaces are five dimensions and we will get five correlation vector pairs with descending order of correlation coefficients $(\rho_1 \rangle \rho_2 \rangle \rho_3)$. The square value of ρ indicates the correlation between the canonical components or variants of a canonical pair. The correlation (ρ^2) is the eigen-value of the following equation system:

$$\begin{aligned} (C_{11}^{-1} C_{12} C_{22}^{-1} C_{21} - \mu I) a_1 = 0 \\ (C_{22}^{-1} C_{21} C_{11}^{-1} C_{12} - v I) a_2 = 0 \end{aligned} \tag{2.18}$$

where $C_{11} = E\lfloor V_1^T V_1 \rfloor$, $C_{22} = E\lfloor V_2^T V_2 \rfloor$ are the variance matrices, $C_{12} = E\lfloor V_1^T V_2 \rfloor$ and $C_{21} = E\lfloor V_2^T V_1 \rfloor$ are cross-variance matrices, and $\mu = v = \rho^2$. The eigen-values are same and equal to correlation of pair (a_1, a_2). The inter-speaker consistency of vowel space depends not only the strong correlations but also the proportion of explained variance or energy of the variants pair. Because, two low energy canonical components may have the higher correlation coefficients. Actually, the canonical correlation gives no information about the energy occupied by the canonical variants pair [41]. The amount of energy can be explained by the loading factor. The loading factor of canonical components can be written by

$$\begin{aligned} u = a_1^T C_{11} \\ v = a_2^T C_{22} \end{aligned} \tag{2.19}$$

where u and v are the loading factors.

In reference [41], a tool has been proposed for determining how much energy of one canonical variate can be predicted by its counterpart partner based on the redundancy index. The redundancy index can be defined by

$$\Re_k = \mu_k \frac{1}{5} \sum_{j=1}^{5} u^2{}_{j,k} \tag{2.20}$$

where, $\Re_k$ refers to the amount energy of first vowel space predicted by the kth canonical variants of the second vowel space, μ_k and $u_{j,k}$ are the eigen-value and loading factor, respectively. The total predicted energy by the second vowel space from the first vowel space can be calculated by taking sum of all the canonical components. The total redundancy index is considered as the speaker independent consistency. By considering the fact, the inter-speaker vowel space consistency can be written

$$VC = \sum_k \Re_k \tag{2.21}$$

CHAPTER 3

Results and Discussion

Chapter Outlines

- Sound capturing
- Software required
- Acoustic property of vocal-tract
- LPC dispersion comparison for Bangla vowel and VCV Sequences
- Transitional energy for Bangla vowel and VCV Sequences
- Correlation among acoustic property of Bangla vowel pairs and VCV sequence
- Perceptual space consistency for Bangla vowels by principal components
- Vowel consistency by canonical correlation

3.1 Sound Capturing

For the investigation of acoustical properties of vocal-tract dynamic shape and inter-speaker similarity of Bangla vowel perceptual space, we have captured ten Bangla vowel sounds (অ, আ, ই, ঈ, উ, ঊ, এ, ঐ, ও, ঔ) from fifteen male speakers. Also five VCV (আলো, আমি, ইতি, অনু, উচু) sounds are collected by different male speakers. The sounds are collected from male speaker of age 19 to 23 by using Praat software in communication engineering laboratory at the Department of Electrical and Electronic Engineering (EEE) in Khulna University of Engineering & Technology (KUET). Cosonic CT- 863 headphone is used for recording. The speech data was digitized at 44100 Hz sampling frequency and stored as wave format. Typical Bangla vowel waveform and its spectrogram are shown in Figure 3.1 and Figure 3.2.

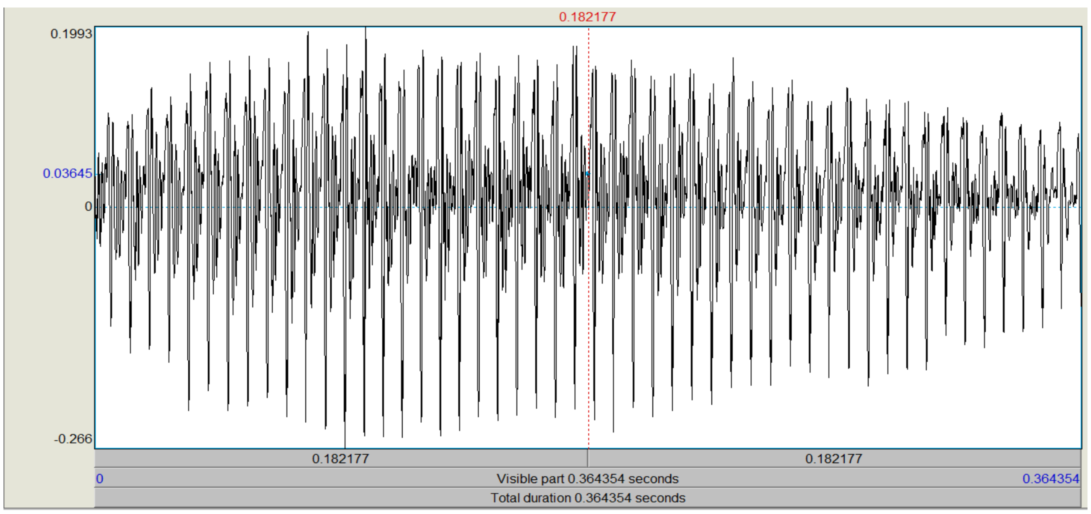


Figure 3.1: Bangla Vowel signal (অ)

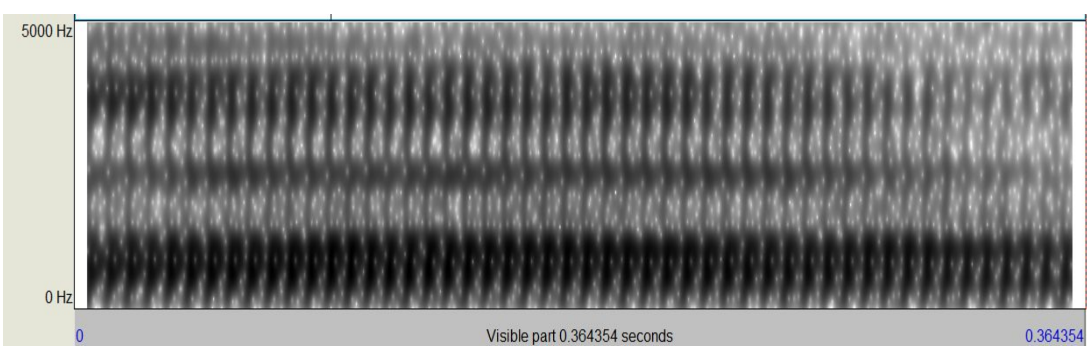


Figure 3.2: Bangla Vowel spectrogram (অ)

3.2 Software required

Here, we analyzed the speech using Praat analysis, Matlab and R softwares. We have evaluated the LPC and formant trajectories from the speech sounds using Praat. Standard deviation and correlation of LPC trajectories of vowel and VCV sequences are computed in Matlab. Transitional tendency of formant frequencies are also calculated in Matlab. The principal component of perceptual space is evaluated utilizing R. Vowel consistency based on canonical redundancy index is also determined using R.

3.3 Acoustic property of vocal-tract

The 16th order LPC filtering coefficients of Bangla vowels (অ, আ, ই, ঈ, উ, ঊ, এ, ঐ, ও, ঔ) and VCV sequences are determined every 0.005 second interval using window length 0.025 second. For quick and easy anticipation of 16 LPC trajectories, first three LPC trajectories for Bangla vowels and VCV sequences are shown in below Fig 3.3, Fig. 3.4 and Fig. 3.5. LPC trajectories calculated from experimentally captured speech consists of only the vocal-tract states related information not its energized source (glottal pulse). The LPC dispersion related comparisons between vowel and VCV sequence indicates the numerical comparison of vocal-tract dynamics. Figure 3.3 shows the LPC trajectories for Bangla vowel (অ). From this figure, we can see that vocal-tract shape remains quasi-stationary. For other Bangla vowels, the LPC trajectories are shown in appendix A.

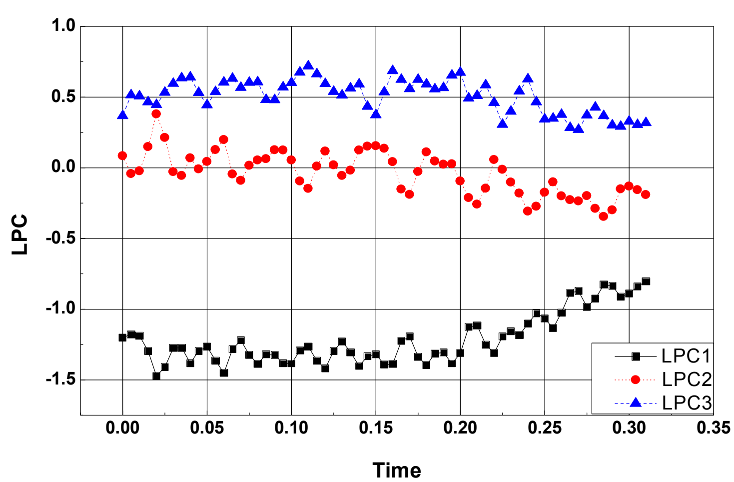


Figure 3.3: Evaluated first three LPC trajectories for Bangla vowel (অ)

On the other hand, from visual observation of Bangla VCV sequences, it is found that more transitions occur during VCV sequence productions. A VCV sequence is formed by a vowel-vowel sequence upon which a consonantal-constriction is superimposed and it can be explained by the LPC trajectories shown in Figure 3.4. Higher dispersive LPC trajectories are obtained in the middle due to consonantal-constriction. Similarly for other Bangla VCV sequences, the LPC trajectories are shown in appendix A.

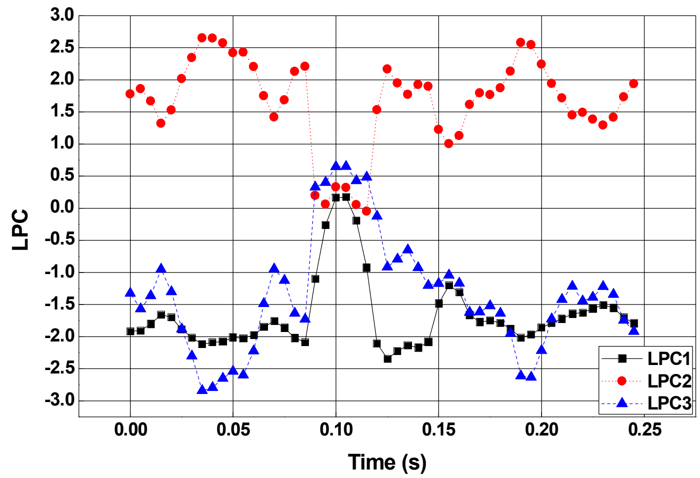


Figure 3.4: Evaluated first three LPC trajectories for Bangla VCV sequence (আলো)

3.4 LPC dispersion comparison for Bangla vowel and VCV sequences

Dispersion of LPC trajectory for Bangla vowels (অ, আ, ই, ঈ, উ, ঊ, এ, ঐ, ও, ঔ) and 5 VCV sequences (আলো, আমি, ইতি, অনু, উচু) are determined. For each speaker, we get 10 vowels each having 16 vector points. It is to be mentioned here that the average values of LPC filtering coefficients of Bangla vowels are shown in Table 3.1. Here we can see that the standard deviations for vowel LPC trajectories are found comparatively small in the range of 0.07 to 0.35. Among the vowels, \অ and \উ have lower-valued and \ঔ has higher-valued standard deviations. From standard-deviation comparisons, it is obvious that \ঔ is the most fricative vowel and \অ and \উ are most sustained vowels among the 10 vowels in Bangla. Vocal-tract dynamic-behavior is the principal factor behind the fricative and sustained nature.

Table 3.1: Average Standard Deviation of LPC filtering coefficients of Bangla vowel

	অ	আ	ই	ঈ	উ	ঊ	এ	ঐ	ও	ঔ
LPC1	0.2382	0.2671	0.2767	0.2864	0.1606	0.2451	0.2615	0.2665	0.2627	0.4044
LPC2	0.1568	0.186	0.1386	0.1722	0.1677	0.1567	0.1895	0.1972	0.1624	0.2444
LPC3	0.1392	0.1557	0.1147	0.1274	0.1045	0.1164	0.1415	0.1369	0.159	0.1649
LPC4	0.1208	0.1378	0.0974	0.1153	0.1009	0.1097	0.1131	0.1137	0.1415	0.126
LPC5	0.0936	0.097	0.0877	0.0902	0.0972	0.1054	0.100	0.1257	0.1031	0.1144
LPC6	0.0972	0.1115	0.0865	0.0988	0.0811	0.0774	0.1049	0.1111	0.0845	0.0909
LPC7	0.0925	0.1099	0.0925	0.0873	0.0782	0.0772	0.1233	0.1124	0.0862	0.0994
LPC8	0.0921	0.100	0.0808	0.0857	0.0729	0.0842	0.0836	0.0913	0.0864	0.0906
LPC9	0.0904	0.0917	0.0789	0.0858	0.0724	0.0748	0.091	0.1036	0.083	0.0922
LPC10	0.0876	0.0917	0.0932	0.0906	0.0786	0.0775	0.0991	0.1058	0.0909	0.1184
LPC11	0.0857	0.0918	0.0925	0.091	0.0738	0.0907	0.1066	0.0975	0.0857	0.0976
LPC12	0.0728	0.0857	0.0906	0.0873	0.0724	0.081	0.0857	0.0954	0.0871	0.0856
LPC13	0.0775	0.0827	0.0694	0.0813	0.0684	0.0686	0.0846	0.0823	0.0818	0.0901
LPC14	0.0853	0.0844	0.0738	0.0867	0.054	0.0658	0.0837	0.0788	0.0918	0.0749
LPC15	0.0771	0.0896	0.0668	0.0871	0.0685	0.0698	0.0777	0.0918	0.0748	0.0747
LPC16	0.0638	0.0675	0.0686	0.0757	0.0812	0.0684	0.0671	0.0958	0.0655	0.0706

On the other hand, for VCV sequences, similar procedure is done and wide range of standard-deviation (0.10-0.82) is estimated for five speakers and average values of standard deviation are shown in Table 3.2. It indicates the nature of imposed consonantal-constrictions in V-V sequence. Although the consonantal-constriction sustains for a short time, it increases the standard-deviation more than two times than the most fricative vowel. In both vowels and VCV sequences the higher-valued standard-deviations are observed for first five LPC trajectories.

Table 3.2: Average Standard Deviation of LPC filtering coefficients of Bangla VCV sequences

	আলো	আমি	ইতি	অনু	উচু
LPC1	0.34569	0.28591	0.65649	0.25329	0.59686
LPC2	0.49436	0.52296	0.81719	0.48509	0.63837
LPC3	0.56124	0.52318	0.6583	0.5838	0.61794
LPC4	0.49093	0.50803	0.4829	0.54698	0.44831
LPC5	0.40083	0.40394	0.46923	0.4524	0.31843
LPC6	0.35665	0.42904	0.43927	0.41769	0.30247
LPC7	0.31249	0.40955	0.35537	0.38509	0.27078
LPC8	0.33835	0.36066	0.25633	0.33371	0.29759
LPC9	0.26207	0.29936	0.25473	0.30315	0.23629
LPC10	0.24413	0.29576	0.26865	0.27424	0.2789
LPC11	0.27568	0.28262	0.25375	0.27892	0.2428
LPC12	0.23513	0.28872	0.2662	0.2912	0.2435
LPC13	0.21542	0.29693	0.23844	0.31797	0.21645
LPC14	0.20256	0.27186	0.20904	0.32707	0.19372
LPC15	0.18688	0.22204	0.17172	0.24811	0.15615
LPC16	0.11117	0.13919	0.10857	0.1433	0.10098

3.5 Transitional energy for Bangla vowel and VCV sequences

Formant frequency of the filter defined by LPC trajectories are also evaluated in Praat. The STFT coefficients of formant trajectory have been calculated using the formula shown in (2.7). Here 1^{st} , 2^{nd} and 3^{rd} formant trajectories are shown in Figure 3.5, Figure 3.6 and Figure 3.7 for Bangla vowel (অ). Similarly, for Bangla VCV sequences, same procedure is done and spectrogram of 1^{st} , 2^{nd} and 3^{rd} formant trajectories are shown in Figure 3.8, Figure 3.9, Figure 3.10 for Bangla VCV sequence (আলো). From these analysis, local frequency components having almost ten times intensity are found in formant frequency spectrogram of VCV sequence than vowel. For other Bangla vowels and VCV sequences, spectrogram of 1^{st} trajectories are shown in appendix A.

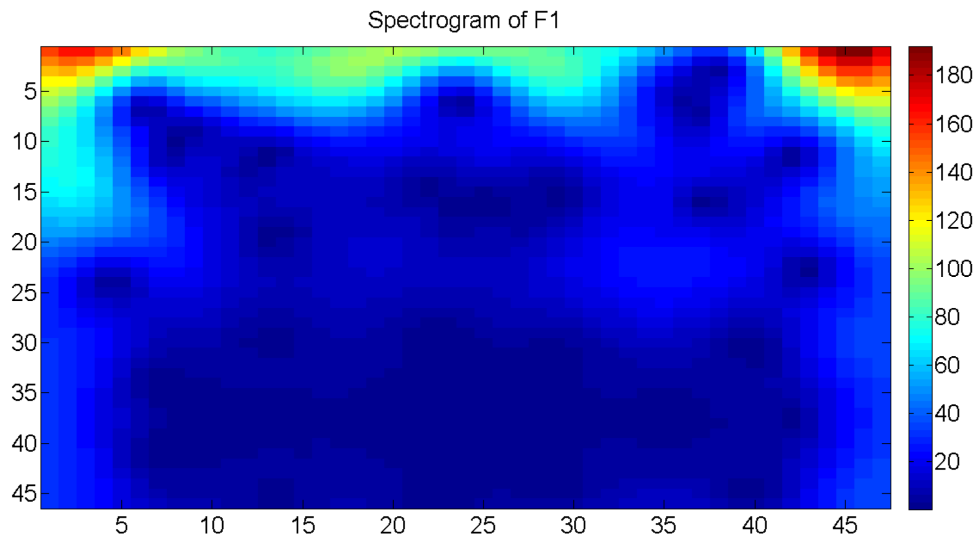


Figure 3.5: Spectrogram of 1st formant trajectory for Bangla vowel (অ)

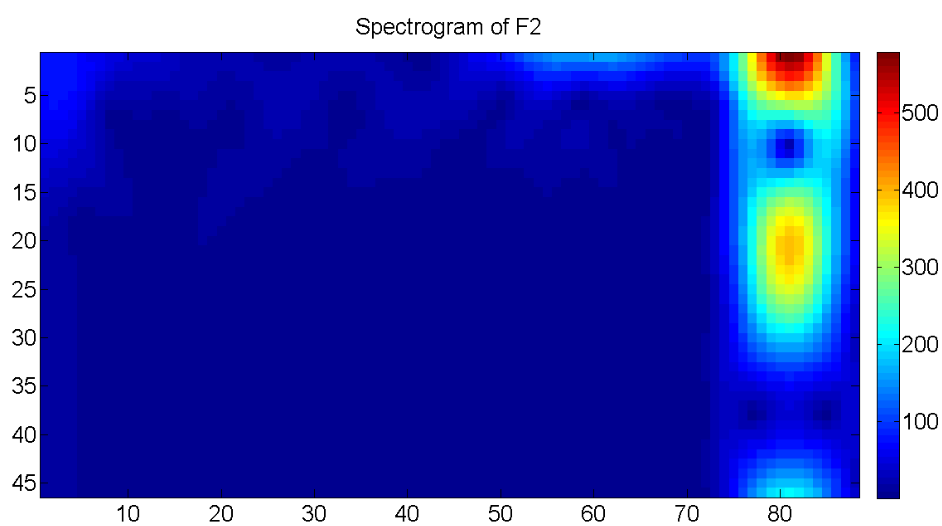


Figure 3.6: Spectrogram of 2nd formant trajectory for Bangla vowel (অ)

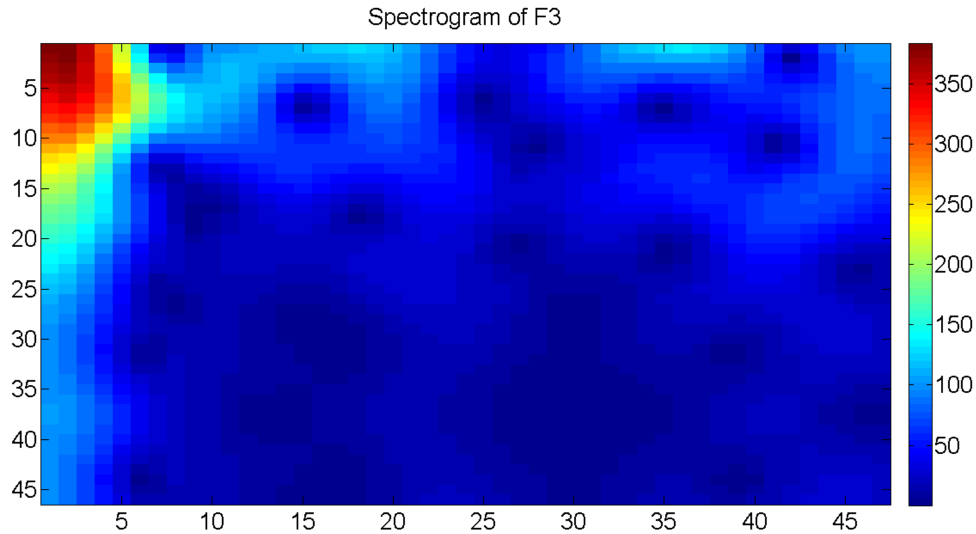


Figure 3.7: Spectrogram of 3rd formant trajectory for Bangla vowel (অ)

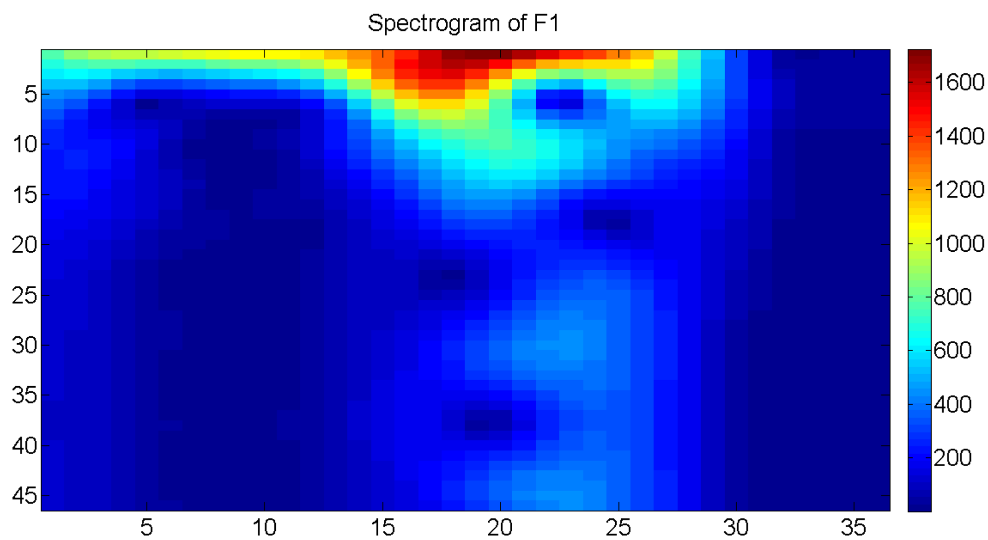


Figure 3.8: Spectrogram of 1st formant trajectory for Bangla VCV sequence (আলো)

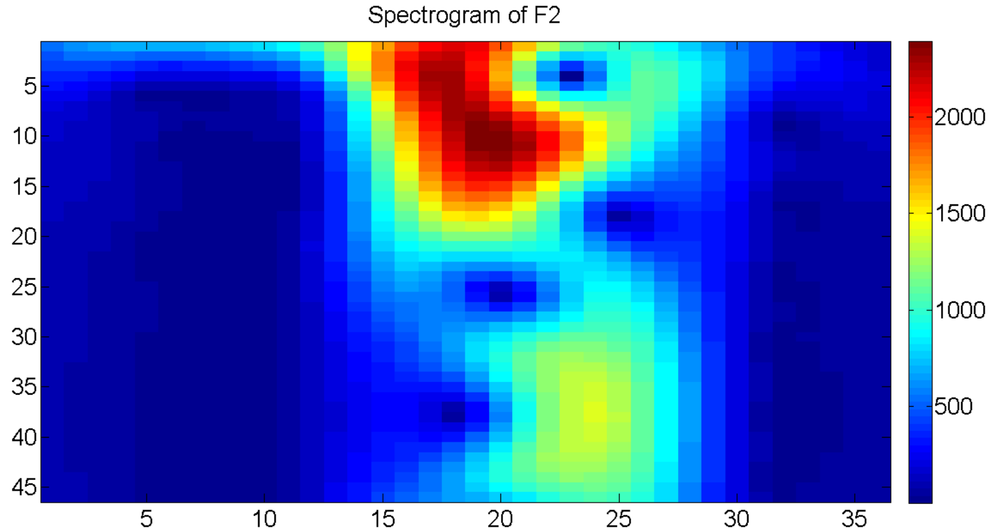


Figure 3.9: Spectrogram of 2nd formant trajectory for Bangla VCV sequence (আলো)

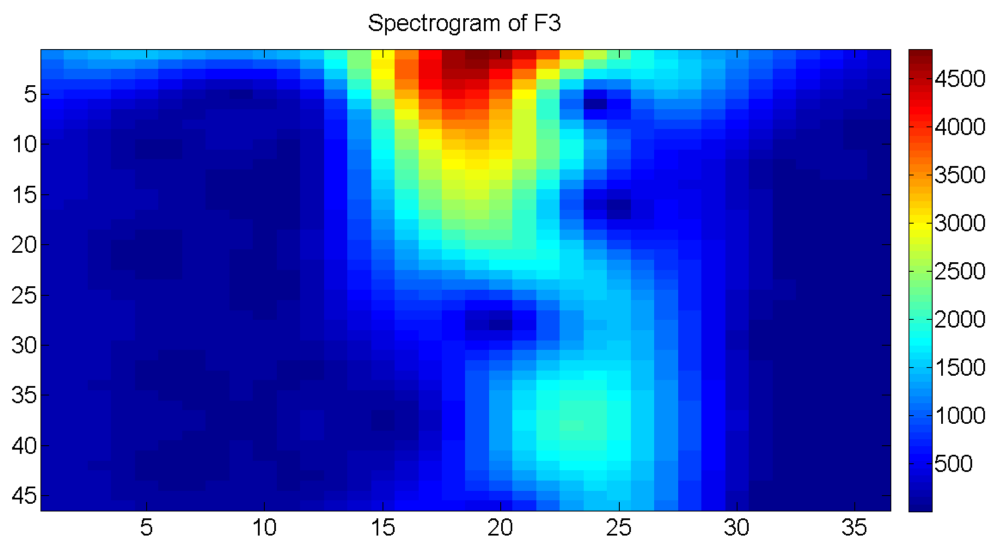


Figure 3.10: Spectrogram of 3rd formant trajectory for Bangla VCV sequence (আলো)

The average transitional energy is calculated by using (2.9) for vowels and VCV sequences uttered by five different speakers. The average transitional energies for vowels and VCV sequences are tabulated in above Table 3.3 and Table 3.4. Considering all five speakers, average formant transitional energy is also ten times higher for VCV sequences due to its consonantal-constriction in vocal-tract. The transitional energies for 10 vowels and 5 VCV sequences are tabulated in appendix B.

