

FREQUENCY OF OCCURRENCE OF PHONEMES IN MALAYALAM

Project funded by A.I.I.S.H Research Fund (ARF)

(2012-2013)



Sanction Number: SH/SLS/ARF/4.60/2012-2013

Total Grants: ₹ 3,31,000

Total Duration of the Project: 10.9.2012 – 9.9.2013

Principal Investigator

Dr. N. Sreedevi
Reader in Speech Sciences
Department of Speech Language Sciences

Research Officer

Ms. Irfana. M
Department of Speech Language Sciences

**All India Institute of Speech and Hearing
Manasagangothri, Mysore – 570 006**

ACKNOWLEDGEMENTS

Our sincere gratitude to our Director, Prof. S R Savithri, All India Institute of Speech and Hearing, Mysore, for funding and providing the infrastructure to carry out the project.

We extend our gratitude to Prof. Y V Geetha, Head, Department of Speech Language Sciences for providing us the departmental facilities to carry out the project work. Our thanks are due to the Head and staff of the AIISH library for their help in completing the project. We also extend our gratefulness to all the subjects who participated in the study.

Dr. N. Sreedevi
Principal Investigator

Ms. Irfana.M
Research Officer

TABLE OF CONTENTS

Sl. No.	Title	Page No.
1.	Introduction	1 - 7
2.	Review of Literature	8- 27
3.	Method	28 - 32
4.	Results	33 – 66
5.	Discussion	67– 69
6.	Summary & Conclusions	70-72
7.	References	73– 76
8.	Appendices	77– 83

LIST OF TABLES

Table No.	Title	Page No.
2.1	Summary of the frequency of occurrence of phonemes in various languages (Yegerlehner & Voegelin, 1957)	10
2.2	Frequency of occurrence of phonemes in Basque (Tambovtsev, 2007)	12
2.3	Percentage of occurrence of phonemes in English (Denes, 1959)	14
2.4	Mean percentage of occurrence of vowels in five different languages (Thomas, 2005)	18
2.5	Mean percentage of occurrence of consonants in five different languages (Thomas, 2005)	19
2.6	Number of occurrences and mean percentage of occurrence of vowels and consonants in spoken and written modes in Castilian Spanish (Sandoval et al, 2008)	21
3.1	Number of participants in each recording session	29
3.2	Topic of conversation in each recording session for the three dialects	30
4.1	Mean percentage and standard deviation of occurrence of vowels and consonants in spoken Calicut dialect of Malayalam in alphabetical order	36
4.2	Mean percentage and standard deviation of occurrence of vowels and consonants in spoken Eranakulam dialect of Malayalam in alphabetical order	43
4.3	Mean percentage and standard deviation of occurrence of vowels and consonants in spoken Thiruvananthapuram dialect of Malayalam	50
4.4	Overall mean percentage and standard deviation of occurrence of vowels and consonants in spoken Malayalam	62

LIST OF FIGURES

Figure No	Title	Page No.
2.1	Occurrence of vowels in Telugu (Kalyani & Sunitha, 2009)	25
2.2	Occurrence of consonants in Telugu (Kalyani & Sunitha, 2009)	25
4.1	Number of phonemes obtained in each recording session	35
4.2	Mean percentage of occurrence of vowels and consonants in Calicut dialect	35
4.3	Mean percentage of occurrence of long and short vowels in Calicut dialect	37
4.4	Mean percentage of occurrence of consonants based on places of articulation in Calicut dialect	38
4.5	Mean percentage of occurrence of consonants based on manner of articulation in Calicut dialect	38
4.6	Mean percentage of occurrence of stops in Calicut dialect	39
4.7	Mean percentage of frequency of occurrence of ten most frequently occurring phonemes in Calicut dialect	40
4.8	Number of phonemes obtained in each recording session in Eranakulam dialect	41
4.9	Mean percentage of occurrence of vowels and consonants in Eranakulam dialect	42
4.10	Mean percentage of occurrence of long and short vowels in Eranakulam dialect	44
4.11	Mean percentage of occurrence of consonants based on places of articulation in Eranakulam dialect	45
4.12	Mean percentage of occurrence of consonants based on manner of articulation in Eranakulam dialect	45

4.13	Mean percentage of occurrence of stops in Eranakulam dialect	46
4.14	Mean percentage of frequency of occurrence of ten most frequently occurring phonemes in Eranakulam dialect	48
4.15	Number of phonemes obtained in each recording session in Thiruvanathapuram dialect	48
4.16	Mean percentage of occurrence of vowels and consonants in Thiruvanathapuram dialect	49
4.17	Mean percentage of occurrence of long and short vowels in Thiruvanathapuram dialect	51
4.18	Mean percentage of occurrence of consonants based on places of articulation in Thiruvanathapuram dialect	52
4.19	Mean percentage of occurrence of consonants based on manner of articulation in Thiruvanathapuram dialect	53
4.20	Mean percentage of occurrence of stops in Thiruvanathapuram dialect	54
4.21	Mean percentage of frequency of occurrence of ten most frequently occurring phonemes in Thiruvanathapuram dialect	55
4.22	Comparison of mean percentage of occurrence of consonants and vowels across dialects	56
4.23	Comparison of mean percentage of occurrence of long and short vowels across dialects	57
4.24	Comparison of mean percentage of occurrence of consonants based on places of articulation across dialects	58
4.25	Comparison of mean percentage of occurrence of consonants based on manner of articulation	59
4.26	Comparison of mean percentage of occurrence of stops across dialects	60
4.27	Mean percentage of occurrence of vowels and consonants in Malayalam	61