Table 3.3: Average formant trajectory transitional energy for Bangla vowels

Name of vowels	Variation tendency				
	F1	F2	F3	F4	F5
অ	1.88E+05	1.87E+05	4.61E+05	4.00E+05	6.12E+05
আ	1.97E+05	2.00E+05	2.38E+05	3.46E+05	4.06E+05
ই	4.13E+04	1.99E+05	3.58E+05	6.12E+05	5.75E+05
ঈ	1.88E+04	2.35E+05	4.88E+05	4.37E+05	5.85E+05
উ	1.27E+05	3.26E+05	4.44E+05	4.99E+05	3.92E+05
ঊ	1.08E+05	2.28E+05	5.87E+05	5.04E+05	1.88E+05
এ	7.69E+04	1.70E+05	3.14E+05	3.64E+05	5.07E+05
ঐ	2.00E+05	3.87E+05	5.94E+05	3.99E+05	5.84E+05
ও	1.26E+05	3.12E+05	4.06E+05	2.88E+05	4.90E+05
ঔ	2.30E+05	3.74E+05	3.78E+05	6.88E+05	4.61E+05

Table 3.4: Average formant trajectory transitional energy for Bangla VCV sequences

Name of VCVs	Variation tendency				
	F1	F2	F3	F4	F5
আলো	1.32E+06	2.31E+06	3.27E+06	2.42E+06	3.25E+06
আমি	1.48E+06	3.92E+06	3.11E+06	2.60E+06	3.88E+06
ইতি	1.59E+06	5.39E+06	3.89E+06	3.20E+06	4.64E+06
অনু	1.33E+06	3.63E+06	4.72E+06	3.26E+06	4.17E+06
উচু	2.37E+06	5.01E+06	4.66E+06	3.01E+06	3.70E+06

3.6 Correlation among acoustic property of Bangla vowel pairs and VCV sequence

To observe the relation among the acoustic property of Bangla vowel, cross-correlation pairs are calculated for each vowel. The correlations of vowels are calculated using (2.10). The correlation coefficient matrix for Bangla vowel pairs is shown in Table 3.5. From these table, it is seen that the values of correlation coefficients among the Bangla vowels is greater than 0.25. So, here is a strong correlation between the acoustic properties of Bangla vowels. Vowels pairs (অ-আ, আ-এ, ই-ঈ, ই-ঐ, ঈ-এ, ঈ-ঐ, উ-ঊ, ঐ-ও) have higher acoustical similarity.

The correlations of vowels and VCV sequences are also calculated and shown in Table 3.6. The average value of correlation coefficient for vowels and VCV sequences are 0.1538 where the average value for vowel-vowel is 0.6516. On average, the consonantal-constriction related vocal-tract dynamics degrade the acoustical similarity 4.23 times. For different speakers, the correlations of vowels and VCV sequences are also calculated and shown in appendix B.

Table 3.5: Correlation of LPC filtering coefficients of Bangla vowel pairs

	অ	আ	ই	ঈ	উ	ঊ	এ	ঐ	ও	ঔ
অ	1									
আ	0.566	1								
ই	0.712	0.694	1							
ঈ	0.524	0.543	0.485	1						
উ	0.758	0.696	0.649	0.701	1					
ঊ	0.648	0.691	0.544	0.621	0.515	1				
এ	0.621	0.622	0.524	0.567	0.394	0.446	1			
ঐ	0.586	0.599	0.496	0.503	0.414	0.454	0.570	1		
ও	0.695	0.664	0.574	0.569	0.468	0.573	0.610	0.539	1	
ঔ	0.649	0.580	0.569	0.540	0.443	0.503	0.605	0.552	0.568	1

Table 3.6: Correlation of LPC filtering coefficients of Bangla vowels and VCV sequences

	আলো	আমি	ইতি	অনু	উচু
অ	0.181	0.163	0.146	0.130	0.215
আ	0.159	0.139	0.147	0.118	0.192
ই	0.180	0.166	0.140	0.129	0.214
ঈ	0.165	0.152	0.151	0.120	0.198
উ	0.171	0.149	0.140	0.120	0.202
ঊ	0.173	0.153	0.147	0.122	0.199
এ	0.148	0.144	0.155	0.107	0.171
ঐ	0.155	0.132	0.139	0.118	0.183
ও	0.170	0.152	0.144	0.119	0.191
ঔ	0.143	0.130	0.138	0.105	0.170

3.7 Perceptual space consistency for Bangla vowels by principal components

When identifying dissimilar sounds such as human vowels, the ears are most sensitive to peaks in the signal spectrum. These resonant peaks in the spectrum are called formants. Each vowel has different formant frequencies. Furthermore, every human being has his/her unique formant frequencies. Five formant trajectories of ten Bangla vowels (অ, আ, ই, ঈ, উ, ঊ, এ, ঐ, ও, ঔ) are determined using Praat script setting window length 0.025s. Figure 3.11 shows the formant tracks of the Bangla vowel. The Table 3.7 shows the first five formant frequencies of Bangla vowel sound.

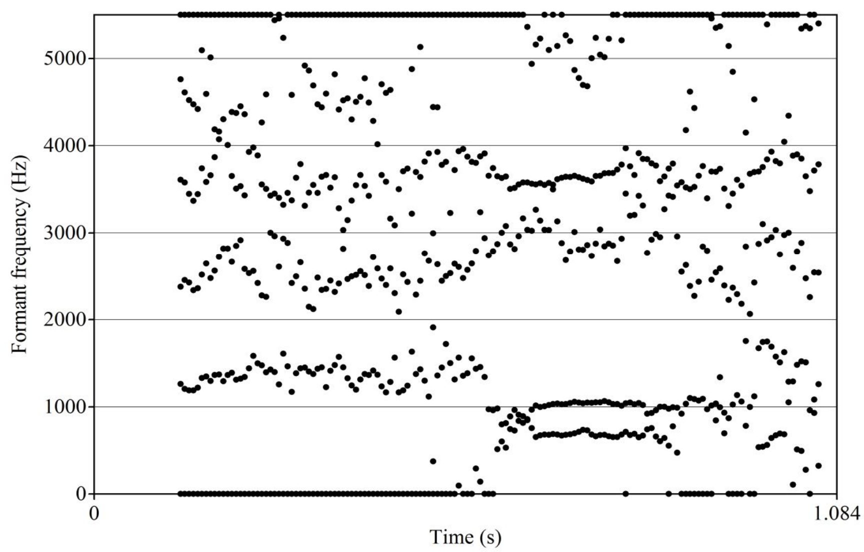


Figure 3.11: Formant contour for vowel transitions

Table 3.7: Formant frequency for Bangla vowel

Time (s)	Formant Frequency				
	F1	F2	F3	F4	F5
0.025	692.27	1046.44	2826.67	3582.8	4211.47
0.031	693.8	1053.38	2875	3583.48	4186.2
0.038	696.71	1058.47	2937.12	3585.32	4145.85
0.044	697.22	1057.4	2963.44	3585.45	4128.62
0.050	698.98	1057.46	2964.89	3595.96	4159.64
0.056	704.08	1060.51	2937.22	3610.37	4194.39
0.063	708.81	1062.31	2898.01	3609.82	4169.4
0.069	707.6	1056.98	2886.92	3607.9	4143.49
0.075	701.31	1052.5	2854.54	3616.32	4155.98
0.081	698.53	1055.65	2839.77	3627.13	4175.16
0.188	698.89	1059.3	2863.5	3620.76	4162.3

For each speaker, nine-dimensional vowel perceptual space, V is formed considering the ten vowels. The Table 3.8 shows the perception vectors of Bangla several vowels.

Table 3.8: Perceptual vector members for Bangla vowels

Bangla Vowels	Perception Vectors								
	F1	F2	F3	F4	F5	F51	F43	F53	F54
অ(a)	363.74	1213.92	2515.93	3541.81	4720.88	295.09	268.45	315.73	256.82
আ(â)	293.61	1177.71	2341.8	3458.62	4612.41	387.19	240.84	332.62	267.98
ই(i)	290.71	1522.83	2761.31	3616.7	4784.75	414.4	380.95	385.89	326.37
ঈ(î)	225.26	1475.6	2611.15	3534.57	4650.02	342.91	385.46	423.85	288.26
উ(u)	198.07	928.25	2115.67	3337.13	4472.47	358.35	308.96	294.8	186.5
ঊ(û)	222.4	1151	2385	3511.63	4651.07	317.37	244.42	279.48	210.61
এ (ê)	332.33	1391.13	2540.85	3491.83	4632.12	476.79	257.63	399.35	339.82
ঐ (ai)	330.8	1385.68	2496.08	3551.76	4729.99	297.37	212.4	228.76	212.83
ও (ô)	295.66	1147.66	2487.2	3545.64	4674.17	257.3	245.63	262.87	243.77
ঔ (au)	261.2	1200.92	2456.47	3509.51	4692.53	342.64	253.7	283.07	244.52

The energies and directions of the principal components are evaluated applying the"Eigen" function of statistical software R on the acoustic covariance matrix given in equation (2.13). Normalized energies or eigen-values μ_i are calculated as

$$\mu_i = \frac{\lambda_i}{\sum_i \lambda_i} \quad (4.1)$$

Normalized eigen-value profiles of vowel perceptual space for the fifteen speakers are evaluated and shown in Figure 3.12. Here,"S" refers the speaker and the number indicates speaker identity. Normalized first and second eigen-energy ranges are 0.87634-0.51354 and 0.3133-0.0627 respectively. Normalized third and fourth eigen-energy ranges are 0.1297-0.0263 and0.0544-0.0175. Normalized fifth and sixth energies are comparatively lower than first and second eigen- energies. Average eigen-value profile is also determined and tabulated in Table 3.9.

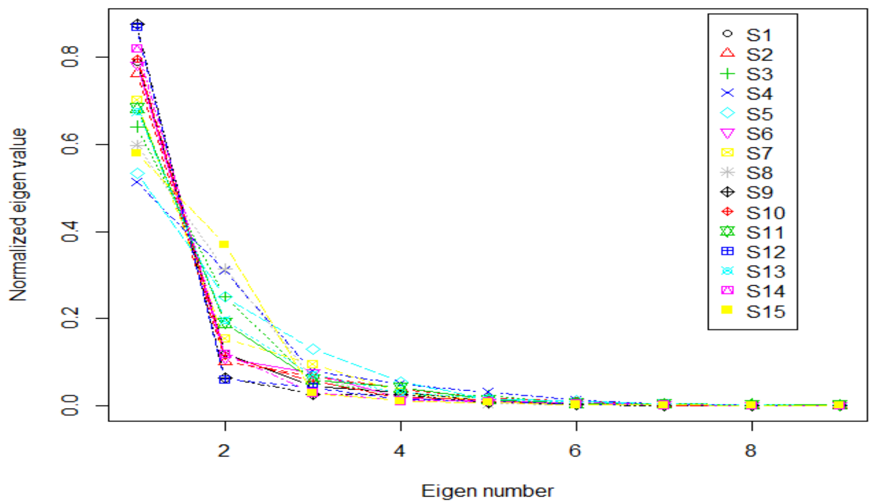


Figure 3.12: Normalized eigen-value profiles for different Speakers

Table 3.9: Average normalized eigen-value of Bangla vowel perceptual space

Eigen number	Mean normalized eigen-value (μ_i)
1	0.708535265
2	0.18116946
3	0.061469749
4	0.027744544
5	0.012726587
6	0.005378277
7	0.002105023
8	0.000672229
9	0.000198867

The summation of first and second eigen-values is 0.88969 and the acceptable limit of the eigen-value summation is 0.800 [11]. Here, first two principal components consist of sufficient eigen-energy or information of Bangla vowel perceptual space.

First eigen-vectors for fifteen perceptual spaces with the contributions of perceptual vector members are shown in Figure 3.13. The contributions of F2, F3, and F4 in first eigen-vector are comparatively higher than the other members of perception space. The average normalized contributions of the members in first principal component are tabulated in Table 3.10.

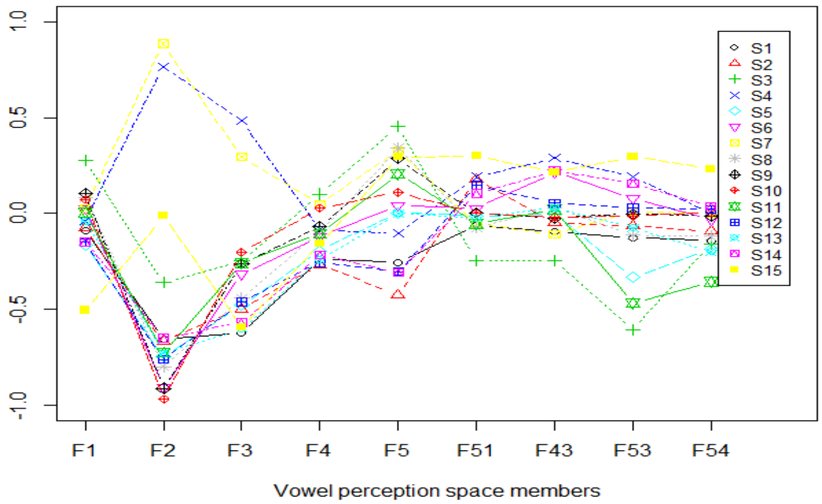


Figure 3.13: First eigen-vectors of fifteen Bangla vowel perceptual spaces

Table 3.10: Average normalized contributions of perceptual space members in first principal component

Perception Space Members	Average Normalized Contributions	Standard Deviation
F1	-0.04069	0.173216
F2	-0.48342	0.581859
F3	-0.31868	0.322552
F4	-0.12459	0.116682
F5	0.044767	0.277094
F51	0.028469	0.134936
F43	0.035678	0.145566
F53	-0.06928	0.242016
F54	-0.06614	0.136543

Similarly, the second principal component which consists of 18.11% average variance of the perceptual space are also evaluated and shown in Figure 3.14. Comparative inconsistencies in contributions are found from the graphical representation of second eigen-vectors with respect to first. The average contributions of perceptual vector members is also determined taking mean value of second eigen-vectors and provided in Table 3.11. The maximum contribution -0.08763) is obtained from F5. Except F5, the average contribution of other perceptual vector members is not significant as second principal components are randomly distributed and do not exhibits consistencies like first principal components.

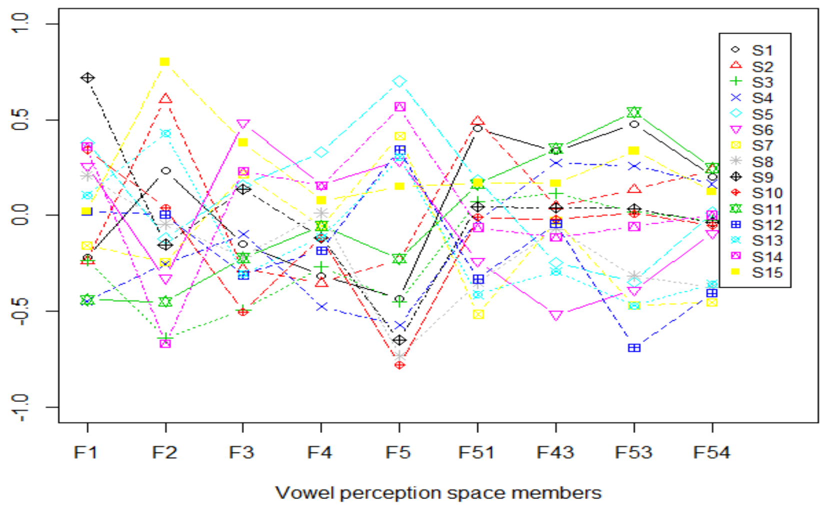


Figure 3.14: Second eigen-vectors of fifteen Bangla vowel perceptual spaces

Table 3.11: Average normalized contributions of perceptual space members in second principal component

Perception Space Members	Average Normalized Contributions	Standard Deviation
F1	0.045287	0.324352
F2	-0.05246	0.411438
F3	-0.0647	0.299894
F4	-0.08736	0.208738
F5	-0.08763	0.488611
F51	-0.02508	0.290571
F43	0.000155	0.229518
F53	-0.06165	0.36109
F54	-0.05393	0.232662
F54	-0.05393	0.232662

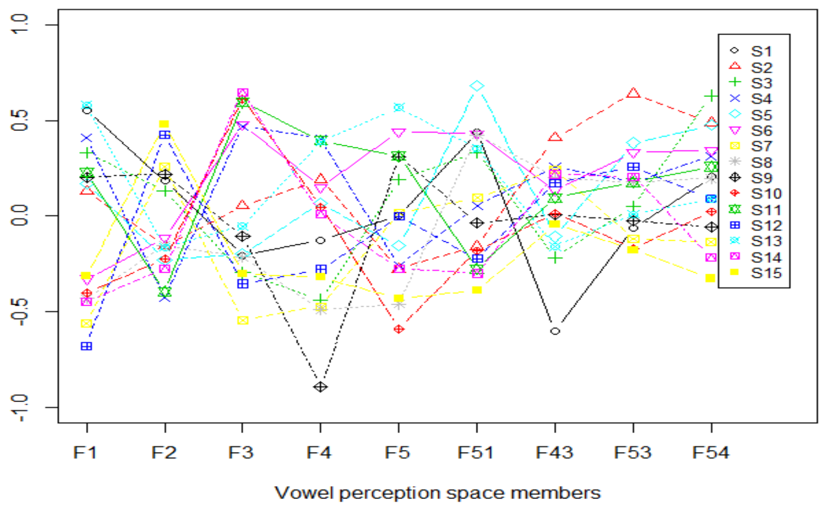


Figure 3.15: Third eigen-vectors of fifteen Bangla vowel perceptual spaces

Table 3.12: Average normalized contributions of perceptual space members in third principal component

Perception Space Members	Average Normalized Contributions	Standard Deviation
F1	-0.03746	0.428607
F2	-0.02916	0.288703
F3	0.038708	0.407271
F4	-0.09051	0.38393
F5	-0.04091	0.349833
F51	0.083102	0.337568
F43	0.041815	0.247786
F53	0.123703	0.226832
F54	0.158999	0.27308

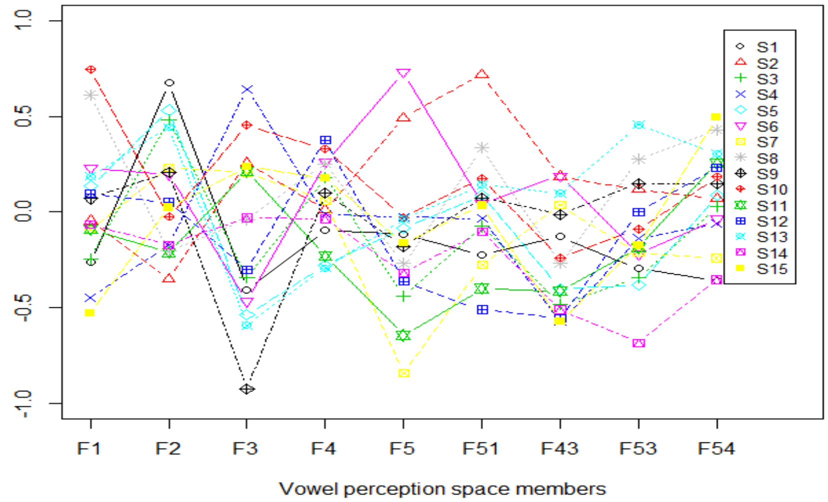


Figure 3.16: Fourth eigen-vectors of fifteen Bangla vowel perceptual spaces

Table 3.13: Average normalized contributions of perceptual space members in fourth principal component

Perception Space Members	Average Normalized Contributions	Standard Deviation
F1	0.01940	0.33267
F2	0.11454	0.29812
F3	-0.10914	0.43432
F4	0.05335	0.20398
F5	-0.15281	0.38756
F51	-0.00068	0.28913
F43	-0.24329	0.29417
F53	-0.11498	0.28159
F54	0.07898	0.24534

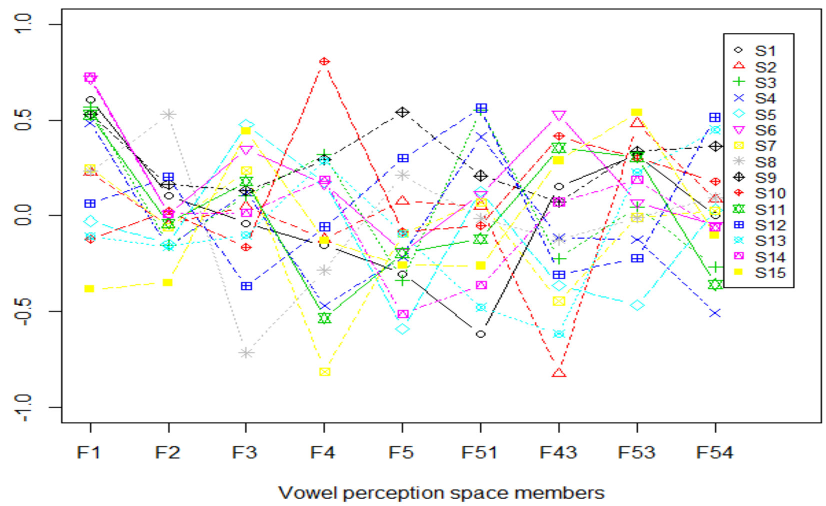


Figure 3.17: Fifth eigen-vectors of fifteen Bangla vowel perceptual spaces

Table 3.14: Average normalized contributions of perceptual space members in fifth principal component

Perception Space Members	Average Normalized Contributions	Standard Deviation
F1	0.28065	0.32683
F2	-0.00499	0.19787
F3	0.04805	0.29788
F4	-0.02240	0.39313
F5	-0.11618	0.29018
F51	0.01271	0.33609
F43	-0.07565	0.38098
F53	0.13163	0.26297
F54	0.02690	0.27293

The correlations among the first eigen-vectors of different perceptual spaces indicate similarity. With higher correlation indicates the information consistency or inter-speaker similarity of the perception space. The correlation matrix of first eigen-vector is evaluated using (2.15) are (2.16) and shown in Figure 3.18. Comparatively higher-valued correlations are found from the matrix and the average value of the correlation is 0.6851. Thus, numerical value of inter-speaker similarity is 68.51% for first eigen-vector of Bangla vowel perceptual space.

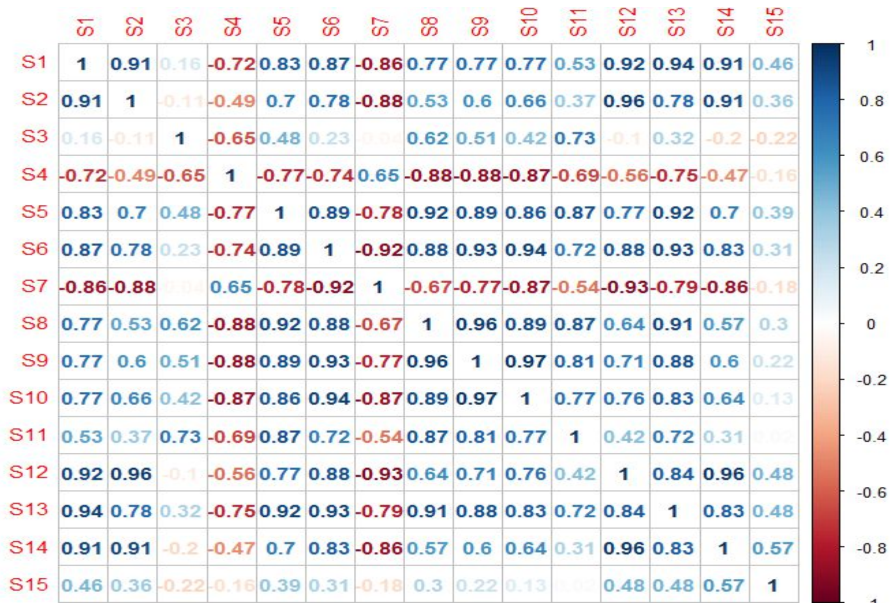


Figure 3.18: Correlation distributions of first eigen-vectors for Bangla vowel perceptual space

To determine the numerical consistency of second principal components throughout the fifteen perceptual spaces, the correlation matrix of second eigen-vectors is also evaluated and shown in Figure 3.19. Comparatively lower-level correlations are observed in the correlation map for the second eigen-vector. The average value of correlation is 0.4711 which is lower than the average of first eigen-vector (0.6851). Hence, Bangla vowel perceptual space, higher variance or information related component is more consistent. The results assure that during perceptual mapping of Bangla vowels, we have to set more concentration or weight on first principal component comparing to second.

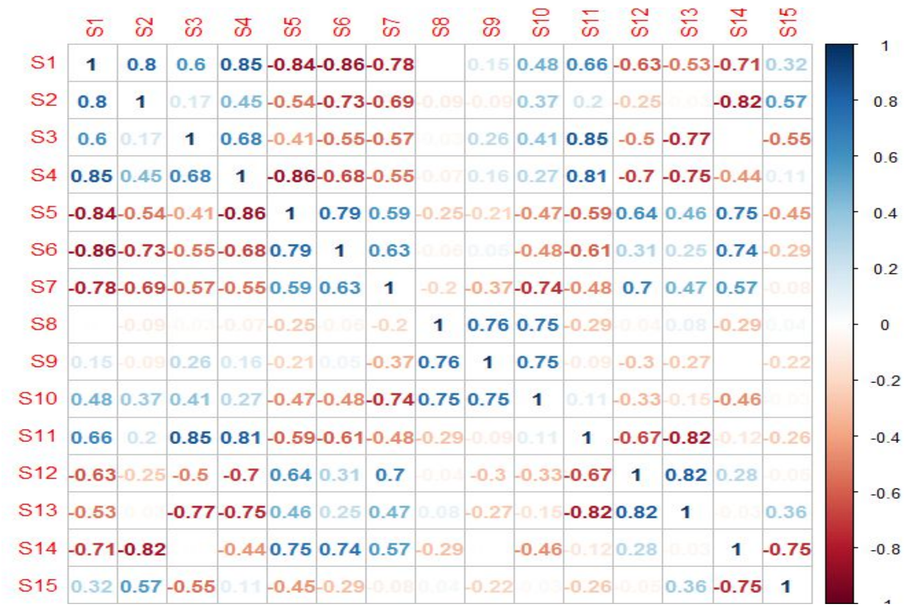


Figure 3.19: Correlation distributions of second eigen-vectors for Bangla vowel perceptual space

Similarly, lower-level correlations are observed in the correlation map for the third eigen-vector and shown in Figure 3.20. The average value of correlation is 0.3850 which is lower than the average of first eigen-vector and second eigen-vector. In the same way, the correlation matrix of fourth eigen-vector and fifth eigen-vector are determined and shown in Figure 3.21 and Figure 3.22 respectively. Their average value is 0.3659 and 0.3315. So, the numerical value of inter-speaker similarity is 36.59% and 33.15% for fourth eigen-vector and fifth eigen-vector of Bangla vowel perceptual space.

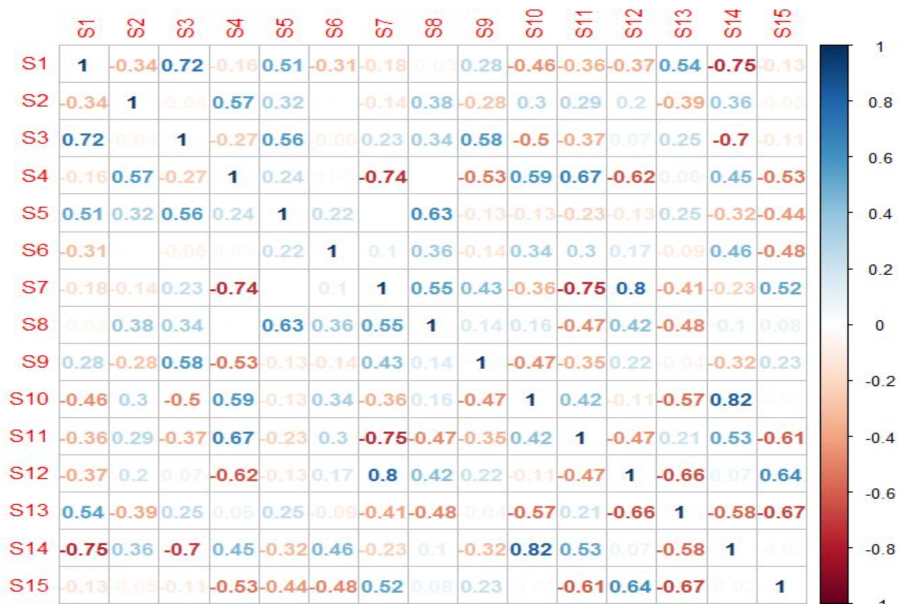


Figure 3.20: Correlation distributions of third eigen-vectors for Bangla vowel perceptual space

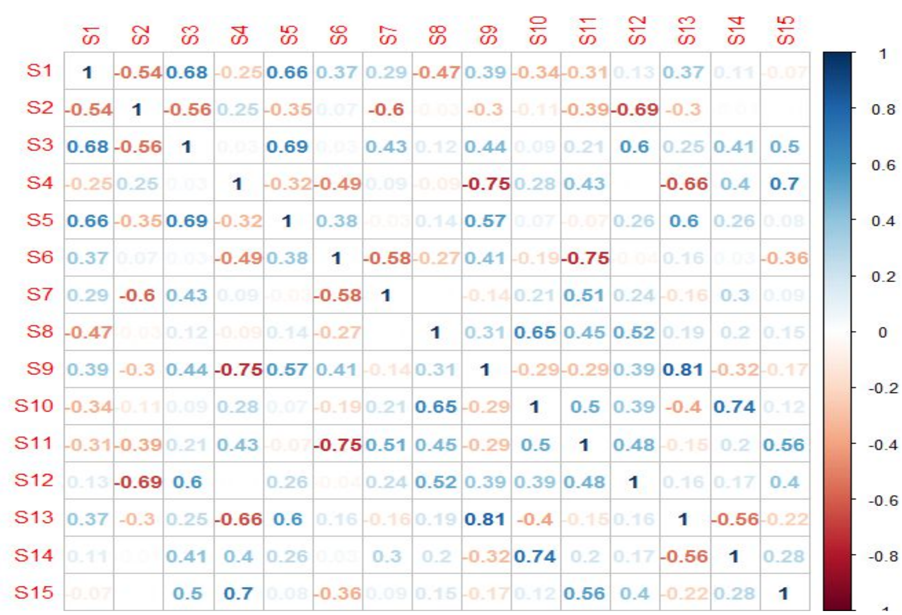


Figure 3.21: Correlation distributions of fourth eigen-vectors for Bangla vowel perceptual space

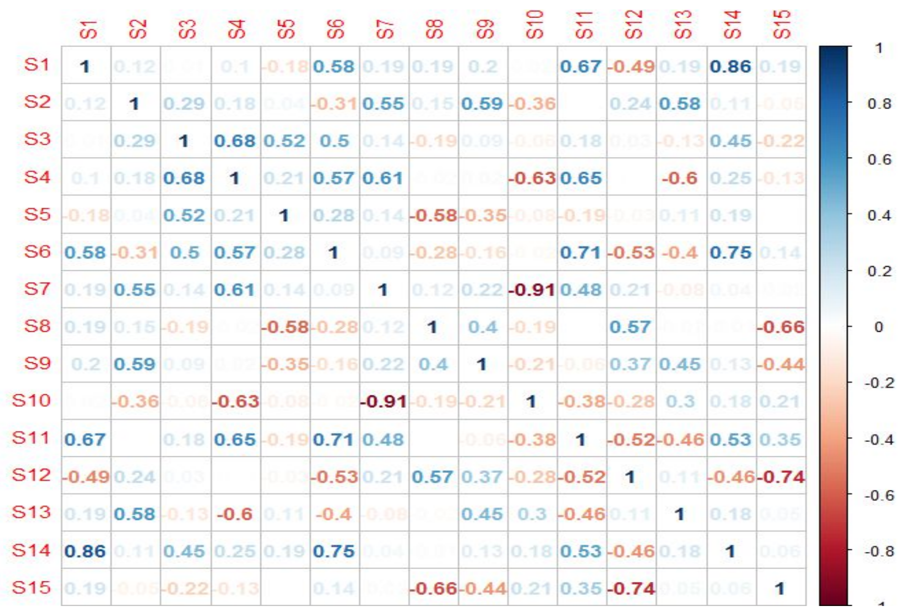


Figure 3.22: Correlation distributions of fifth eigen-vectors for Bangla vowel perceptual space

On the other hand, to determine the numerical consistency of first principal component throughout the nine perceptual space members, the correlation matrix of first eigen-vectors is computed and shown in Figure 3.23. The average value is 0.4717. Similarly, lower-level correlations are observed in the correlation map for the second eigen-vectors. The average value of correlation is 0.4322 for the second eigen-vectors of perceptual space members and shown in Figure 3.24. Correspondingly, the correlation matrix of third, fourth and fifth eigen-vectors are also determined and shown in Figure 3.25, Figure 3.26 and Figure 3.27 respectively. Their average values are 0.3761, 0.3272, and 0.3061. That result reveals that during perceptual mapping of Bangla space members, more concentration or weight have to set on first principal component that others principal component.

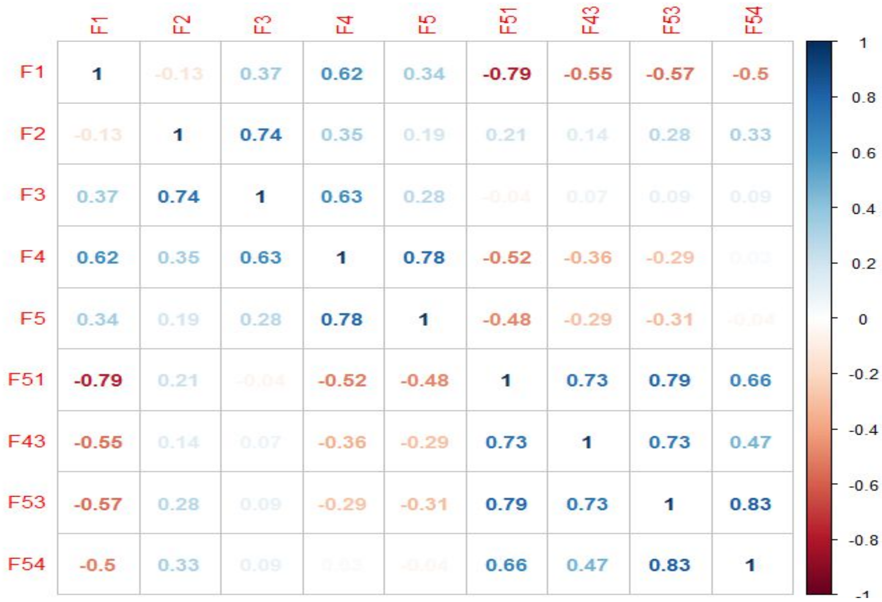


Figure 3.23: Correlation distributions of first eigen-vectors of perceptual space members for Bangla vowel

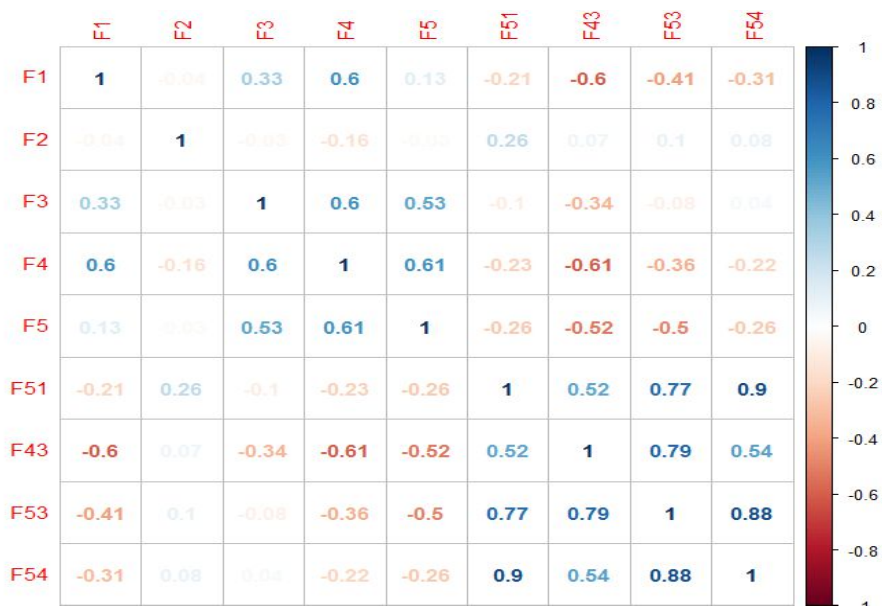


Figure 3.24: Correlation distributions of second eigen-vectors of perceptual space members for Bangla vowel

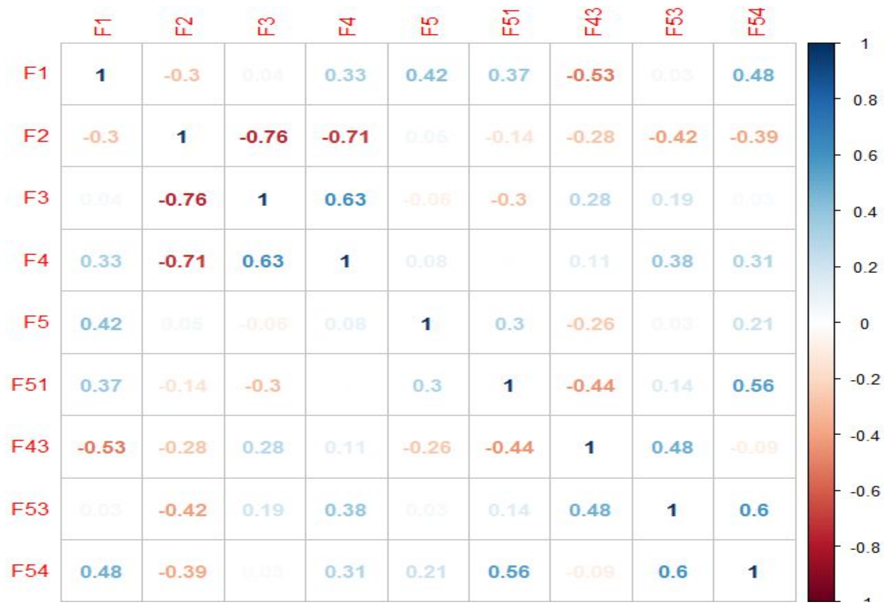


Figure 3.25: Correlation distributions of third eigen-vectors of perceptual space members for Bangla vowel

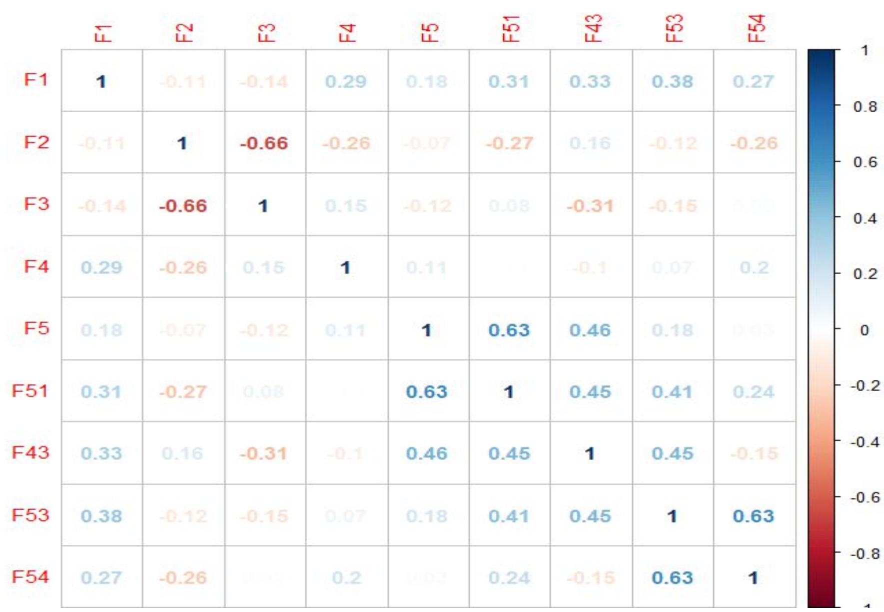


Figure 3.26: Correlation distributions of fourth eigen-vectors of perceptual space members for Bangla vowel

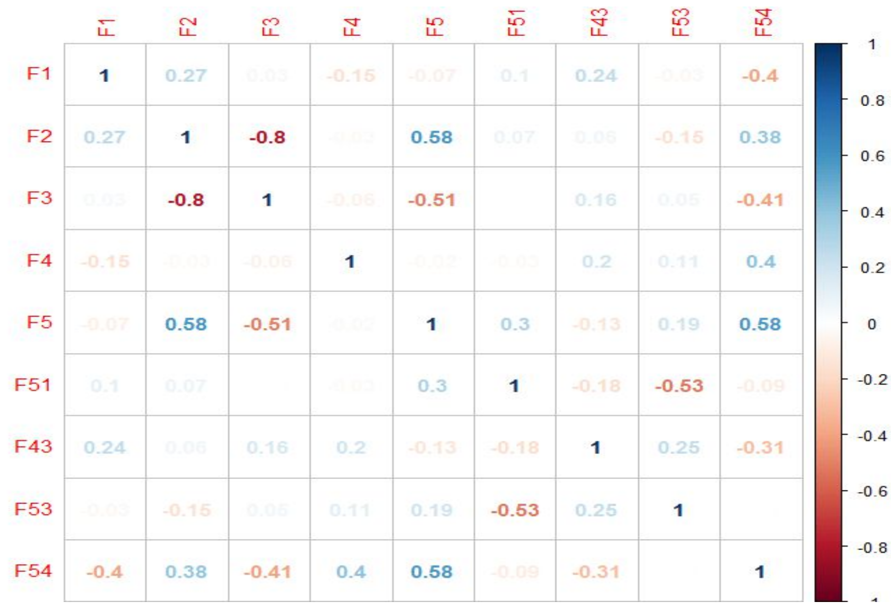


Figure 3.27: Correlation distributions of fifth eigen-vectors of perceptual space members for Bangla vowel

3.8 Vowel consistency by canonical correlation

The canonical correlation between the two subjects of acoustical content has been determined for the vowel consistency using the CCA package [42] in R software environment [43]. The reason for reducing five dimensions from nine dimensions has been discussed in mathematical formulation. The canonical correlation is associated with the correlations and cross-correlations of the inter-speaker consistency in acoustic spaces. As a preliminary step, the correlation matrices have been evaluated for the two vowel spaces. The correlation matrices are presented by concatenating speaker related of acoustic spaces by

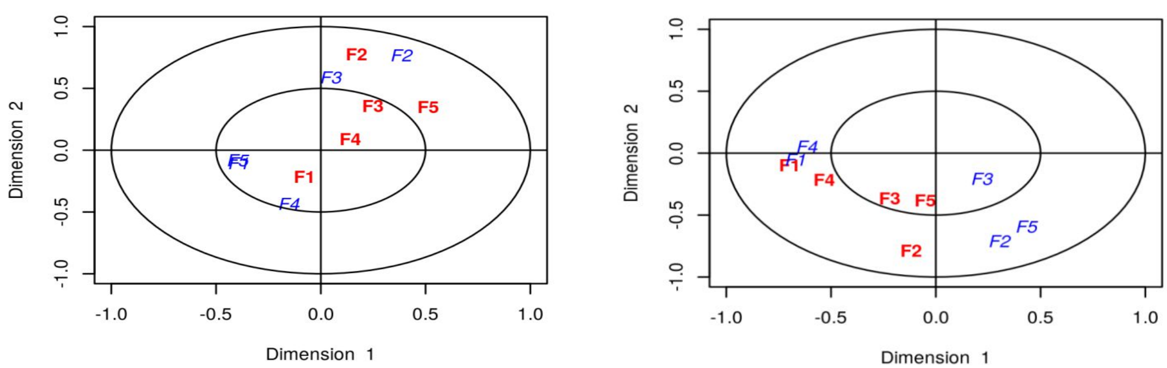
$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \quad (4.2)$$

In Figure 3.28, the off-diagonal sections (C_{12} and C_{21}) are the cross-correlation components and the brighter and darker colors represent positive and negative correlation. Comparatively faded color (0 in color bar) indicates the insignificant correlation. For the acoustic vector element-wise description subscripts 1 and 2 define the first and second speaker members of acoustic elements. The noticeable positive correlations are (F_2, F_2), (F_3, F_2), and (F5, F_2) for first and sixth speaker which is shown in Fig 4.62 (a). Similarly, for third and tenth speaker, the positive correlation is only (F_2, F_2) and (F_2, F_3) and negative correlations is (F_2, F_5).

(a) Between first and sixth speaker (b) Between third and tenth speaker

Figure 3.28: The correlation matrices between two vowel spaces (speakers)

The canonical variates are arranged according to their canonical correlation coefficients. As first and second canonical variates are highly correlated comparing to others, the acoustic vector elements are represented in first and second canonical variates plane in Figure 3.29. The red-regular and green-italic letters refers the number of first and second vowel spaces (speakers). The acoustic components are projected same directions and greater distances from the origin have stronger correlations. The inner and outer circles in Figure 3.29 are called correlation circles of radius 0.5 and 1.0 respectively.



(a) Between first and third speaker (b) Between first and sixth speaker

Figure 3.29: Acoustic components representation on the plane by two canonical variates

The salient acoustic features are located between these two circumferences. The numbers of salient acoustic features are 4, 7, and 9 between the first and third, first and sixth and second and third respectively and shown in Table 3.15.