4.28	Mean percentage of occurrence of short and long vowels in Malayalam	63
4.29	Mean percentage of occurrence of consonants based on places of articulation in Malayalam	63
4.30	Mean percentage of occurrence of consonants based on manner of articulation in Malayalam	64
4.31	Mean percentage of occurrence of stops in the total corpus obtained in Malayalam	65
4.32	Mean percentages of frequency of occurrence of the ten most frequent phonemes in Malayalam	66

CHAPTER I

INTRODUCTION

Communication is central to our lives. We communicate in a number of ways; from text messaging to facial expressions. It is a complex, two way and intentional process of passing the message from one end to the other using a channel. It refers to any act in which information is given to or received from one person to another person concerning that person's needs, desires, perceptions, knowledge, or affective states. For this, we need a dynamic system of conventional symbol that can be used in various modes. Language is a purely human and non-instinctive mode of communicating ideas, emotions and desires by means of voluntarily produced symbols (Sapir, 1921). According to Chomsky (1957), all natural languages in either spoken or written form, are languages where each natural language has a finite number of sounds in it and although there may be infinitely many distinct sentences in the languages, each sentence can be represented as a finite sequence of these sounds.

The study of the sound system of language which includes the rules that govern its spoken form is called phonology (Perkin & Riley, 2005). It is the blue print of each sound type, which serves as the constant basis of all the variations in different physical articulations of that sound type in different contexts. When we think of the /t/ sound in the words *tar*, *star*, *writer* and *eighth* as being the same, in the phonology of English, they would be represented in the same way. In the actual conversational speech we are using it in different ways. Considering this difference, we see that phonology is concerned with the abstract set of sounds in a language that allows us to distinguish meaning in the actual physical sounds we say and hear. Phonology

analyzes which sound units are within a language. The sound system of English contains different vowels and consonants than that of other languages in the world.

Phoneme is the smallest linguistic unit that is able, when combined with other such units, to establish word meanings and distinguish between them. It is an abstraction that is based on the many variations that occur for a particular sound as it is heard in different contexts of conversational speech. Allophones are variations in phoneme realizations that do not change the meaning of a word when they are produced in differing contexts.

Each language has its own phonological system and phoneme inventory. A standard language is actually an idealized variety, but exists for most people as the version that is accepted as the official language of their community or country. It is clearly associated with education and broadcasting in public contexts and is more easily described in terms of the written language and varies from spoken language. When we speak we always follow specific pronunciation pattern based on the region or the society where we are from. These language variations are quite normal in a society composed of a multitude of social groups that have become quite diversified.

Dialect is a neutral label that refers to any variety of a language that is shared by a group of speakers. The dimension of language variations may vary from vocabulary, grammar and pronunciation. The factors that may correlate with a particular dialect usage may be as simple as geographical locality or as complex as a notion of cultural identity. Those dialects corresponding to various geographical locations considered as regional dialects and traditionally dialectologist study languages dividing the region into northern, midland and southern. There are

studies reporting that there are certain phonological changes associated with each dialect of a language (Christian, Wolfram & Nudé, 1988).

Malayalam is a Dravidian language primarily spoken in the southwest of India. According to Summer (2009, as cited in Ishtiaq, 2012), it is the official language of Kerala state and Lakshadweep union territory. Within India alone there were over 35 million speakers of Malayalam, not including the other nearly 500,000 speakers outside India. Malayalam phoneme system has 11 monothongs and 2 diphthongs and 52 consonant phonemes, encompassing 9 places of articulation which are bilabial, labiodentals, dental, alveolar, alveolo-palatal, retroflex, palatal, velar and glottal and 8 manners of articulation which include plosive, nasal, trill, tap/flap, fricative, affricate, central approximant and lateral approximant (Jian, 2010). Based on reports of Asher and Kumari (1997), places of articulations in Malayalam are labials, dentals, alveolars, retroflex, palatals, velars, glottal and manner of articulation include stops, nasals, fricatives, taps/trills, lateral and glides.