Table 3.15: Number of salient features for vowel spaces

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	10														
2	9	10													
3	4	9	10												
4	6	7	8	10											
5	9	8	6	3	10										
6	7	7	8	8	6	10									
7	5	8	5	3	8	7	10								
8	9	9	9	7	9	6	8	10							
9	8	5	9	9	9	7	7	9	10						
10	6	8	8	7	6	7	7	7	7	10					
11	10	8	8	2	6	7	9	7	8	7	10				
12	9	9	8	2	3	2	8	8	10	8	9	10			
13	7	6	8	3	7	9	6	8	3	10	6	8	10		
14	10	6	6	7	9	7	8	9	10	8	4	6	3	10	
15	5	6	6	4	7	5	10	8	6	7	9	6	9	7	10

The canonical correlation coefficient indicates the proportion of variance of a pair canonical variate predictable from other member of the pair. In other words, the five dimensional canonical correlation coefficient refers the energy or variance redundancy of the variate pair. The canonical correlation coefficients for the vowel spaces are given in Table 3.16.

Table 3.16: Canonical correlation coefficient for vowel spaces (speakers)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.00														
2	0.72	1.00													
3	0.53	0.77	1.00												
4	0.32	0.68	0.76	1.00											
5	0.83	0.88	0.75	0.44	1.00										
6	0.64	0.68	0.78	0.49	0.73	1.00									
7	0.44	0.61	0.74	0.58	0.73	0.71	1.00								
8	0.76	0.65	0.67	0.62	0.81	0.60	0.88	1.00							
9	0.85	0.69	0.77	0.64	0.88	0.60	0.64	0.82	1.00						
10	0.82	0.70	0.78	0.66	0.71	0.65	0.43	0.78	0.86	1.00					
11	0.86	0.92	0.81	0.35	0.60	0.58	0.74	0.78	0.73	0.73	1.00				
12	0.87	0.79	0.69	0.60	0.79	0.52	0.70	0.83	0.84	0.69	0.80	1.00			
13	0.74	0.80	0.69	0.29	0.79	0.67	0.54	0.68	0.76	0.83	0.78	0.82	1.00		
14	0.78	0.89	0.67	0.46	0.85	0.70	0.69	0.90	0.84	0.72	0.77	0.83	0.87	1.00	
15	0.64	0.69	0.60	0.51	0.77	0.62	0.63	0.79	0.82	0.71	0.61	0.89	0.83	0.53	1.00

CHAPTER 4

Conclusion

Chapter Outlines

- Conclusion
- Recommendation of future work

4.1 Conclusion

In this study, at first Bangla vowels and VCV sequences are analyzed considering the acoustical features related to vocal-tract dynamics during uttering. Several research works have been done in speech perception field in the past, which are based on formant frequency and its dispersion. The LPC filtering coefficients provide important information about the vocal-tract dynamics as well as vowel and consonantal acoustic cues. So, the dynamic-shape of vocal-tract is main factors for VCV production, the statistical moments of LPC trajectories are the important cues for VCV recognition. It is observed that from LPC filtering coefficients dispersion perspective, \অ and \উ are most stable vowels and \ও is the most fricative vowel. Consonantal-constriction on V-V sequence gesture accelerates the transitional state of vocal-tract more than two-time than the most fricative vowel. Formant transitional energy is found more than ten times higher for VCV sequence with respect to vowel. The consonantal-transition of vocal-tract reduces the acoustical correlation with vowel 4.23 times on an average.

In speech perception, nine-dimensional Bangla vowel perceptual spaces have been formed considering linguistic information of ten vowels. Average eigen-value profile of principal components is evaluated considering perceptual spaces of fifteen speakers. It is noticed that first and second eigen-values are comparatively higher than the next eigen-values profile. From these values, we can see that summation of first and second eigen-values is 88.97%, which is higher than acceptable limit (80%) and this indicates that these two components are sufficient to represent the information of Bangla vowel perceptual space. Speaker independency has been modeled using correlation of corresponding eigen-vectors of fifteen perceptual spaces. The average correlations of first and second eigen-vectors are 0.6853 and 0.4711 respectively which is higher than others eigen-vector. So, in perceptual mapping with vowels, more concentration or weight have to set on first principal component, which consists 70.853% information. In first principal component, the average contributions of F2, F3, and F4 are significant, but in second principal component only F5 contributes significantly. As the second eigen-vectors are comparatively less invariant, the average contributions become negligible. On the other hand, the canonical correlation between speakers related of acoustical content has been developed using the redundancy of the canonical variates for vowel consistency. The average canonical correlation coefficient is 0.74 for vowel spaces whereas maximum correlation coefficient is 1. These salient features play the principle roles in enhancing the acoustic similarity.

4.2 Recommendation of future work

The future research related to this work can be carried out as:

- Morphed Bangla vowel and VCV sequences will be produced and analyzed.
- Recognition of morphed sounds on the basis of the statistical developed in this research work.

References

[1] Saeed V. Vaseghi, *Speech processing in multimedia signal processing: Theory and applications in speech, music and communications,* John Wiley & Sons, Ltd, 2007.

[2] K. N. Stevens, *Acoustic phonetics,* The MIT Press, Cambridge, Massachusetts, USA 1998.

[3] T. F. Quatieri, *Discrete-time speech signal processing: Principles and practice,* Prentice Hall, 2002.

[4] K. R. Scherer, "Vocal communication of emotion: A Review of Research Paradigms," *Speech Communication*, vol. 40, no. 1, pp. 227-256, April 2003.

[5] R. J. Moran, R. B. Reilly, P. de Chazal, and P. D. Lacy, "Telephony-based voice pathology assessment using automated speech analysis," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 3, pp. 468-477, March 2006.

[6] X. Yao, T. Jitsuhiro, C. Miyajima, N. Kitaoka, and K. Takeda, "Classification of speech under stress based on modeling of the vocal folds and vocal tract," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2013:17, 2013.

[7] R. B. Shinde and V. P. Pawar, "Vowel classification based on LPC and ANN," *International Journal of Computer Applications*, vol. 50, no. 6, pp. 1-9, July 2012.

[8] J.D. Markel and A.H. Gray Jr., "Linear prediction of speech," *Journal of Sound and Vibration*, vol. 51, no. 4, pp. 595-595, April 1976.

[9] http:/www.statista.com/statistics/266808/the-most-spoken-language-worldwide.

[10] Lori L. Holt, Andrew J. Lotto, "Speech perception as categorization," *Atten Percept Psychophys*, vol. 72, no. 5, pp. 1218–1227, doi:10.3758/APP.72.5.1218, July 2010.

[11] O. Baumann and P. Belin, "Perceptual scaling of voice identity: common dimensions for different vowels and speakers," *Psychological Research*, vol. 74, pp. 110-120, January 2010.

[12] Alamgir Kabir, Rebecca D. Merrill , Abu Ahmed Shamim , Rolf D. W. Klemn , Alain B. Labrique , Parul Christian , Keith P. West Jr , Mohammed Nasser, " Canonical correlation analysis of infant's size at birth and maternal factors: A Study in Rural Northwest Bangladesh," *PLoS ONE 9(4): e94243,* doi:10.1371/journal. pone. 0094243, April 7, 2014.

[13] B. H. Story, "Vowel and consonant contributions to vocal tract shape," *Journal of Acoustical Society of America*, vol. 126, no. 2, pp. 825–836, August 2009.

[14] S. E. G. Öhman, "Coarticulation in VCV utterances: spectrographic measurements," *The Journal of the Acoustical Society of America*, vol. 39, no. 1, pp. 151–168, 1966.

[15] B. Yegnanarayana and R. N. J. Veldhuis, "Extraction of vocal-tract system characteristics from speech signals," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 4, pp. 313-327, July 1998.

[16] A. Hasanat, Md. R. Karim, Md. S. Rahman, and Md. Z. Iqbal, "Recognition of spoken letters in Bangla," *5^{th} International Conference on Computer and Information Technology (ICCIT)*, pp. 213-226, 2002.

[17] S. A. Hossain, M. L. Rahman, and F. Ahmed, "Bangla vowel characterization based on analysis by synthesis," *International Journal of Computer, Information, Systems and Control Engineering*, vol. 1, no. 2, pp. 354-357, 2007.

[18] S. A. Hossain, M. L. Rahman, and F. Ahmed, "Acoustic space of Bangla vowels," *5th WSEAS International Conerence on Signal, Speech and Image Processing*, pp. 138-142, Corfu, Greece, August 2005.

[19] A. K. Paul, D. Das, and Md. M. Kamal, "Bangla speech recognition system using LPC and ANN," *Seventh International Conference on Advances in Pattern Recognition*, Kolkata, 2009.

[20] G. E. Peterson, and H. L. Barney, "Control methods used in a study of the vowels," *Journal of the Acoustical Society of America*, vol. 24, pp. 175-184, 1952.

[21] P. Belin, S. Fecteau, and C. Bedard, "Thinking the voice: neural correlates of voice perception," *Trends in Cognitive Science*, vol. 8, pp. 129-135, 2004.

[22] M. L. Erickson, "Acoustic properties of the voice source and the vocal tract: Are they perceptually independent?," *Journal of Voice*, (in press), 2016.

[23] D. B. Pisoni, "Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops," *Journal of the Acoustical Society of America*, vol. 61, pp. 352-1361,1977.

[24] T. M. Nearey, "Static, dynamic, and relational properties in vowel perception," *Journal of the Acoustical Society of America*, vol. 85, pp. 2088-2113, 1989.

[25] P. G. Vilda, J. M. F. Vincente, and V. R. Biarge, "Simulating the phonological auditory cortex from vowel representation spaces to categories," *Neurocomputing*, vol. 114, pp. 63-75, August 2013.

[26] H. Matsumoto, S. Hiki, T. Sone, and T. Nimura, "Multidimensional representation of personal quality of vowels and its acoustical correlates," *IEEE Transacrions on Audio and Electroacoustics*, vol. AU-21, October, 1973.

[27] K. J. Reilly and K. E. Dougherty, "The role of vowel perceptual cues in compensatory responses to perturbations of speech auditory feedback," *Journal of the Acoustical Society of America*, vol. 134, pp. 1314-1323, August 2013.

[28] B. E. Walden, A. A. Montgomery, G. T. Gibeily, R. A. Prosek, and D. M. Schwartz, "Correlates of psychological dimensions in talker similarity," *Journal of Speech and Hearing Research*, vol. 21, pp. 265275, 1978.

[29] J. Kreiman, B. R. Gerratt, K. Precoda, and G. S. Berke, "Individual differences in voice quality perception," *Journal of Speech and Hearing Research*, vol. 35, pp. 512-520, 1992.

[30] D. Giuliani, M. Gerosa, and F. Brugnara, "Improved automatic speech recognition through speaker normalization," *Computer Speech & Language*, vol. 20, pp. 107-124, January 2006.

[31] J. D. Harnsberger, M. A. Svirsky, A. R. Kaiser, D. B. Pisoni, R. Wright, and T. A. Meyer, "Perceptual vowel spaces of cochlear implant users:Implications for the study of auditory adaptation to spectral shift," *Journal of the Acoustical Society of America*, vol. 109, pp. 2135-2145, May 2001.

[32] Gorden E. Pelton, *Voice Processing,* Mc. Graw. Hill, New York, 1993.

[33] J.L. Flanagan, K. ishizaka, and k. L. Shipley, "Synthesis of speech from a dynamc model of the vocal cords and vocal tract," *The Bell System Technical Journal*, vol. 54, no. 3, March 1975, U.S.A

[34] G. Fant, *Acoustic Theory of Speech Production*, Mouton: The Hague, 1960.

[35] H. Wakita, "Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms," *IEEE Transactions on Audio and Electroacoustics*, vol. AU-21, no. 5, pp. 417-427, October 1973.

[36] R.D. Kent, and F. D. Minifie, "Coarticulation in recent speech production models," *Journal of Phonetics*, vol. 5, pp. 115-133, 1977.

[37] A. M. Liberman, and F. S. Cooper, and D. P. Shankweiler, and M. Studdert-Kennedy, "Perception of the speech code," *Psychological Review*, vol. 74, pp. 431-461, 1967.

[38] F. Li, A. Menon, and J. B. Allen, "A psychoacoustic method to find the perceptual cues of stop consonants in natural speech," *Journal of Acoustical Society of America*,vol. 127, pp. 2599-2610, April 2010.

[39] S. Lopez, P. Riera, M. F. Assaneo, M. Eguia, M. Sigman, and M. A. Trevisan, "Vocal caricatures reveal signatures of speaker identity," *Nature Scientific Reports 3*, 3407,doi:10.1038/srep03407, December 2013.

[40] Olivier Ledoit, and Michael Wolf, "A well-conditioned estimator for large-dimensional covariance matrices", *Journal of Multivariate Analysis*, vol. 88, pp. 365-411, 2004.

[41] D. Stewart, and W. Love, "A general canonical correlation index," *Psychological Bulletin*, vol.70, pp.160-163, 1968.

[42] I. Gonzalez, S. Dejean, P. G. P. Martin, and A. Baccini, "CCA: An R package to extend canonical correlation analysis," *Journal of Statistical Software*, vol. 23,pp.1-14, January 2008.

[43] R Core Team, R: *A language and environment for statistical computing, R foundation for statistical computing,* Vienna, Austria, 2013. (http: //www.R-project.org/)

Publication from this Thesis

[1] **Sathi Rani Mitra** and Md. Mahbub Hasan, "Comparison of vocal tract dynamics for Bangla vowel and Vowel-Consonant-Vowel sequence,"*1st International Conference on Advanced Information and Communication Technology,* May 2016.

[2] **Sathi Rani Mitra**, Md. Mahbub Hasan and Kenbu Teramoto, "Speaker invariant principal component determination for Bangla vowel perceptual space,"*5th International Conference on Informatics, Electronics and Vision*, May 2016.

[3] Md. Mahbub Hasan, **Sathi Rani Mitra** and Kenbu Teramoto, "Canonical correlation based impersonation quality determination algorithm for natural morphed speech," *1st IEEE International Conference on Telecom & Photonics*, September 2015.

Appendix A

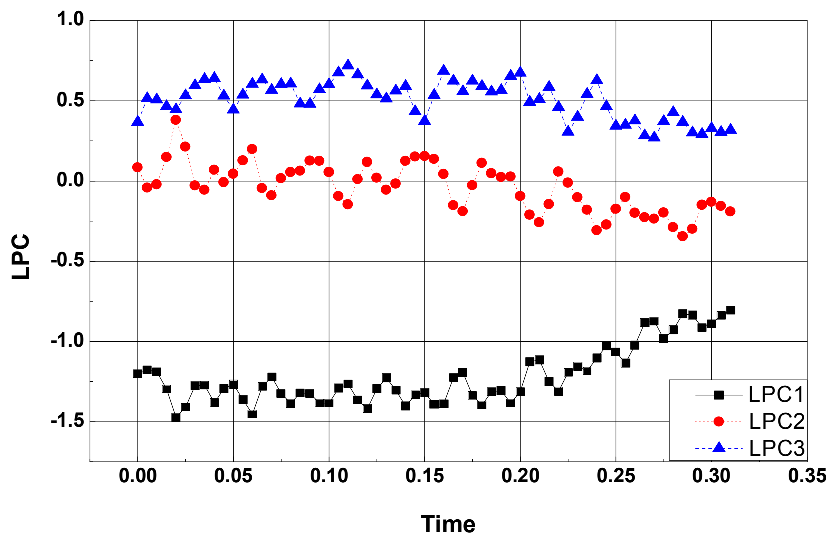


Figure A-1: Evaluated first three LPC trajectories for Bangla vowel (অ)

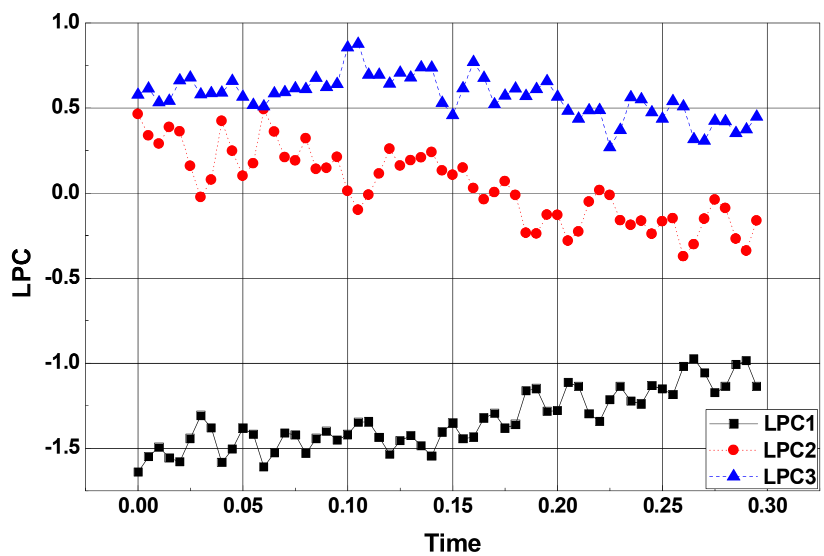


Figure A-2: Evaluated first three LPC trajectories for Bangla vowel (আ)

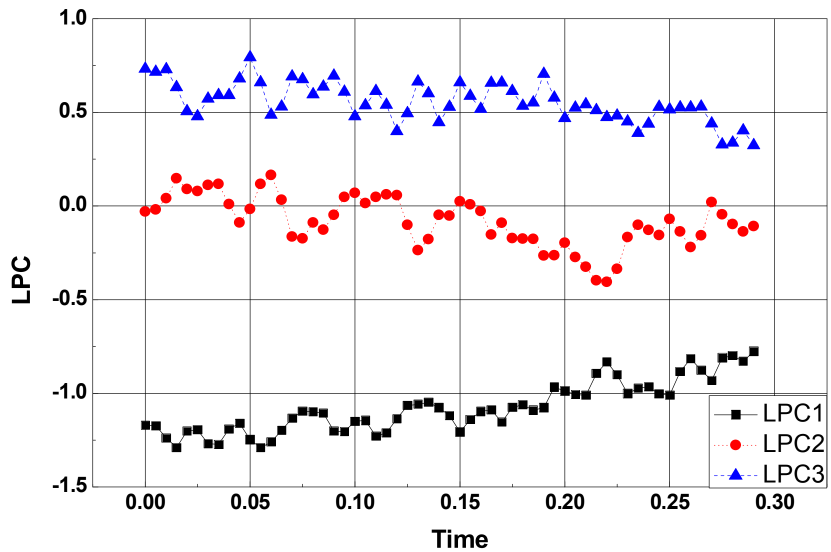


Figure A-3: Evaluated first three LPC trajectories for Bangla vowel (ই)

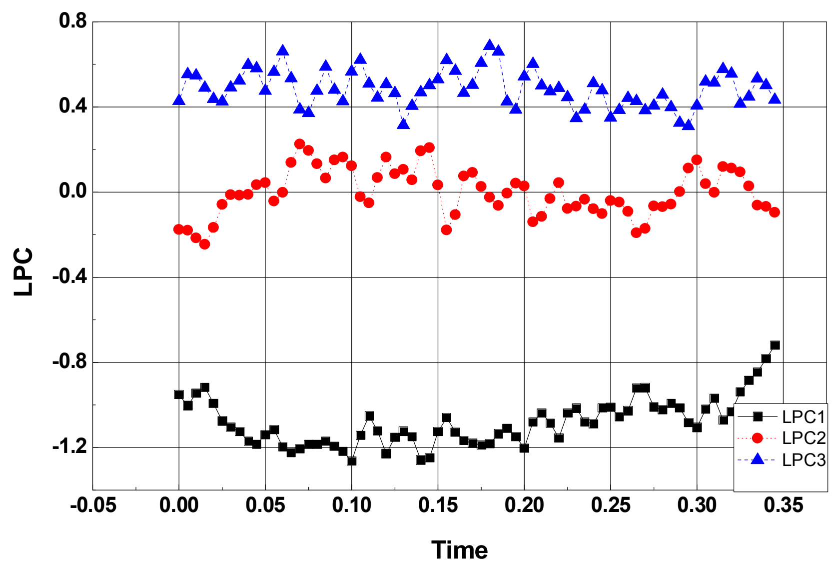


Figure A-4: Evaluated first three LPC trajectories for Bangla vowel (ঈ)

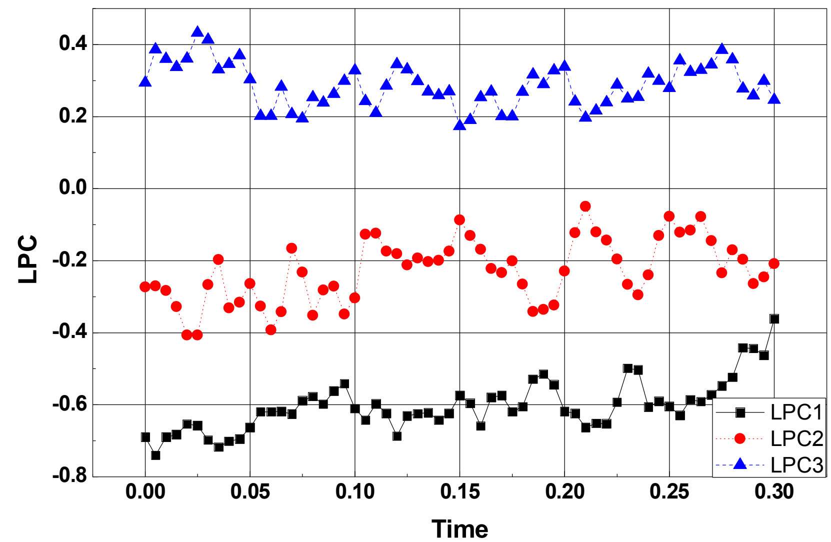


Figure A-5: Evaluated first three LPC trajectories for Bangla vowel (উ)

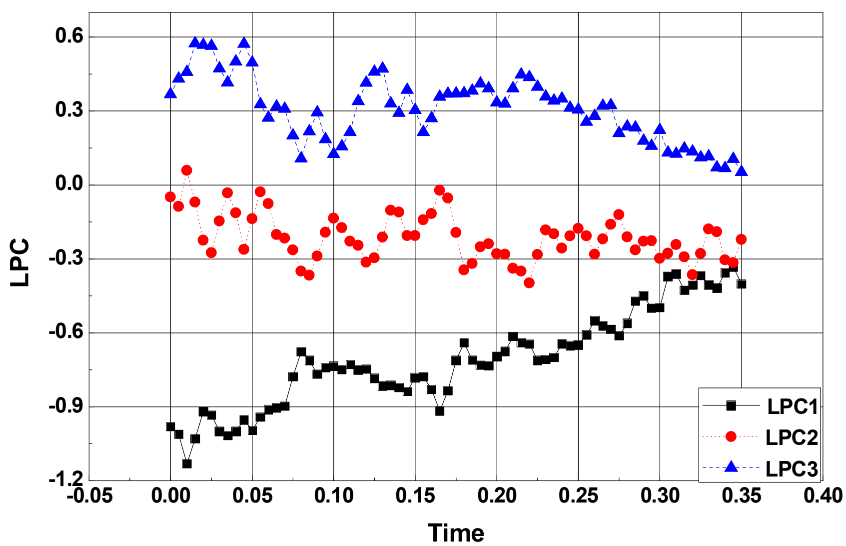


Figure A-6: Evaluated first three LPC trajectories for Bangla vowel (ঊ)

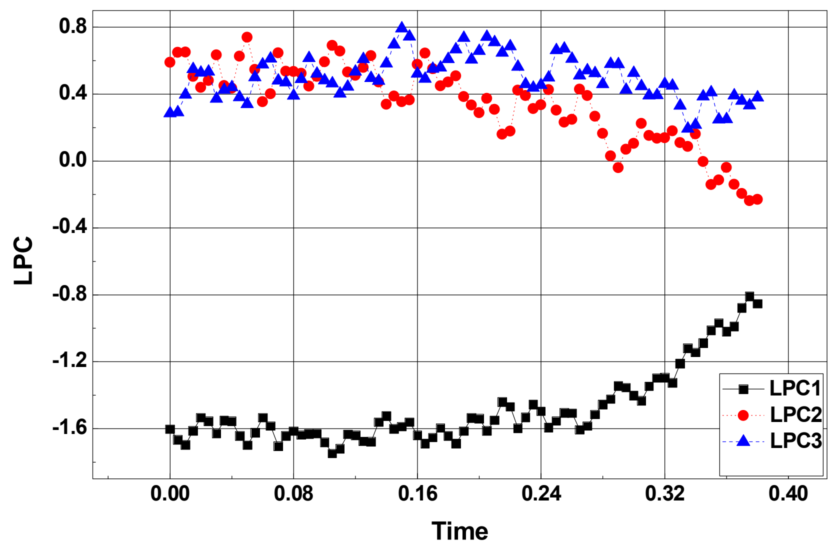


Figure A-7: Evaluated first three LPC trajectories for Bangla vowel (এ)

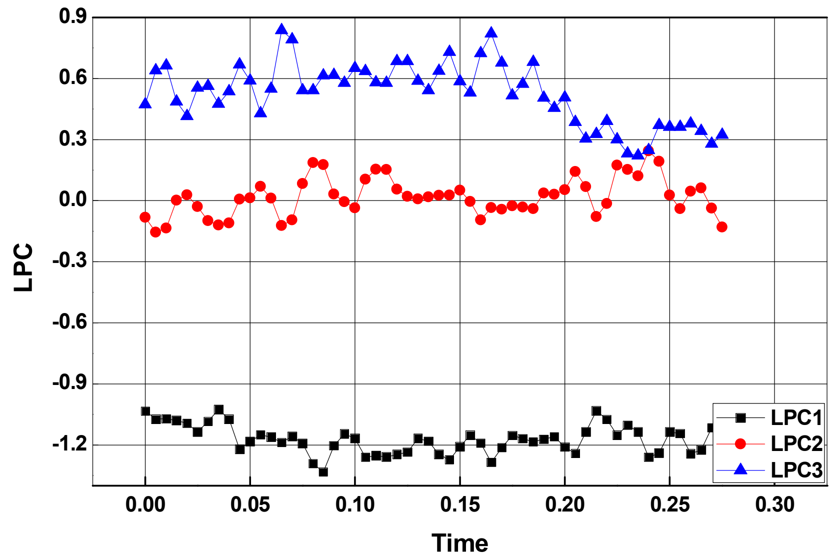


Figure A-8: Evaluated first three LPC trajectories for Bangla vowel (ঐ)

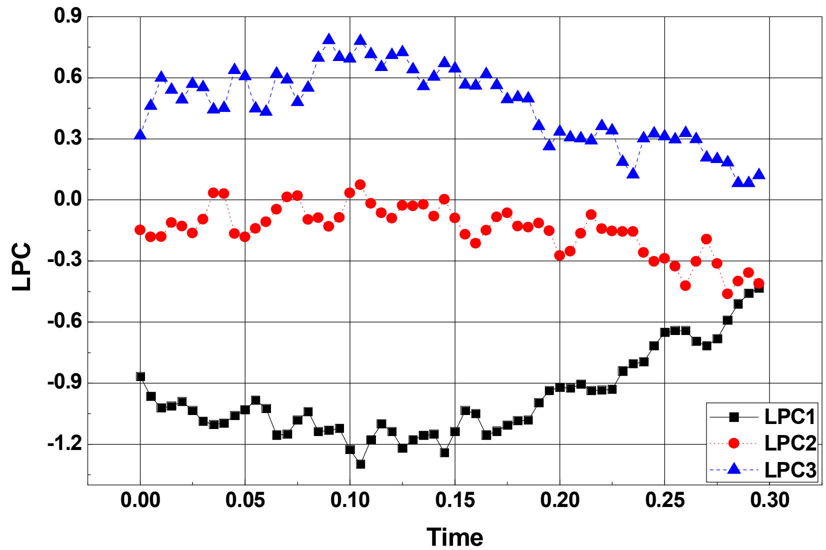


Figure A-9: Evaluated first three LPC trajectories for Bangla vowel (ও)

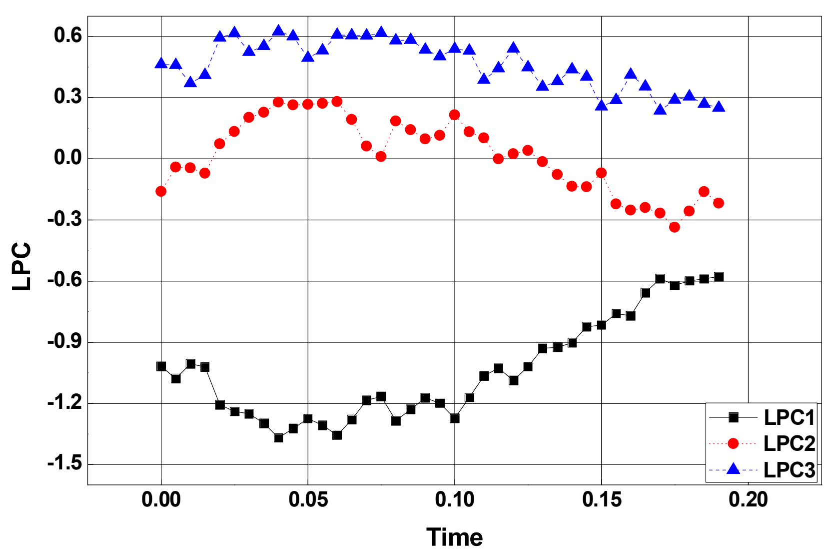


Figure A-10: Evaluated first three LPC trajectories for Bangla vowel (ঔ)

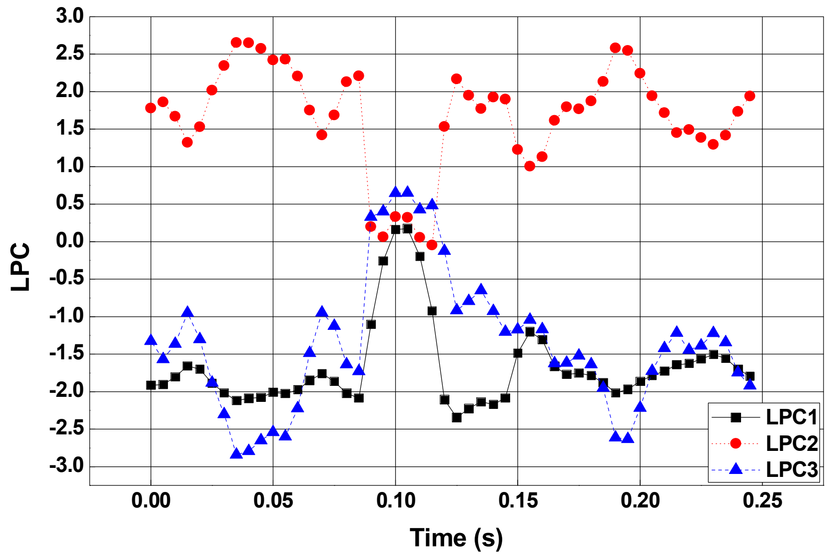


Figure A-11: Evaluated first three LPC trajectories for Bangla VCV sequence (আলো)

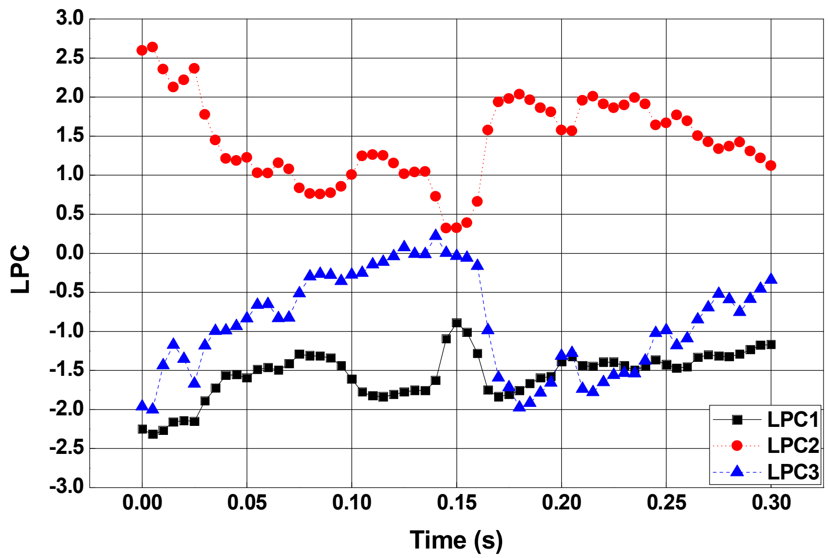


Figure A-12: Evaluated first three LPC trajectories for Bangla VCV sequence (আমি)

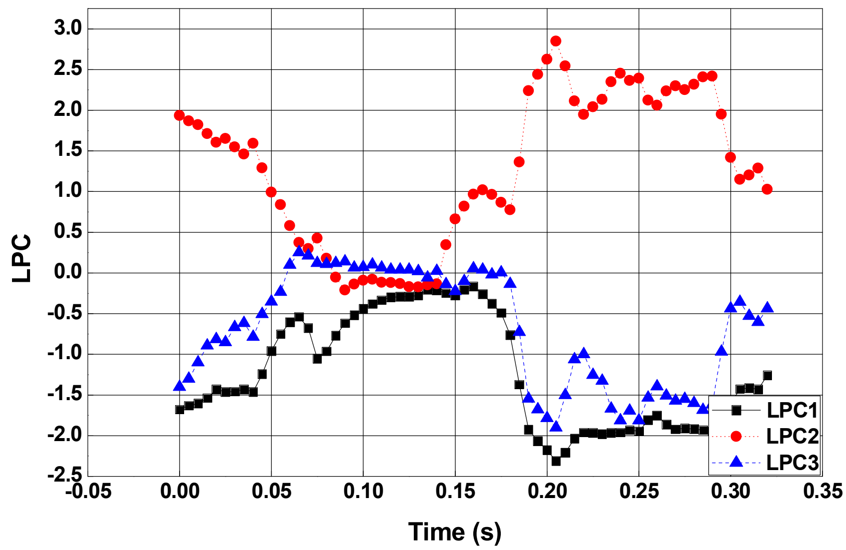


Figure A-13: Evaluated first three LPC trajectories for Bangla VCV sequence (ইতি)

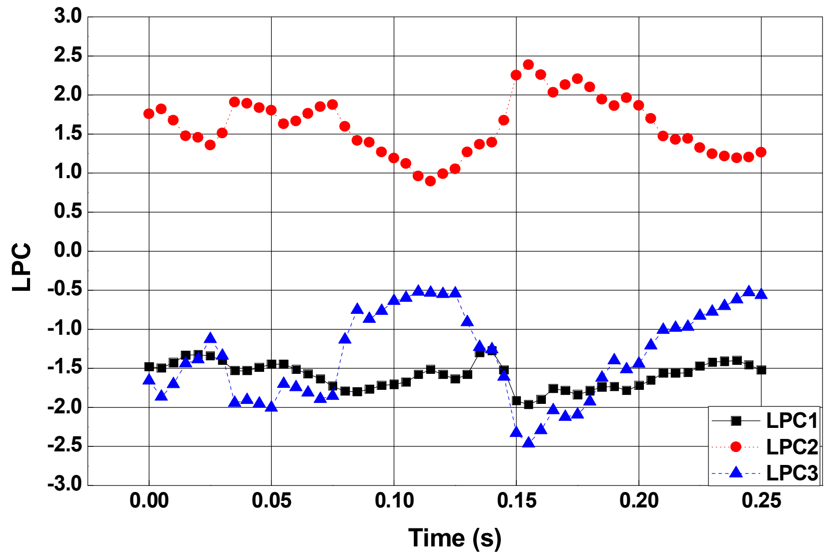


Figure A-14: Evaluated first three LPC trajectories for Bangla VCV sequence (অনু)

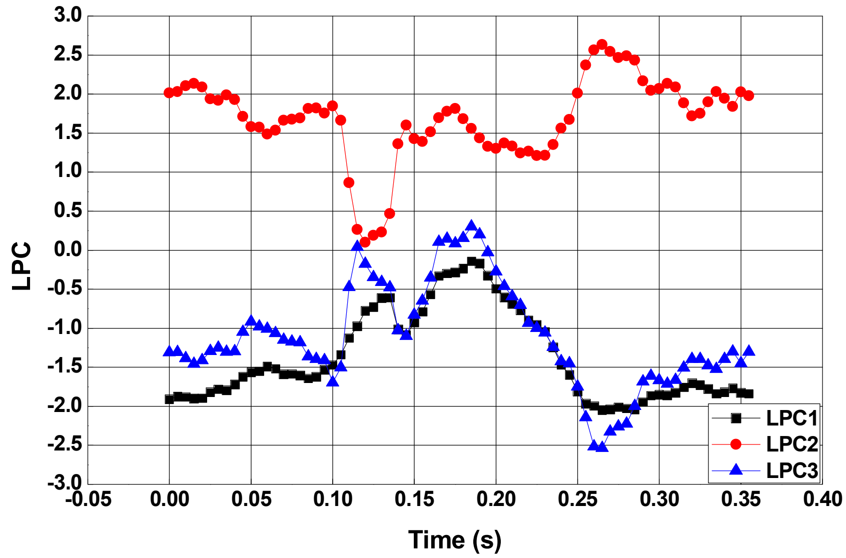


Figure A-15: Evaluated first three LPC trajectories for Bangla VCV sequence (উচু)

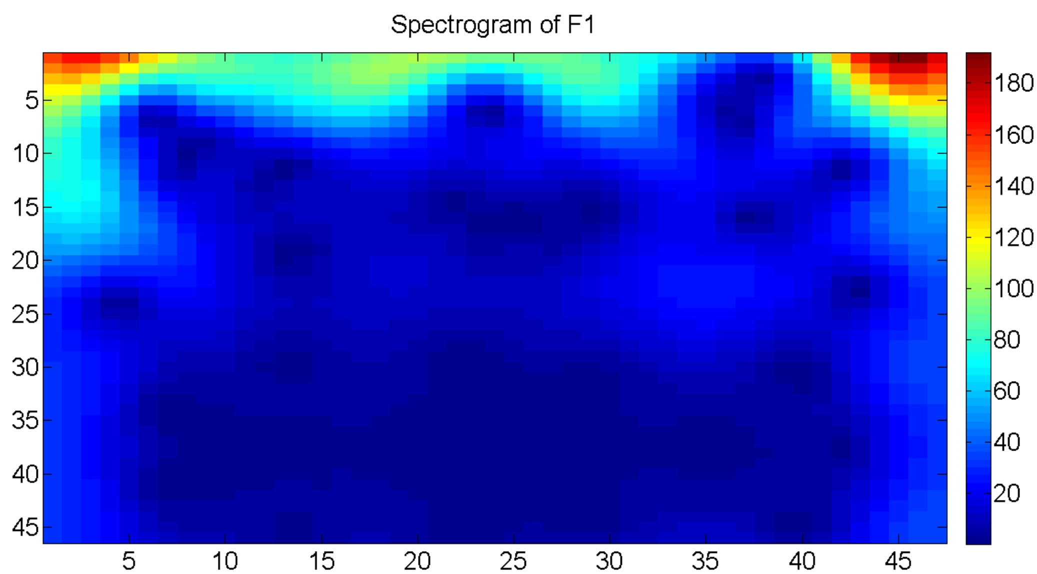


Figure A-16: Spectrogram of 1st formant trajectory for Bangla vowel (অ)

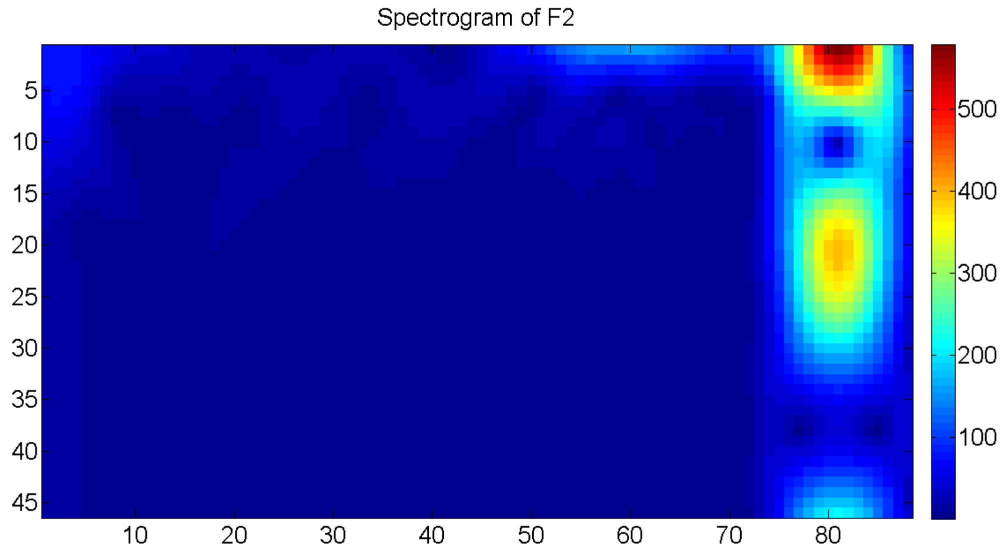


Figure A-17: Spectrogram of 2nd formant trajectory for Bangla vowel (অ)

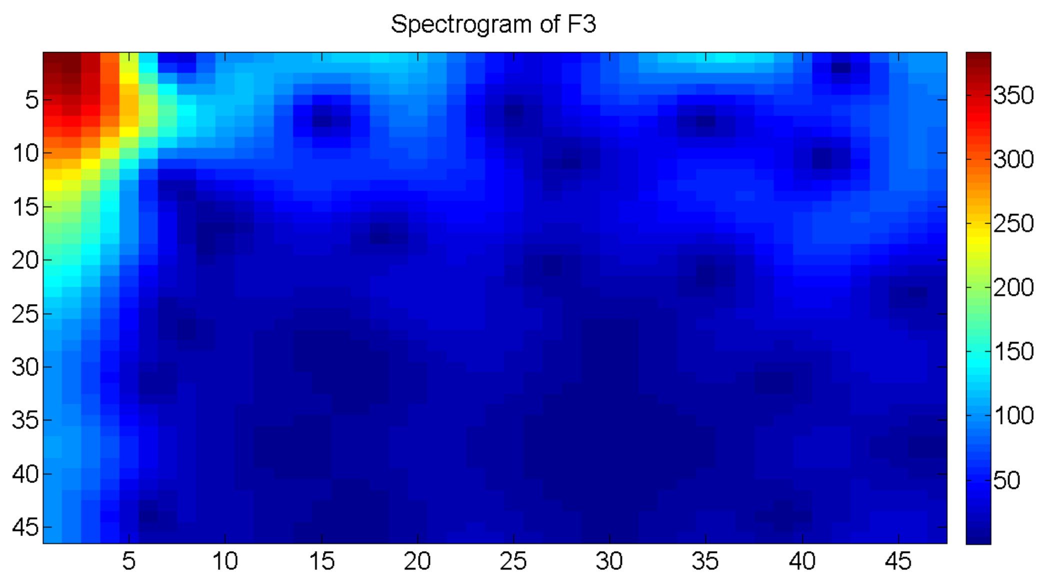


Figure A-18: Spectrogram of 3rd formant trajectory for Bangla vowel (অ)

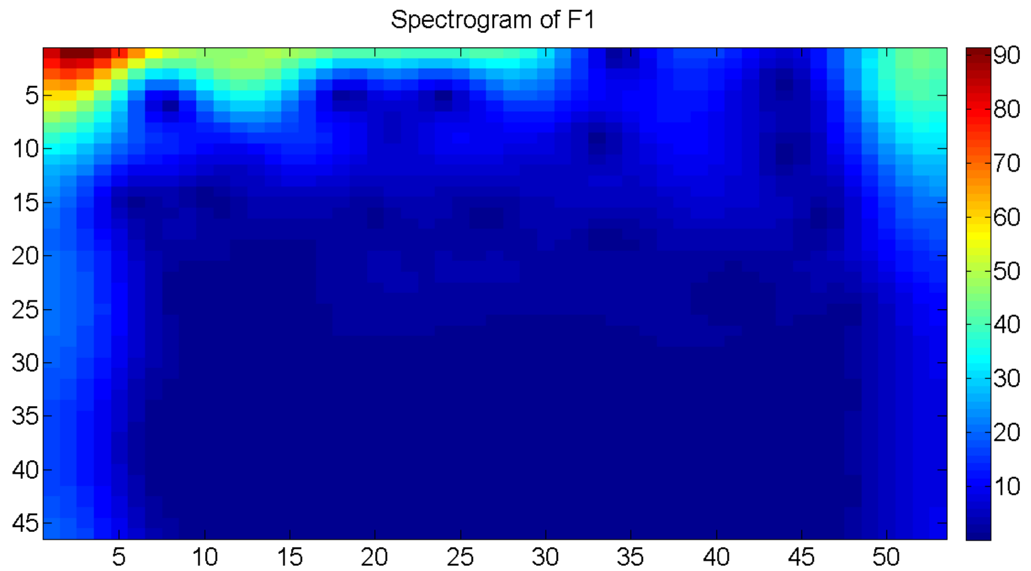


Figure A-19: Spectrogram of 1st formant trajectory for Bangla vowel (আ)

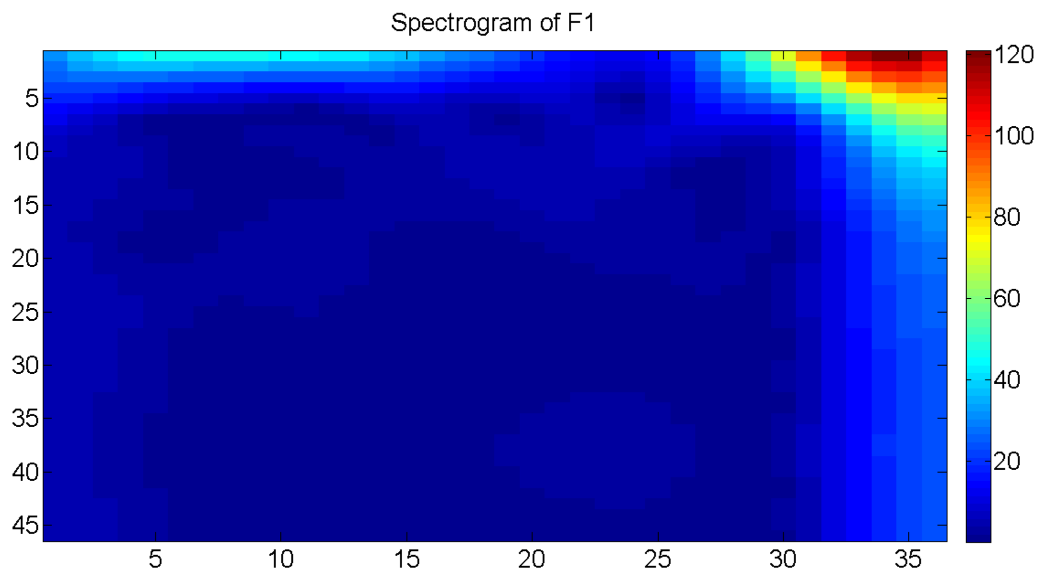


Figure A-20: Spectrogram of 1st formant trajectory for Bangla vowel (ই)

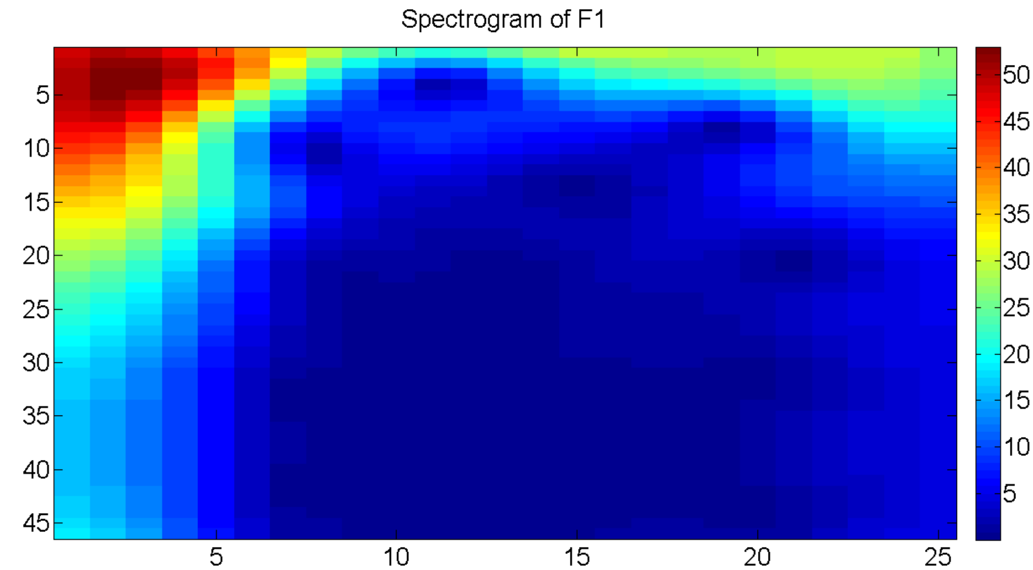


Figure A-21: Spectrogram of 1st formant trajectory for Bangla vowel (ঈ)

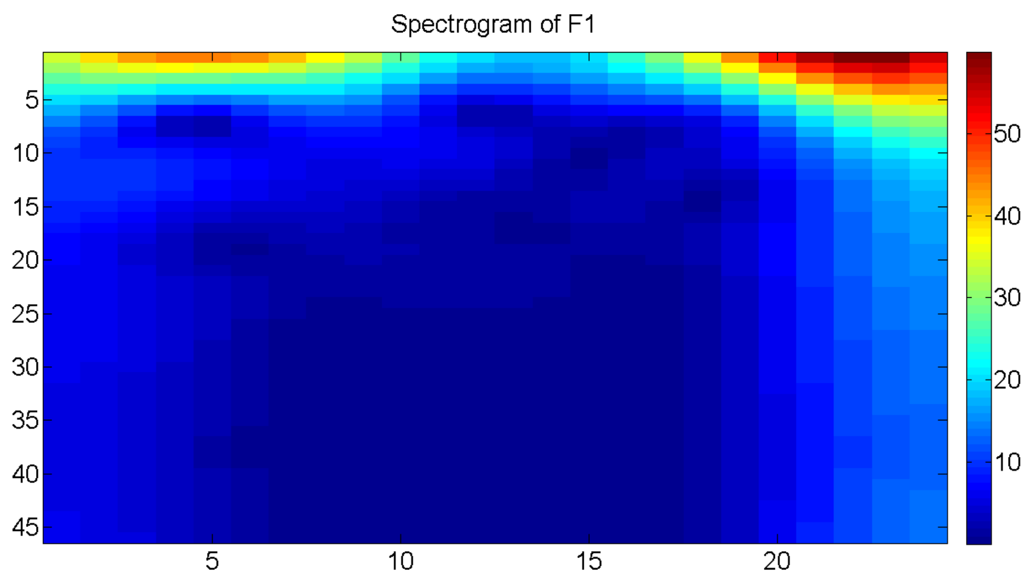


Figure A-22: Spectrogram of 1st formant trajectory for Bangla vowel (উ)

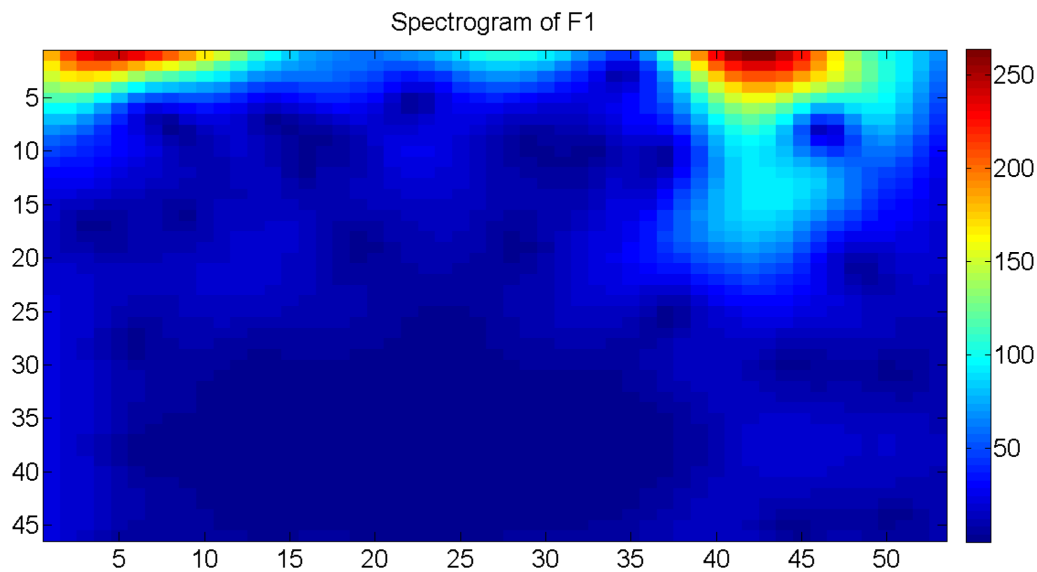


Figure A-23: Spectrogram of 1st formant trajectory for Bangla vowel (ঊ)

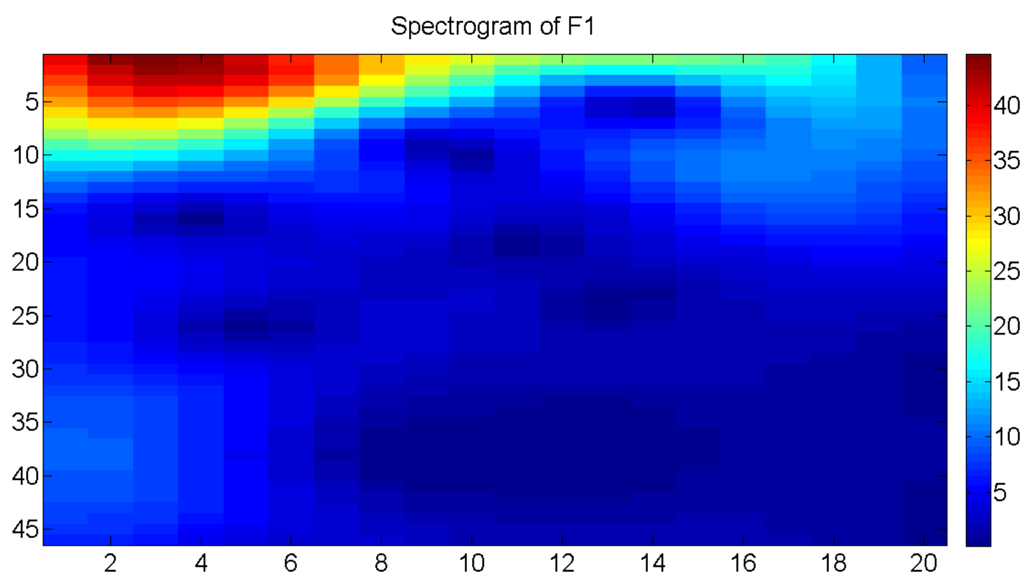


Figure A-24: Spectrogram of 1st formant trajectory for Bangla vowel (এ)

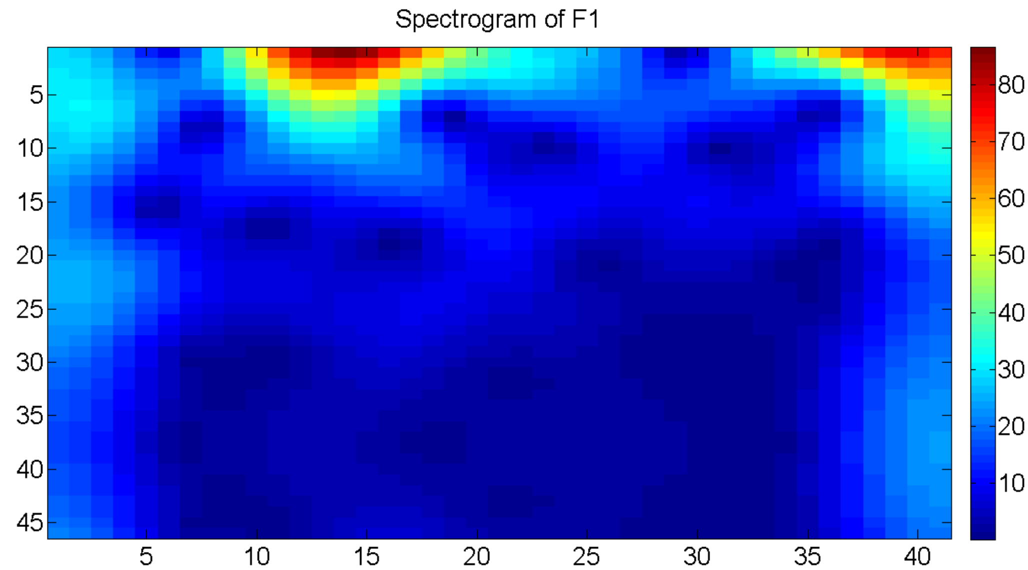


Figure A-25: Spectrogram of 1st formant trajectory for Bangla vowel (ঐ)

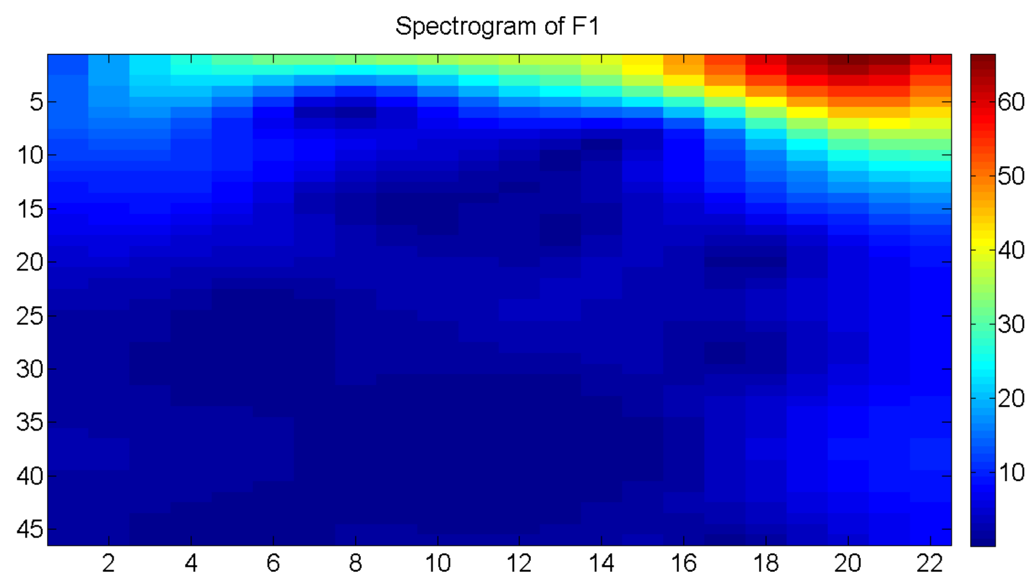


Figure A-26: Spectrogram of 1st formant trajectory for Bangla vowel (ও)

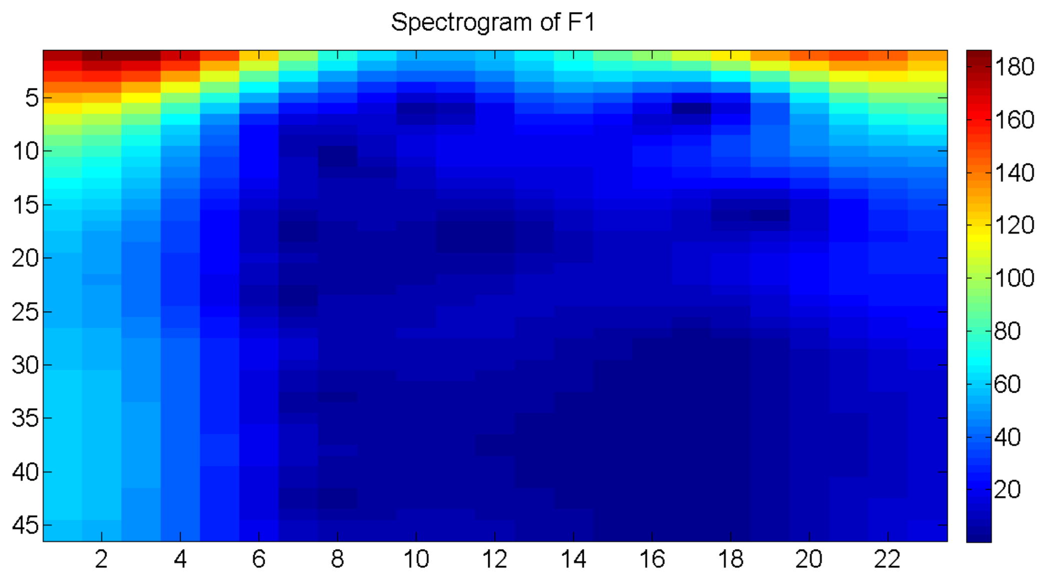


Figure A-27: Spectrogram of 1st formant trajectory for Bangla vowel (ঔ)

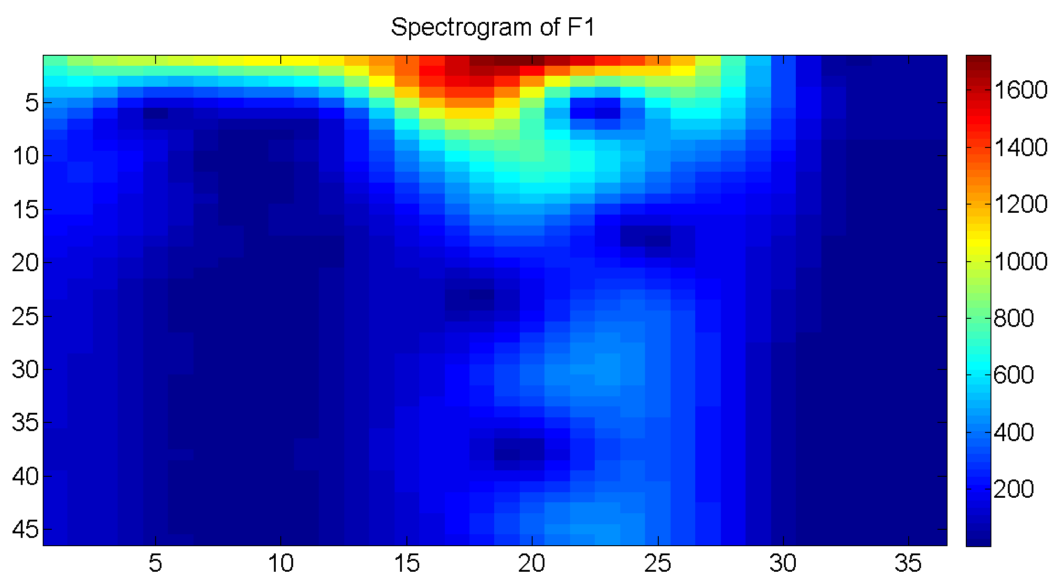


Figure A-28: Spectrogram of 1st formant trajectory for Bangla VCV sequence (আলো)

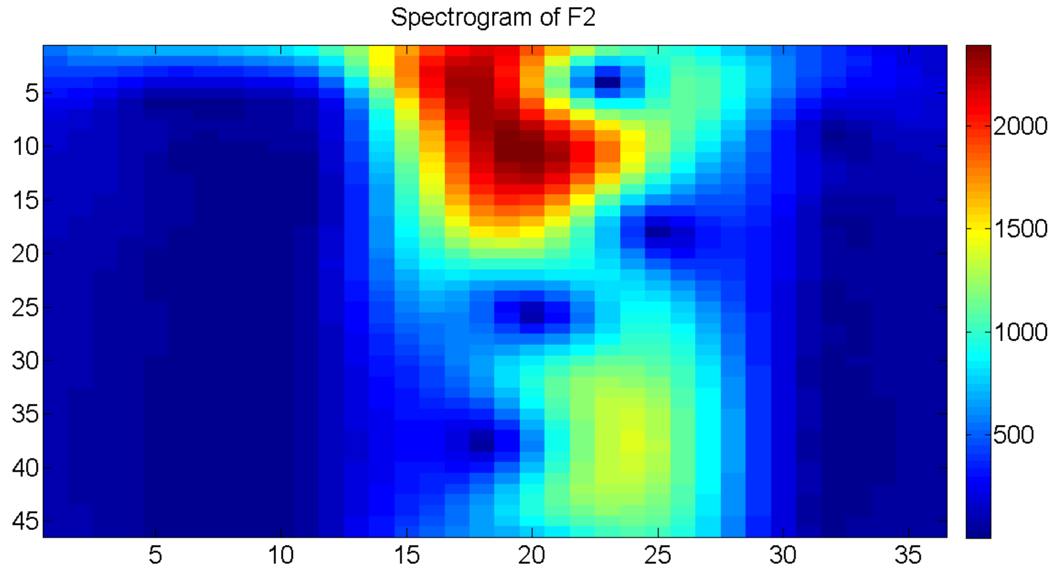


Figure A-29: Spectrogram of 2^{nd} formant trajectory for Bangla VCV sequence (আলো)

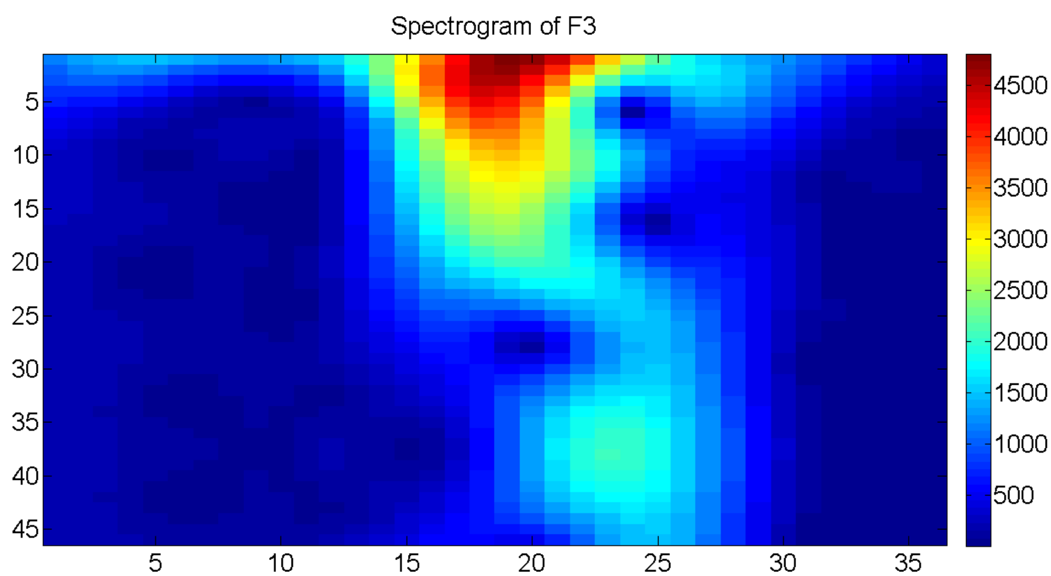


Figure A-30: Spectrogram of 3^{rd} formant trajectory for Bangla VCV sequence (আলো)

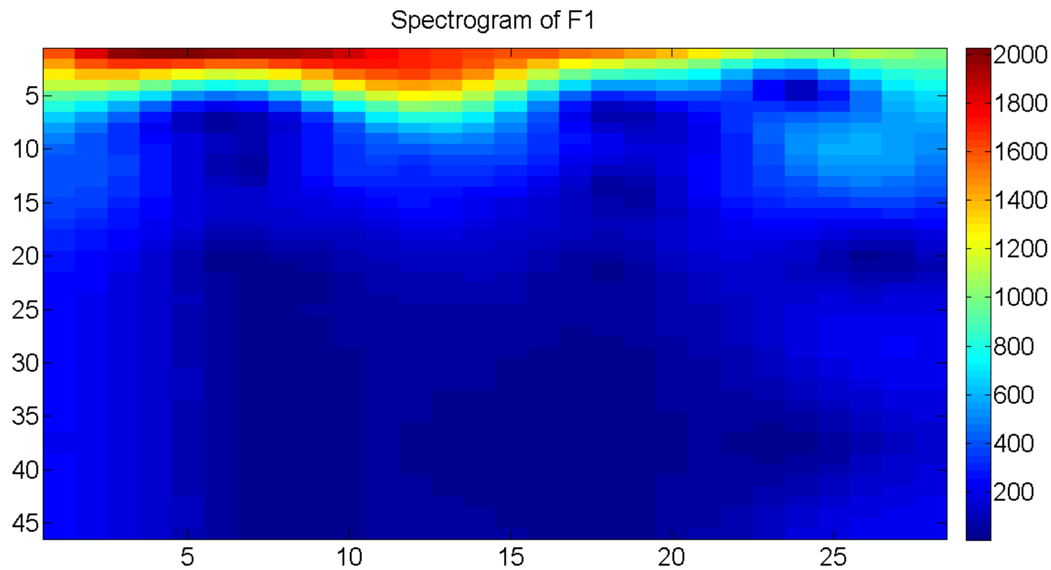


Figure A-31: Spectrogram of 1st formant trajectory for Bangla VCV sequence (আমি)

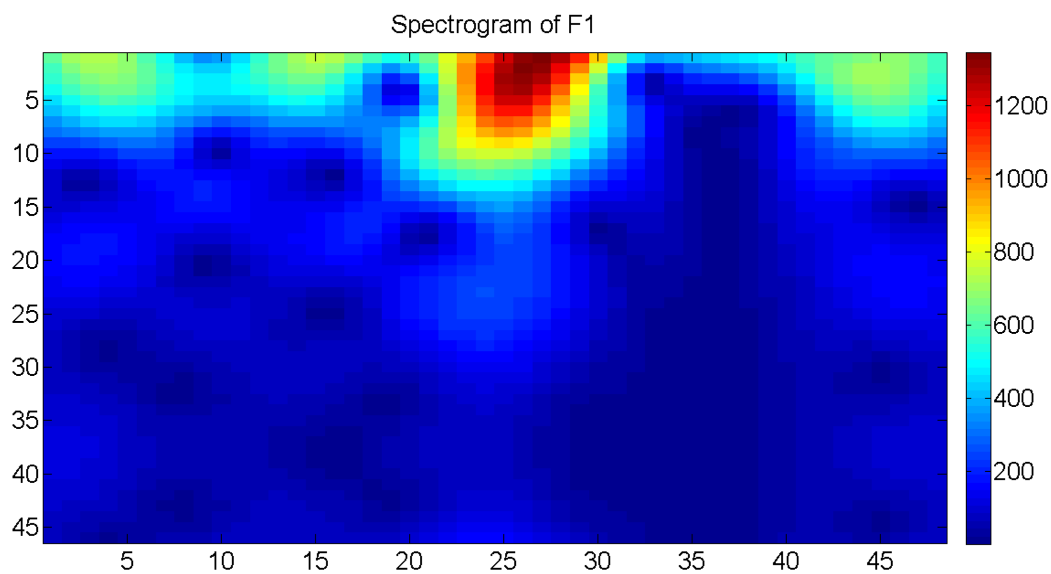


Figure A-32: Spectrogram of 1st formant trajectory for Bangla VCV sequence (ইতি)

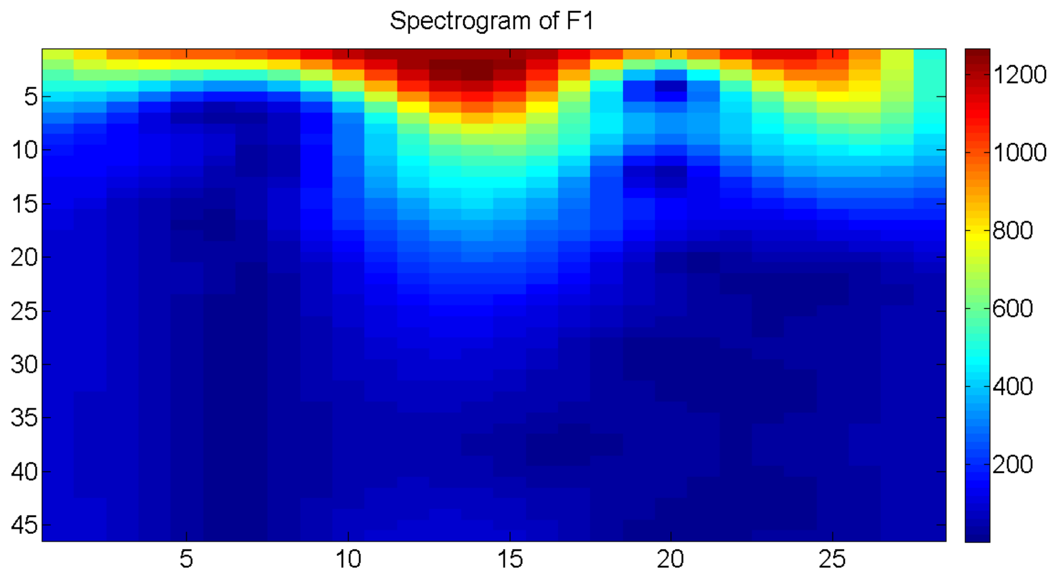


Figure A-33: Spectrogram of 1st formant trajectory for Bangla VCV sequence (অনু)

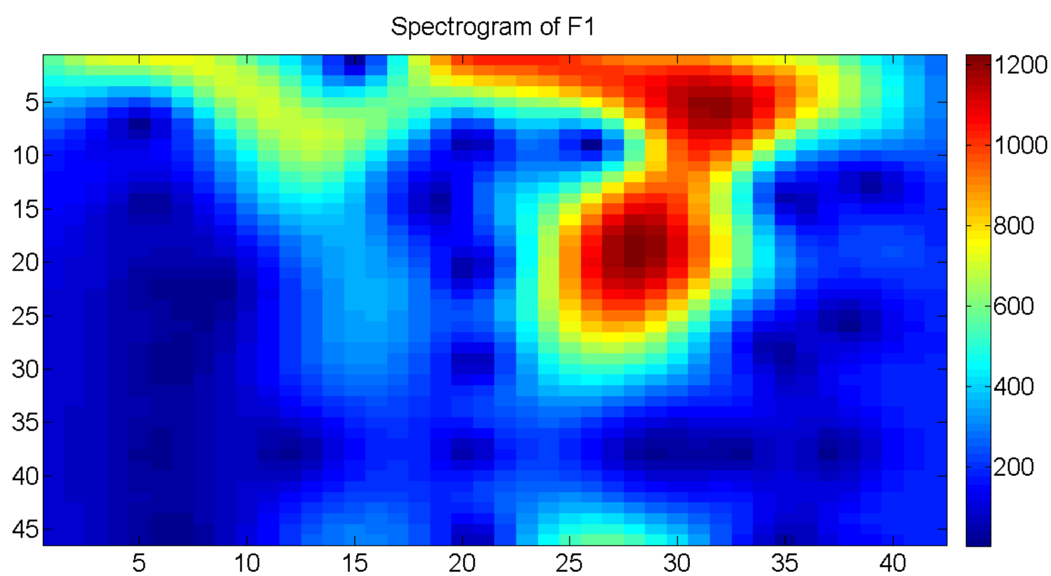


Figure A-34: Spectrogram of 1st formant trajectory for Bangla VCV sequence (উচু)

Appendix-B

Table B-1: Standard Deviation of LPC filtering coefficients of Bangla vowel (speaker 1)

	অ	আ	ই	ঈ	উ	ঊ	এ	ঐ	ও	ঔ
LPC1	0.1790	0.1713	0.1388	0.1118	0.0711	0.1957	0.2244	0.0721	0.2056	0.2480
LPC2	0.1514	0.2180	0.1345	0.1110	0.0874	0.0948	0.2408	0.0937	0.1218	0.1812
LPC3	0.1201	0.1250	0.1059	0.0855	0.0601	0.1325	0.1302	0.1519	0.1880	0.1210
LPC4	0.0960	0.1279	0.1035	0.1003	0.0622	0.0705	0.1239	0.0789	0.0794	0.1197
LPC5	0.1059	0.1208	0.0753	0.1095	0.0723	0.1233	0.1421	0.2161	0.0864	0.0917
LPC6	0.1009	0.1347	0.1056	0.1027	0.0894	0.0906	0.1605	0.1071	0.0851	0.0788
LPC7	0.0922	0.1132	0.0889	0.0914	0.0966	0.0815	0.1519	0.1077	0.0937	0.1052
LPC8	0.0972	0.0982	0.0964	0.0888	0.0632	0.0702	0.1045	0.0857	0.0878	0.0827
LPC9	0.0921	0.1321	0.0501	0.1034	0.0670	0.0701	0.1224	0.1016	0.0635	0.0723
LPC10	0.0906	0.0652	0.0900	0.0813	0.0572	0.0642	0.0786	0.0804	0.0817	0.0714
LPC11	0.0690	0.0978	0.0711	0.0813	0.0558	0.0805	0.1169	0.0958	0.0960	0.0898
LPC12	0.0884	0.0886	0.0991	0.0810	0.0745	0.0818	0.1222	0.1031	0.1121	0.1018
LPC13	0.1052	0.0933	0.0703	0.0865	0.0590	0.0532	0.0899	0.0646	0.0745	0.1225
LPC14	0.0674	0.1010	0.0949	0.1013	0.0515	0.0559	0.0877	0.0755	0.0933	0.0680
LPC15	0.0777	0.1179	0.0681	0.0690	0.0517	0.0601	0.0946	0.0941	0.0746	0.1162
LPC16	0.0784	0.0970	0.0791	0.1181	0.0825	0.0649	0.0596	0.0933	0.0679	0.0656

Table B-2: Standard Deviation of LPC filtering coefficients of Bangla vowel (speaker 2)

	অ	আ	ই	ঈ	উ	ঊ	এ	ঐ	ও	ঔ
LPC1	0.0881	0.066	0.1338	0.2181	0.0561	0.118	0.0809	0.1963	0.103	0.2525
LPC2	0.0789	0.1026	0.0864	0.3161	0.04	0.1005	0.1645	0.2196	0.1018	0.2276
LPC3	0.1114	0.0907	0.1372	0.1132	0.0577	0.0813	0.0845	0.1118	0.1184	0.1804
LPC4	0.0883	0.101	0.1161	0.1716	0.0752	0.0689	0.1253	0.1084	0.0917	0.0847
LPC5	0.0816	0.0748	0.0964	0.1121	0.0614	0.0921	0.1088	0.1415	0.0999	0.147
LPC6	0.0793	0.1142	0.0797	0.0954	0.0524	0.0681	0.1095	0.1391	0.0858	0.0889
LPC7	0.098	0.0842	0.1071	0.0882	0.0515	0.0753	0.105	0.1038	0.0654	0.1135
LPC8	0.0694	0.124	0.0884	0.1137	0.0566	0.0973	0.0733	0.1357	0.0902	0.1254
LPC9	0.0677	0.0844	0.0831	0.0969	0.0527	0.0642	0.0844	0.1676	0.0902	0.0883
LPC10	0.1105	0.1128	0.08	0.1067	0.0513	0.0617	0.1074	0.1362	0.079	0.1016
LPC11	0.0856	0.1169	0.0879	0.0951	0.0555	0.0915	0.1403	0.1253	0.0655	0.0752
LPC12	0.0902	0.0907	0.0803	0.1119	0.0591	0.0826	0.1073	0.1261	0.0836	0.1068
LPC13	0.0683	0.0864	0.0795	0.0959	0.0499	0.0783	0.0987	0.106	0.0768	0.0749
LPC14	0.0786	0.0792	0.0662	0.0966	0.0492	0.0875	0.1002	0.0874	0.0787	0.0632
LPC15	0.0574	0.0882	0.0582	0.112	0.0534	0.0671	0.0923	0.1306	0.0723	0.0577
LPC16	0.0412	0.0549	0.0395	0.0532	0.0474	0.0585	0.0654	0.0883	0.0437	0.043

Table B-3: Standard Deviation of LPC filtering coefficients of Bangla vowel (speaker 3)

	অ	আ	ই	ঈ	উ	ঊ	এ	ঐ	ও	ঔ
LPC1	0.3884	0.3765	0.3368	0.4445	0.2701	0.3651	0.2868	0.4116	0.372	0.4547
LPC2	0.1753	0.2134	0.1097	0.1409	0.254	0.0954	0.1055	0.2721	0.1693	0.314
LPC3	0.1263	0.1316	0.084	0.1218	0.1339	0.0776	0.1229	0.1206	0.133	0.1103
LPC4	0.1037	0.1061	0.0762	0.0939	0.1153	0.1383	0.0801	0.1038	0.2025	0.1511
LPC5	0.06	0.0611	0.0794	0.0683	0.1031	0.0627	0.0737	0.081	0.0902	0.1169
LPC6	0.1002	0.1005	0.0932	0.0907	0.0927	0.0563	0.0547	0.1155	0.0666	0.0663
LPC7	0.0623	0.1018	0.0966	0.0594	0.0817	0.0703	0.0828	0.1344	0.0747	0.1054
LPC8	0.0974	0.096	0.073	0.0834	0.115	0.1136	0.0764	0.0854	0.1106	0.1052
LPC9	0.0974	0.0829	0.0784	0.0671	0.1023	0.0768	0.0662	0.0729	0.1103	0.0888
LPC10	0.0653	0.0689	0.0797	0.0629	0.1212	0.0894	0.0614	0.1036	0.1125	0.1943
LPC11	0.0853	0.0791	0.0886	0.0578	0.0892	0.0793	0.0736	0.0601	0.0678	0.1288
LPC12	0.0559	0.1009	0.0622	0.0742	0.0853	0.0662	0.0622	0.0865	0.0701	0.0568
LPC13	0.0693	0.0694	0.0578	0.0726	0.0932	0.084	0.067	0.1104	0.1223	0.1215
LPC14	0.1023	0.0734	0.0829	0.1176	0.0531	0.0537	0.0721	0.0878	0.0979	0.0973
LPC15	0.1	0.1109	0.0873	0.1517	0.1099	0.0832	0.0661	0.1019	0.0642	0.0589
LPC16	0.0766	0.0924	0.1023	0.0827	0.1305	0.0949	0.0745	0.1359	0.0906	0.1169

Table B-4: Standard Deviation of LPC filtering coefficients of Bangla vowel (speaker 4)

	অ	আ	ই	ঈ	উ	ঊ	এ	ঐ	ও	ঔ
LPC1	0.215	0.3287	0.3973	0.4	0.2002	0.3264	0.4332	0.2968	0.3581	0.5875
LPC2	0.1078	0.0946	0.1441	0.1104	0.1855	0.2644	0.1781	0.1181	0.1522	0.2493
LPC3	0.1354	0.218	0.1437	0.2066	0.0587	0.0857	0.2095	0.106	0.1384	0.2121
LPC4	0.1809	0.2104	0.1137	0.1259	0.089	0.1218	0.1402	0.1442	0.1769	0.1235
LPC5	0.1008	0.1126	0.0917	0.0893	0.1199	0.1331	0.0931	0.0956	0.1209	0.096
LPC6	0.1006	0.1078	0.0913	0.1482	0.0742	0.0724	0.1229	0.1046	0.1008	0.1142
LPC7	0.1112	0.1709	0.0956	0.1412	0.0784	0.0617	0.1932	0.1402	0.1169	0.0768
LPC8	0.1018	0.1086	0.0683	0.0806	0.0568	0.0467	0.0953	0.0761	0.0746	0.0633
LPC9	0.1055	0.0837	0.0742	0.0894	0.0583	0.0681	0.0821	0.0814	0.0696	0.1244
LPC10	0.1102	0.1619	0.124	0.1217	0.0886	0.0931	0.1733	0.1184	0.111	0.1388
LPC11	0.1177	0.1041	0.0965	0.1098	0.0769	0.1131	0.1101	0.1075	0.116	0.104
LPC12	0.0716	0.095	0.1201	0.0722	0.0628	0.1048	0.083	0.0687	0.1123	0.0947
LPC13	0.0874	0.1102	0.0572	0.0753	0.0662	0.0703	0.1068	0.044	0.0853	0.0655
LPC14	0.1181	0.1198	0.0521	0.0587	0.0653	0.0771	0.0851	0.062	0.1331	0.0872
LPC15	0.0913	0.0838	0.0606	0.0554	0.0519	0.071	0.0824	0.0714	0.1001	0.0748
LPC16	0.0607	0.0441	0.0678	0.0723	0.0635	0.063	0.0722	0.1058	0.0497	0.0594

Table B-5: Standard Deviation of LPC filtering coefficients of Bangla vowel (speaker 5)

	অ	আ	ই	ঈ	উ	ঊ	এ	ঐ	ও	ঔ
LPC1	0.3204	0.3928	0.377	0.2574	0.2058	0.2202	0.2822	0.3557	0.2751	0.4792
LPC2	0.2707	0.3013	0.2182	0.1825	0.2719	0.2285	0.2588	0.2822	0.2669	0.2498
LPC3	0.2026	0.2133	0.1026	0.1097	0.2122	0.2047	0.1605	0.1942	0.2175	0.2005
LPC4	0.1351	0.1436	0.0776	0.0849	0.1625	0.1488	0.0958	0.1333	0.1569	0.1509
LPC5	0.1196	0.1154	0.0957	0.072	0.1292	0.1155	0.082	0.0943	0.118	0.1203
LPC6	0.1051	0.1001	0.0629	0.057	0.0966	0.0997	0.077	0.0892	0.0845	0.1063
LPC7	0.099	0.0794	0.0743	0.0563	0.0826	0.0975	0.0835	0.0761	0.0806	0.0961
LPC8	0.0947	0.0734	0.078	0.0623	0.0729	0.0931	0.0686	0.0735	0.069	0.0765
LPC9	0.0895	0.0753	0.1085	0.0723	0.0819	0.0948	0.0999	0.0943	0.0812	0.0875
LPC10	0.0612	0.0498	0.0923	0.0807	0.0747	0.0793	0.0749	0.0902	0.0704	0.0859
LPC11	0.0707	0.061	0.1184	0.1112	0.0915	0.0889	0.0921	0.0987	0.0833	0.0901
LPC12	0.058	0.0531	0.0913	0.097	0.0804	0.0696	0.054	0.0926	0.0576	0.068
LPC13	0.0574	0.0544	0.082	0.0761	0.0738	0.0572	0.0609	0.0864	0.05	0.066
LPC14	0.06	0.0485	0.0729	0.0595	0.0507	0.0548	0.0736	0.0817	0.056	0.0587
LPC15	0.059	0.0474	0.0598	0.0476	0.076	0.0678	0.0533	0.0613	0.0629	0.066
LPC16	0.062	0.0491	0.0543	0.0523	0.0821	0.061	0.0636	0.0556	0.0759	0.0681

Table B-6: Standard Deviation of LPC filtering coefficients of Bangla VCV sequences (speaker 1)

	আলো	আমি	ইতি	অনু	উচু
LPC1	0.56679	0.30746	0.68854	0.1706	0.5756
LPC2	0.6934	0.54366	0.93695	0.36894	0.53422
LPC3	0.90495	0.63572	0.70631	0.57321	0.66924
LPC4	0.68136	0.50762	0.27914	0.50032	0.45077
LPC5	0.47722	0.45308	0.32272	0.40872	0.43955
LPC6	0.37184	0.47005	0.41398	0.30203	0.50408
LPC7	0.38491	0.40025	0.29416	0.28183	0.38498
LPC8	0.40453	0.3508	0.22937	0.28313	0.47968
LPC9	0.36729	0.26904	0.24312	0.32487	0.32927
LPC10	0.33113	0.26884	0.22935	0.24537	0.41462
LPC11	0.30797	0.29115	0.24169	0.24495	0.28631
LPC12	0.29571	0.34642	0.29836	0.31255	0.30742
LPC13	0.27159	0.30678	0.22293	0.40954	0.21318
LPC14	0.20714	0.289	0.18666	0.44321	0.18468
LPC15	0.18418	0.26146	0.17065	0.33859	0.18842
LPC16	0.11163	0.13836	0.08794	0.16533	0.11358

Table B-7: Standard Deviation of LPC filtering coefficients of Bangla VCV sequence (speaker 2)

	আলো	আমি	ইতি	অনু	উচু
LPC1	0.35078	0.27487	0.40327	0.29849	0.39556
LPC2	0.48059	0.57534	0.48701	0.56144	0.6154
LPC3	0.49282	0.63368	0.36427	0.56953	0.481
LPC4	0.51014	0.47101	0.25573	0.4367	0.42378
LPC5	0.46546	0.47501	0.21399	0.40471	0.17669
LPC6	0.44784	0.40187	0.26761	0.40777	0.19389
LPC7	0.41137	0.38442	0.19438	0.32436	0.17551
LPC8	0.45574	0.41862	0.16578	0.30043	0.23319
LPC9	0.29299	0.32863	0.1634	0.38015	0.17146
LPC10	0.24612	0.49301	0.17788	0.41372	0.25923
LPC11	0.44467	0.45754	0.18669	0.34171	0.27404
LPC12	0.40275	0.34815	0.21669	0.36063	0.25412
LPC13	0.31593	0.3835	0.27487	0.37176	0.23599
LPC14	0.28688	0.3309	0.18603	0.47194	0.27992
LPC15	0.26132	0.25989	0.23164	0.3778	0.19844
LPC16	0.15677	0.15021	0.13285	0.17366	0.11297

Table B-8: Standard Deviation of LPC filtering coefficients of Bangla VCV sequence (speaker 3)

	আলো	আমি	ইতি	অনু	উচু
LPC1	0.29686	0.26178	0.68524	0.32172	0.58334
LPC2	0.66942	0.4572	1.00982	0.72605	0.67915
LPC3	0.77243	0.51538	0.99866	0.90526	0.93277
LPC4	0.64196	0.55738	0.87914	0.86762	0.66143
LPC5	0.38999	0.49011	0.87139	0.80286	0.51637
LPC6	0.38984	0.47658	0.79698	0.80863	0.36776
LPC7	0.38455	0.35114	0.62821	0.76274	0.37292
LPC8	0.32329	0.30794	0.46097	0.58287	0.4061
LPC9	0.30018	0.35384	0.47834	0.39723	0.38371
LPC10	0.28398	0.30275	0.48291	0.26488	0.29674
LPC11	0.24256	0.24719	0.38813	0.26396	0.32144
LPC12	0.21227	0.24723	0.31007	0.36953	0.31385
LPC13	0.2502	0.28088	0.25324	0.39903	0.2785
LPC14	0.26999	0.28886	0.25338	0.33081	0.19555
LPC15	0.19543	0.25922	0.1974	0.22578	0.10602
LPC16	0.10684	0.1948	0.13523	0.14186	0.08115

Table B-9: Standard Deviation of LPC filtering coefficients of Bangla VCV sequence (speaker 4)

	আলো	আমি	ইতি	অনু	উচু
LPC1	0.24788	0.16884	0.60024	0.15807	0.50648
LPC2	0.35183	0.48497	0.90484	0.44182	0.76637
LPC3	0.37432	0.65257	0.90526	0.68743	0.62924
LPC4	0.50586	0.69855	0.78169	0.77577	0.44656
LPC5	0.41712	0.51154	0.70738	0.52559	0.30717
LPC6	0.34691	0.57577	0.57788	0.40309	0.31812
LPC7	0.2656	0.70485	0.42887	0.436	0.22089
LPC8	0.24717	0.64568	0.29994	0.39531	0.23183
LPC9	0.21993	0.41647	0.26621	0.30219	0.20048
LPC10	0.22431	0.24211	0.36784	0.32156	0.28758
LPC11	0.24633	0.28469	0.3451	0.38688	0.18945
LPC12	0.21237	0.35814	0.39275	0.30428	0.24014
LPC13	0.13635	0.37132	0.31924	0.29377	0.2682
LPC14	0.15421	0.33444	0.25892	0.26837	0.21399
LPC15	0.18889	0.22373	0.15028	0.18143	0.12902
LPC16	0.11465	0.10992	0.09935	0.09678	0.11351

Table B-10: Standard Deviation of LPC filtering coefficients of Bangla VCV sequence (speaker 5)

	আলো	আমি	ইতি	অনু	উচু
LPC1	0.26614	0.4166	0.90514	0.31757	0.9233
LPC2	0.27658	0.55365	0.74732	0.32721	0.59669
LPC3	0.2617	0.17857	0.31701	0.18355	0.37743
LPC4	0.11534	0.30557	0.21882	0.15447	0.25903
LPC5	0.25436	0.08994	0.23068	0.12012	0.15239
LPC6	0.22681	0.22093	0.13989	0.16694	0.12849
LPC7	0.11602	0.2071	0.23122	0.12054	0.19959
LPC8	0.261	0.08024	0.12558	0.10681	0.13715
LPC9	0.12995	0.12882	0.1226	0.11133	0.09653
LPC10	0.13513	0.17209	0.08527	0.12569	0.13631
LPC11	0.13689	0.13253	0.10715	0.15709	0.14278
LPC12	0.05257	0.14366	0.11311	0.10901	0.10196
LPC13	0.10304	0.14215	0.12194	0.11573	0.08639
LPC14	0.09456	0.11608	0.16023	0.121	0.09448
LPC15	0.10459	0.10592	0.10864	0.11694	0.15883
LPC16	0.06594	0.10266	0.0875	0.13887	0.08369

Table B-11: Formant trajectory transitional energy for Bangla vowel (অ)

Speaker Name	Variation Tendency				
	F1	F2	F3	F4	F5
Speaker 1	4.09E+05	2.55E+05	8.90E+05	9.14E+05	9.87E+05
Speaker 2	1.63E+05	1.56E+05	3.06E+05	2.56E+05	6.90E+05
Speaker 3	2.59E+05	2.93E+05	5.70E+05	5.16E+05	8.14E+05
Speaker 4	7.36E+04	1.85E+05	2.09E+05	1.52E+05	3.09E+05
Speaker 5	3.44E+04	4.39E+04	3.32E+05	1.61E+05	2.61E+05

Table B-12: Formant trajectory transitional energy for Bangla vowel (আ)

Speaker Name	Variation Tendency				
	F1	F2	F3	F4	F5
Speaker 1	7.57E+04	5.28E+05	1.76E+05	1.92E+05	5.22E+05
Speaker 2	7.14E+04	1.09E+05	2.88E+05	2.20E+05	4.83E+05
Speaker 3	5.73E+05	1.72E+05	3.00E+05	6.44E+05	7.47E+05
Speaker 4	5.28E+04	4.76E+04	1.53E+05	4.77E+05	1.14E+05
Speaker 5	2.13E+05	1.41E+05	2.73E+05	1.95E+05	1.63E+05

Table B-13: Formant trajectory transitional energy for Bangla vowel (ই)

Speaker Name	Variation Tendency				
	F1	F2	F3	F4	F5
Speaker 1	4.04E+04	1.34E+05	1.99E+05	1.75E+05	4.20E+05
Speaker 2	6.01E+04	1.36E+05	1.74E+05	8.08E+05	3.69E+05
Speaker 3	5.98E+04	1.50E+05	7.53E+05	6.15E+05	8.64E+05
Speaker 4	3.05E+04	4.32E+05	3.05E+05	8.89E+05	3.48E+05
Speaker 5	1.58E+04	1.45E+05	3.57E+05	5.75E+05	8.76E+05

Table B-14: Formant trajectory transitional energy for Bangla vowel (ঈ)

Speaker Name	Variation Tendency				
	F1	F2	F3	F4	F5
Speaker 1	2.04E+04	2.25E+05	2.04E+05	3.58E+05	5.5E+05
Speaker 2	1.67E+04	1.95E+05	3.57E+05	4.57E+05	4.6E+05
Speaker 3	1.80E+04	3.47E+05	8.80E+05	5.30E+05	6.20E+05
Speaker 4	2.50E+04	2.20E+05	7.28E+05	3.77E+05	5.15E+05
Speaker 5	1.38E+04	1.87E+05	2.69E+05	4.63E+05	7.81E+05

Table B-15: Formant trajectory transitional energy for Bangla vowel (উ)

Speaker Name	Variation Tendency				
	F1	F2	F3	F4	F5
Speaker 1	1.94E+04	8.88E+04	2.58E+05	3.93E+05	1.89E+05
Speaker 2	2.20E+05	4.14E+05	4.47E+05	6.50E+05	3.92E+05
Speaker 3	1.36E+04	3.64E+05	6.42E+05	6.50E+05	2.33E+05
Speaker 4	2.10E+05	4.19E+05	3.21E+05	4.16E+05	4.67E+05
Speaker 5	1.73E+05	3.45E+05	5.54E+05	3.84E+05	6.79E+05

Table B-16: Formant trajectory transitional energy for Bangla vowel (ঊ)

Speaker Name	Variation Tendency				
	F1	F2	F3	F4	F5
Speaker 1	5.07E+04	9.91E+04	1.76E+05	2.40E+05	2.63E+05
Speaker 2	1.78E+05	2.67E+05	4.47E+05	7.86E+05	1.85E+05
Speaker 3	8.54E+04	3.01E+05	7.87E+05	7.89E+05	1.22E+05
Speaker 4	1.14E+05	2.37E+05	7.57E+05	3.52E+05	1.87E+05
Speaker 5	1.12E+05	2.37E+05	7.70E+05	3.52E+05	1.85E+05

Table B-17: Formant trajectory transitional energy for Bangla vowel (এ)

Speaker Name	Variation Tendency				
	F1	F2	F3	F4	F5
Speaker 1	1.14E+05	2.64E+05	5.31E+05	5.49E+05	4.24E+05
Speaker 2	1.61E+04	6.76E+04	1.27E+05	1.56E+05	8.39E+04
Speaker 3	2.07E+04	1.12E+05	2.01E+05	1.47E+05	9.73E+05
Speaker 4	1.52E+05	3.49E+05	5.18E+05	8.52E+05	7.71E+05
Speaker 5	8.17E+04	5.91E+04	1.93E+05	1.16E+05	2.82E+05

Table B-18: Formant trajectory transitional energy for Bangla vowel (ঐ)

Speaker Name	Variation Tendency				
	F1	F2	F3	F4	F5
Speaker 1	8.19E+04	5.96E+05	6.92E+05	3.76E+05	2.30E+05
Speaker 2	1.35E+05	7.03E+05	4.94E+05	6.07E+05	8.33E+05
Speaker 3	3.99E+05	1.73E+05	7.02E+05	4.46E+05	8.01E+05
Speaker 4	1.91E+05	2.20E+05	4.36E+05	4.38E+05	4.40E+05
Speaker 5	1.93E+05	2.45E+05	6.44E+05	1.26E+05	6.15E+05

Table B-19: Formant trajectory transitional energy for Bangla vowel (ও)

Speaker Name	Variation Tendency				
	F1	F2	F3	F4	F5
Speaker 1	3.36E+05	3.64E+05	3.41E+05	3.04E+05	6.53E+05
Speaker 2	8.57E+04	3.59E+05	5.04E+05	3.70E+05	6.82E+05
Speaker 3	6.02E+04	9.04E+04	5.04E+05	3.59E+05	4.18E+05
Speaker 4	9.73E+04	2.20E+05	2.79E+05	1.82E+05	5.24E+05
Speaker 5	5.07E+04	5.28E+05	4.00E+05	2.27E+05	1.74E+05

Table B-20: Formant trajectory transitional energy for Bangla vowel (ঔ)

Speaker Name	Variation Tendency				
	F1	F2	F3	F4	F5
Speaker 1	1.06E+05	3.51E+05	3.38E+05	8.49E+05	9.31E+05
Speaker 2	1.78E+05	2.74E+05	5.31E+05	1.79E+06	7.54E+05
Speaker 3	1.09E+05	1.81E+05	3.26E+05	2.88E+05	2.14E+05
Speaker 4	6.30E+05	8.69E+05	3.17E+05	3.31E+05	2.90E+05
Speaker 5	1.26E+05	1.95E+05	3.78E+05	1.81E+05	1.16E+05

Table B-21: Formant trajectory transitional energy for Bangla VCV sequence (আলো)

Speaker Name	Variation Tendency				
	F1	F2	F3	F4	F5
Speaker 1	1.22E+06	2.52E+06	3.77E+06	2.59E+06	4.33E+06
Speaker 2	1.27E+06	2.66E+06	2.11E+06	1.87E+06	2.63E+06
Speaker 3	1.36E+06	1.09E+06	2.27E+06	1.13E+06	2.39E+06
Speaker 4	1.49E+06	3.32E+06	4.94E+06	4.21E+06	5.33E+06
Speaker 5	1.25E+06	1.97E+06	3.25E+06	2.29E+06	1.56E+06

Table B-22: Formant trajectory transitional energy for Bangla VCV sequence (আমি)

Speaker Name	Variation Tendency				
	F1	F2	F3	F4	F5
Speaker 1	1.70E+06	4.21E+06	3.19E+06	2.58E+06	5.36E+06
Speaker 2	1.61E+06	4.80E+06	3.68E+06	3.74E+06	3.21E+06
Speaker 3	1.20E+06	3.46E+06	2.40E+06	1.98E+06	2.70E+06
Speaker 4	1.26E+06	3.05E+06	4.08E+06	2.03E+06	3.90E+06
Speaker 5	1.62E+06	4.09E+06	2.22E+06	2.66E+06	4.21E+06

Table B-23: Formant trajectory transitional energy for Bangla VCV sequence (ইতি)

Speaker Name	Variation Tendency				
	F1	F2	F3	F4	F5
Speaker 1	1.90E+06	5.57E+06	5.76E+06	4.19E+06	7.21E+06
Speaker 2	1.21E+06	7.11E+06	4.38E+06	3.42E+06	5.43E+06
Speaker 3	1.11E+06	3.40E+06	1.84E+06	2.57E+06	2.56E+06
Speaker 4	1.84E+06	4.47E+06	3.12E+06	2.69E+06	4.27E+06
Speaker 5	1.87E+06	6.42E+06	4.37E+06	3.12E+06	3.74E+06

Table B-24: Formant trajectory transitional energy for Bangla VCV sequence (অনু)

Speaker Name	Variation Tendency				
	F1	F2	F3	F4	F5
Speaker 1	1.11E+06	3.39E+06	4.54E+06	2.67E+06	5.82E+06
Speaker 2	1.11E+06	4.26E+06	3.98E+06	4.05E+06	3.79E+06
Speaker 3	1.69E+06	3.41E+06	6.34E+06	3.82E+06	4.31E+06
Speaker 4	1.31E+06	3.51E+06	3.78E+06	1.94E+06	4.24E+06
Speaker 5	1.41E+06	3.58E+06	4.97E+06	3.84E+06	2.69E+06

Table B-25: Formant trajectory transitional energy for Bangla VCV sequence (উচু)

Speaker 1	2.37E+06	3.87E+06	4.35E+06	2.92E+06	3.26E+06
Speaker 2	2.20E+06	6.77E+06	5.82E+06	4.30E+06	5.57E+06
Speaker 3	2.72E+06	5.60E+06	3.76E+06	2.94E+06	4.23E+06
Speaker 4	2.51E+06	4.09E+06	5.24E+06	1.87E+06	3.34E+06
Speaker 5	2.05E+06	4.74E+06	4.11E+06	3.00E+06	2.10E+06

Table B-26: Correlation of LPC filtering coefficients of Bangla vowel pairs (speaker 1)

	অ	আ	ই	ঈ	উ	ঊ	এ	ঐ	ও	ঔ
অ	1									
আ	0.5539	1								
ই	0.7196	0.5979	1							
ঈ	0.5541	0.52	0.4358	1						
উ	0.8842	0.7601	0.6424	0.7234	1					
ঊ	0.7111	0.6321	0.579	0.5728	0.5606	1				
এ	0.4935	0.4351	0.3628	0.4439	0.3139	0.3596	1			
ঐ	0.7032	0.6046	0.5583	0.5548	0.4342	0.5443	0.7697	1		
ও	0.807	0.6548	0.6287	0.6098	0.5366	0.6445	0.8256	0.588	1	
ঔ	0.7925	0.73	0.6073	0.6371	0.464	0.5294	0.8784	0.6728	0.5343	1

Table B-27: Correlation of LPC filtering coefficients of Bangla vowel pairs (speaker 2)

	অ	আ	ই	ঈ	উ	ঊ	এ	ঐ	ও	ঔ
অ	1									
আ	0.718	1								
ই	0.7193	0.6829	1							
ঈ	0.6018	0.678	0.7121	1						
উ	0.903	0.9593	0.7481	0.7832	1					
ঊ	0.8791	0.8183	0.6494	0.5791	0.6074	1				
এ	0.6801	0.7183	0.5896	0.6498	0.36	0.3557	1			
ঐ	0.6424	0.632	0.5825	0.5937	0.4572	0.4884	0.6106	1		
ও	0.7894	0.7428	0.5892	0.5376	0.4976	0.5807	0.6409	0.6562	1	
ঔ	0.4696	0.4718	0.5081	0.4323	0.2865	0.3486	0.4704	0.5106	0.3407	1

Table B-28: Correlation of LPC filtering coefficients of Bangla vowel pairs (speaker 3)

	অ	আ	ই	ঈ	উ	ঊ	এ	ঐ	ও	ঔ
অ	1									
আ	0.6454	1								
ই	0.7557	0.7566	1							
ঈ	0.4128	0.3905	0.4055	1						
উ	0.6505	0.6113	0.5235	0.5278	1					
ঊ	0.5006	0.4688	0.4389	0.6342	0.599	1				
এ	0.4557	0.5527	0.4766	0.4353	0.446	0.406	1			
ঐ	0.4325	0.5607	0.4572	0.4415	0.5488	0.4841	0.5617	1		
ও	0.5909	0.6092	0.561	0.3553	0.6068	0.5105	0.585	0.526	1	
ঔ	0.603	0.52	0.5524	0.4824	0.6347	0.5795	0.5929	0.548	0.7521	1

Table B-29: Correlation of LPC filtering coefficients of Bangla vowel pairs (speaker 4)

	অ	আ	ই	ঈ	উ	ঊ	এ	ঐ	ও	ঔ
অ	1									
আ	0.4577	1								
ই	0.5896	0.4783	1							
ঈ	0.606	0.4872	0.5364	1						
উ	0.5464	0.4036	0.489	0.5653	1					
ঊ	0.5629	0.546	0.5778	0.605	0.6099	1				
এ	0.6955	0.5995	0.5936	0.6599	0.5514	0.6862	1			
ঐ	0.4173	0.4428	0.3961	0.4617	0.3544	0.383	0.391	1		
ও	0.5574	0.3699	0.4607	0.5236	0.5089	0.6366	0.4972	0.428	1	
ঔ	0.483	0.3079	0.5161	0.4944	0.5904	0.5719	0.4621	0.4215	0.6645	1

Table B-30: Correlation of LPC filtering coefficients of Bangla vowel pairs (speaker 5)

	অ	আ	ই	ঈ	উ	ঊ	এ	ঐ	ও	ঔ
অ	1									
আ	0.4559	1								
ই	0.7745	0.9548	1							
ঈ	0.4458	0.6401	0.3327	1						
উ	0.805	0.745	0.8407	0.9076	1					
ঊ	0.5864	0.9915	0.475	0.7159	0.1973	1				
এ	0.7807	0.8064	0.5982	0.647	0.2986	0.4219	1			
ঐ	0.7334	0.7539	0.4837	0.4622	0.2731	0.3692	0.5159	1		
ও	0.7286	0.945	0.6282	0.8183	0.1897	0.4927	0.4995	0.4977	1	
ঔ	0.8967	0.8684	0.659	0.6542	0.2382	0.4865	0.6221	0.6083	0.5464	1

Table B-31: Correlation of LPC filtering coefficients of Bangla vowel and VCV sequence (speaker 1)

	আলো	আমি	ইতি	অনু	উচু
অ	0.2453	0.2044	0.1238	0.1476	0.2334
আ	0.2463	0.2025	0.1188	0.1610	0.2512
ই	0.1750	0.1684	0.0964	0.1253	0.1822
ঈ	0.1927	0.1878	0.1153	0.1193	0.2217
উ	0.1703	0.1397	0.0780	0.0864	0.1703
ঊ	0.1722	0.1512	0.0950	0.1113	0.1749
এ	0.2532	0.2728	0.1670	0.1841	0.2606
ঐ	0.2136	0.1822	0.1022	0.1614	0.1969
ও	0.2063	0.1752	0.0903	0.1258	0.1682
ঔ	0.1159	0.1098	0.0632	0.0996	0.1070

Table B-32: Correlation of LPC filtering coefficients of Bangla vowel and VCV sequence (speaker 2)

	আলো	আমি	ইতি	অনু	উচু
অ	0.1890	0.1762	0.2148	0.1539	0.2166
আ	0.1718	0.1556	0.2133	0.1328	0.2031
ই	0.2302	0.2475	0.2075	0.1740	0.2338
ঈ	0.1661	0.1676	0.1999	0.1344	0.2153
উ	0.1802	0.1787	0.2312	0.1498	0.2379
ঊ	0.1970	0.1768	0.2509	0.1389	0.2567
এ	0.1984	0.1855	0.2266	0.1415	0.2599
ঐ	0.1727	0.1198	0.1673	0.1394	0.2043
ও	0.1819	0.1410	0.2334	0.1442	0.2563
ঔ	0.2009	0.1844	0.2233	0.1363	0.2501

Table B-33: Correlation of LPC filtering coefficients of Bangla vowel and VCV sequence (speaker 3)

	আলো	আমি	ইতি	অনু	উচু
অ	0.1799	0.193	0.2067	0.1039	0.1398
আ	0.1361	0.1273	0.2389	0.0728	0.1099
ই	0.1971	0.1902	0.2076	0.1188	0.1732
ঈ	0.1733	0.1715	0.2426	0.1017	0.1457
উ	0.2427	0.2173	0.2345	0.1343	0.1739
ঊ	0.2202	0.2148	0.2086	0.12	0.161
এ	0.0891	0.0994	0.2434	0.0511	0.0862
ঐ	0.1455	0.1519	0.2597	0.0848	0.1201
ও	0.2346	0.2441	0.2372	0.1387	0.1654
ঔ	0.1804	0.1685	0.2412	0.1037	0.1462

Table B-34: Correlation of LPC filtering coefficients of Bangla vowel and VCV sequence (speaker 4)

	আলো	আমি	ইতি	অনু	উচু
অ	0.1531	0.1242	0.0862	0.1156	0.3496
আ	0.107	0.1022	0.0681	0.1025	0.2741
ই	0.1608	0.1089	0.0828	0.1082	0.343
ঈ	0.1538	0.1118	0.0877	0.1195	0.2907
উ	0.1427	0.1084	0.0713	0.1208	0.3072
ঊ	0.1108	0.0856	0.062	0.0918	0.2667
এ	0.0608	0.0433	0.0337	0.0446	0.1324
ঐ	0.101	0.0821	0.0567	0.0815	0.2541
ও	0.1075	0.0954	0.0659	0.0856	0.246
ঔ	0.0965	0.0803	0.0561	0.0799	0.2158

Table B-35: Correlation of LPC filtering coefficients of Bangla vowel and VCV sequence (speaker 5)

	আলো	আমি	ইতি	অনু	উচু
অ	0.1387	0.1166	0.0975	0.1265	0.1366
আ	0.1357	0.1092	0.0963	0.1191	0.1239
ই	0.1376	0.1169	0.1054	0.1184	0.1354
ঈ	0.1377	0.1226	0.1071	0.1233	0.1142
উ	0.1181	0.0985	0.0856	0.1091	0.1223
ঊ	0.1649	0.1384	0.1208	0.1459	0.1348
এ	0.1379	0.1172	0.1063	0.1152	0.1135
ঐ	0.1403	0.1228	0.1106	0.1244	0.1382
ও	0.1221	0.1063	0.093	0.103	0.1182
ঔ	0.1196	0.1082	0.1067	0.1071	0.1304

Table C-1: Normalized eigen-value of Bangla vowel perceptual space for fifteen Bangla voice speakers

Speakers Number	Eigen Number								
	1	2	3	4	5	6	7	8	9
S1	0.7879	0.1197	0.0435	0.0301	0.0102	0.0055	0.0023	0.0005	0.0002
S2	0.7634	0.1003	0.0673	0.0392	0.0166	0.0091	0.0030	0.0010	0.0003
S3	0.6411	0.2494	0.0551	0.0244	0.0225	0.0037	0.0023	0.0015	0.0000
S4	0.5135	0.3098	0.0806	0.0491	0.0305	0.0127	0.0031	0.0006	0.0001
S5	0.5346	0.2502	0.1297	0.0544	0.0155	0.0099	0.0030	0.0018	0.0010
S6	0.7833	0.1095	0.0745	0.0175	0.0073	0.0058	0.0020	0.0002	0.0000
S7	0.7021	0.1536	0.0953	0.0314	0.0101	0.0051	0.0022	0.0001	0.0000
S8	0.5978	0.3133	0.0613	0.0213	0.0034	0.0021	0.0008	0.0001	0.0000
S9	0.8763	0.0627	0.0263	0.0258	0.0061	0.0027	0.0001	0.0001	0.0000
S10	0.7970	0.1144	0.0561	0.0179	0.0130	0.0008	0.0007	0.0000	0.0000
S11	0.6837	0.1898	0.0599	0.0405	0.0133	0.0056	0.0036	0.0025	0.0011
S12	0.8698	0.0601	0.0419	0.0133	0.0100	0.0029	0.0012	0.0005	0.0001
S13	0.6755	0.1957	0.0711	0.0297	0.0154	0.0074	0.0042	0.0008	0.0001
S14	0.8216	0.1189	0.0301	0.0112	0.0102	0.0060	0.0019	0.0003	0.0000
S15	0.5802	0.3702	0.0294	0.0104	0.0068	0.0016	0.0013	0.0001	0.0001

Table C-2: Normalized contributions of perceptual space members in first principal component

Speaker Numbers	Perception Space Members								
	F1	F2	F3	F4	F5	F51	F43	F53	F54
S1	-0.088	-0.655	-0.626	-0.241	-0.256	-0.060	-0.093	-0.126	-0.142
S2	-0.073	-0.667	-0.498	-0.268	-0.427	0.178	-0.049	-0.065	-0.093
S3	0.277	-0.361	-0.247	0.102	0.456	-0.245	-0.245	-0.607	-0.159
S4	-0.037	0.767	0.485	-0.085	-0.102	0.191	0.291	0.191	0.004
S5	-0.164	-0.749	-0.478	-0.196	0.004	-0.014	-0.003	-0.331	-0.187
S6	0.027	-0.911	-0.316	-0.119	0.041	0.027	0.215	0.080	-0.031
S7	0.027	0.887	0.296	0.049	0.328	-0.052	-0.110	0.008	0.002
S8	0.082	-0.799	-0.443	-0.092	0.342	-0.075	0.028	-0.114	-0.119
S9	0.106	-0.912	-0.262	-0.068	0.287	0.002	-0.024	-0.002	-0.013
S10	0.075	-0.970	-0.202	0.027	0.110	0.001	-0.020	-0.012	0.004
S11	0.006	-0.727	-0.261	-0.106	0.206	-0.054	0.021	-0.469	-0.356
S12	-0.151	-0.762	-0.461	-0.257	-0.303	0.148	0.055	0.029	0.023
S13	-0.043	-0.729	-0.601	-0.241	0.000	-0.023	0.030	-0.075	-0.199
S14	-0.149	-0.651	-0.570	-0.217	-0.310	0.105	0.223	0.156	0.039
S15	-0.505	-0.013	-0.596	-0.157	0.295	0.299	0.216	0.296	0.233

Table C-3: Normalized contributions of perceptual space members in second principal component

Speaker Numbers	Perception Space Members								
	F1	F2	F3	F4	F5	F51	F43	F53	F54
S1	-0.220	0.233	-0.149	-0.316	-0.434	0.453	0.336	0.477	0.203
S2	-0.237	0.608	-0.280	-0.354	-0.234	0.491	0.048	0.135	0.236
S3	-0.237	-0.638	-0.493	-0.267	-0.450	0.073	0.115	0.021	-0.026
S4	-0.448	-0.252	-0.097	-0.476	-0.573	-0.028	0.275	0.260	0.164
S5	0.380	-0.120	0.159	0.330	0.701	0.184	-0.248	-0.347	0.016
S6	0.257	-0.328	0.484	0.161	0.285	-0.239	-0.517	-0.390	-0.091
S7	-0.156	-0.245	0.214	-0.054	0.415	-0.516	-0.028	-0.469	-0.456
S8	0.210	-0.044	-0.215	0.013	-0.734	-0.353	-0.071	-0.315	-0.375
S9	0.718	-0.152	0.136	-0.116	-0.650	0.046	0.042	0.037	-0.037
S10	0.343	0.039	-0.504	-0.116	-0.781	-0.011	-0.018	0.010	-0.053
S11	-0.440	-0.450	-0.222	-0.055	-0.226	0.163	0.349	0.540	0.247
S12	0.020	0.004	-0.311	-0.183	0.342	-0.332	-0.042	-0.690	-0.406
S13	0.106	0.429	-0.304	-0.113	0.304	-0.413	-0.292	-0.472	-0.359
S14	0.360	-0.670	0.231	0.157	0.568	-0.062	-0.112	-0.057	0.002
S15	0.023	0.801	0.380	0.078	0.151	0.170	0.166	0.337	0.125

Table C-4: Normalized contributions of perceptual space members in third principal component

Speaker Numbers	Perception Space Members								
	F1	F2	F3	F4	F5	F51	F43	F53	F54
S1	0.554	0.183	-0.208	-0.126	0.001	0.440	-0.601	-0.065	0.206
S2	0.128	-0.156	0.053	0.192	-0.277	-0.159	0.410	0.640	0.489
S3	0.330	0.131	-0.297	-0.438	0.190	0.334	-0.217	0.050	0.630
S4	0.411	-0.424	0.471	0.405	-0.258	0.051	0.260	0.177	0.313
S5	0.169	-0.230	-0.198	0.068	-0.156	0.680	-0.108	0.383	0.479
S6	-0.327	-0.113	0.479	0.151	0.440	0.432	0.136	0.332	0.345
S7	-0.561	0.256	-0.544	-0.474	0.014	0.096	0.241	-0.117	-0.135
S8	-0.431	-0.154	-0.216	-0.489	-0.461	0.425	0.201	0.179	0.197
S9	0.204	0.219	-0.107	-0.893	0.310	-0.035	0.007	-0.025	-0.058
S10	-0.403	-0.223	0.612	0.048	-0.591	-0.183	0.015	-0.170	0.023
S11	0.228	-0.396	0.595	0.395	0.314	-0.278	0.095	0.176	0.258
S12	-0.681	0.425	-0.353	-0.278	-0.002	-0.222	0.172	0.257	0.092
S13	0.581	-0.163	-0.053	0.390	0.569	0.353	-0.159	0.010	0.090
S14	-0.450	-0.275	0.648	0.011	-0.274	-0.300	0.218	0.205	-0.218
S15	-0.314	0.481	-0.301	-0.319	-0.432	-0.389	-0.044	-0.177	-0.326

Table C-5: Normalized contributions of perceptual space members in fourth principal component

Speaker Numbers	Perception Space Members								
	F1	F2	F3	F4	F5	F51	F43	F53	F54
S1	-0.263	0.679	-0.406	-0.096	-0.117	-0.223	-0.128	-0.295	-0.360
S2	-0.046	-0.350	0.257	0.011	0.492	0.716	0.187	0.117	0.069
S3	-0.247	0.484	-0.343	0.182	-0.442	-0.076	-0.482	-0.345	0.029
S4	-0.450	-0.173	0.643	-0.011	-0.026	-0.035	-0.574	-0.138	-0.060
S5	0.135	0.533	-0.539	-0.279	-0.084	0.087	-0.402	-0.381	0.088
S6	0.231	0.196	-0.467	0.260	0.732	0.047	0.189	-0.219	-0.037
S7	-0.094	0.233	0.206	0.060	-0.842	-0.278	0.037	-0.212	-0.241
S8	0.612	-0.203	-0.039	0.252	-0.270	0.337	-0.269	0.276	0.429
S9	0.070	0.209	-0.928	0.100	-0.179	0.073	-0.013	0.148	0.146
S10	0.744	-0.024	0.455	0.330	-0.028	0.174	-0.239	-0.089	0.184
S11	-0.091	-0.213	0.212	-0.231	-0.646	-0.400	-0.414	-0.187	0.255
S12	0.094	0.051	-0.302	0.375	-0.364	-0.510	-0.556	0.002	0.234
S13	0.187	0.444	-0.592	-0.293	-0.035	0.143	0.098	0.457	0.304
S14	-0.066	-0.171	-0.030	-0.035	-0.318	-0.101	-0.510	-0.684	-0.353
S15	-0.527	0.024	0.237	0.176	-0.164	0.034	-0.574	-0.173	0.497

Table C-6: Normalized contributions of perceptual space members in fifth principal component

Speaker Numbers	Perception Space Members								
	F1	F2	F3	F4	F5	F51	F43	F53	F54
S1	0.531	0.104	-0.038	-0.155	-0.303	-0.617	0.152	0.317	0.004
S2	0.226	-0.048	0.050	-0.121	0.076	0.051	-0.825	0.482	0.088
S3	0.570	-0.153	0.104	0.319	-0.338	0.551	-0.223	0.045	-0.266
S4	0.483	-0.161	0.124	-0.470	-0.214	0.412	-0.115	-0.124	-0.510
S5	-0.028	-0.149	0.477	0.173	-0.593	0.125	-0.364	-0.468	0.031
S6	0.716	0.006	0.349	0.166	-0.190	0.113	0.531	0.069	-0.052
S7	0.247	-0.058	0.237	-0.816	-0.092	0.071	-0.446	-0.009	0.027
S8	0.232	0.530	-0.717	-0.285	0.214	-0.014	-0.125	-0.009	0.094
S9	0.531	0.165	0.131	0.293	0.541	0.210	0.071	0.337	0.366
S10	-0.122	0.026	-0.164	0.807	-0.084	-0.052	0.417	0.300	0.179
S11	0.528	-0.040	0.181	-0.534	-0.194	-0.121	0.356	0.306	-0.360
S12	0.064	0.203	-0.367	-0.058	0.300	0.563	-0.307	-0.224	0.515
S13	-0.108	-0.161	-0.102	0.286	-0.093	-0.478	-0.616	0.226	0.450
S14	0.725	0.010	0.016	0.187	-0.513	-0.361	0.072	0.188	-0.062
S15	-0.385	-0.351	0.440	-0.130	-0.259	-0.263	0.289	0.538	-0.102