According to Ladefoged and Maddieson (1996) a difference between apical dental and interdental gestures with different manners of articulation occurs in Malayalam. This may be related to the fact that Malayalam has a contrast between dentals and alveolars for both stops and nasals. Acoustically, both dental and interdental stops are quite distinct from the contrasting alveolar stops in their burst as well as their formant transitions. The nasals have virtually no bursts, and are distinguished almost entirely by their formant transition. Those speakers of Malayalam who have interdental nasals might thus increase the difference by producing more distinct formant transitions as results of the interdental articulation. Dental stops are usually laminal rather than apical, with contact on both teeth and the front part of the alveolar ridge;

whereas in Malayalam, the alveolar stops are often apical with contact usually on the centre of the alveolar ridge.

As per the geographical locations, in Malayalam there are mainly three regional dialects such as Calicut, Eranakulam and Thiruvananthapuram dialects. Calicut is a northern district of Kerala where more of Muslim community is accommodated. The dialect of Malayalam spoken by mappila Muslims shows deep influence of Arabic language. Eranakulam is a middle district of Kerala where more of middle class and upper middle class community is accommodated and has more influence of English. Also Eranakulam dialect varies with respect to rate of speech and intonation pattern from standard Malayalam. Thiruvananthapuram is a southern district of Kerala where the economical status is similar to Eranakulam. There are some peculiar vocabularies present in this dialect and the usage of these words varies from area to area. And these dialects are partially different from mainstream Malayalam.

The statistics of phoneme occurrence varies from language to language and dialects to dialect. This phoneme frequency is useful in automatic generation of speech for reading machine for the blind. For the same reason the knowledge about the frequency of phoneme provides a database for phoneme identification process of an automatic speech recognition system. The frequency of occurrence of phonemes plays a crucial role in the development of linguistic theories in a number of areas including the grammatical relations, semantic structure etc. Audiologists use several speech materials in the form of word lists, involving phonetically balanced phonemes in a language for hearing assessment and intervention. As these test materials are language specific, they need to be based on frequently occurring phonemes in that particular language (Egan, 1948; Campbell, 1965). For speech language pathologists, these

studies would provide a database for developing speech materials for assessment and selecting treatment targets for various communication disorders, knowledge about the most frequent phonemes aids in targeting those phonemes in speech therapy for the hearing impaired which leads to better intelligibility of their speech.

India with a multicultural and multi-linguistic background, there are dearth of studies on the frequency of phonemes in Indian languages. These studies would provide a database for developing speech materials for assessment and selecting treatment targets for various communication disorders, knowledge about the most frequently occurring phonemes can help in targeting those phonemes in therapy for hearing impaired. The phonetically balanced word lists that audiologists use for assessing auditory processing disorders like staggered spondaic words (SSW), for checking speech identification scores (SIS), speech in noise test (SPIN) and speech recognition scores (SRT) in routine audiological evaluations are based on such phoneme frequency information and they are highly language specific. Such information can also help in the development of different aids and devices like text to speech converters for individuals with communication disorders. Gelfand (2009) has listed the following clinical functions of speech recognition score testing: 1) to describe the extent of hearing impairment in terms of how it affects speech understanding 2) to differentially diagnose auditory disorders by determining the anatomical site of lesion 3) for determining the needs for amplification and other forms of aural rehabilitation devices like cochlear implants, bone anchored hearing aids etc. 4) for making comparisons between various hearing aids, amplification approaches and other forms of aural rehabilitation devices 5) for verifying the benefits of hearing aid use and other forms of aural rehabilitation devices 6) for monitoring patient performance over time for either diagnostic or rehabilitative purposes.

The phoneme statistics are also applicable to the fields of linguistics and speech, to teach the language as a foreign language. India is a country with diverse languages. And these languages have different phoneme system and there are different dialects under each language system. A database of different dialects with its phoneme frequency will help us to know a language better. The earlier study on phoneme frequency in Malayalam by Ghatage (1994) was from various written materials in the language. Also his findings may not be apt at present as there are a lot of new words, modified and borrowed words used in day to day conversation. Hence the present study is planned to obtain the frequency of occurrence of phonemes in conversational speech samples in Malayalam.

Need for the study

There is a wide application using the statistics of phoneme occurrence in a language. Machines that speak and recognize speech has been a dream for communication engineers. The realization of these machines, be it a phonetic type writer or a computer which responds to spoken words, depends upon the availability of a great deal of information about the acoustical characteristics of sounds in each language spoken by its community. Hence for the same reason the knowledge about the frequency of phoneme provides a database for phoneme identification process of an automatic speech recognition system. These statistics are also applicable to the fields of linguistics and speech, to teach the language as a foreign language and to speech pathology. Furthermore, the possibility of translation from one language into another by computer, points out the need for the study of languages. These data may also be of use in automatic generation of speech for reading machine for the blind.

Aim

To obtain the frequency of occurrence of phonemes in conversational speech samples of three major dialects of Malayalam: Calicut, Eranakulam and Thiruvananthapuram dialects.

Objectives

1. To obtain a database with the occurrence of various phonemes in the three major dialects of Malayalam
2. To calculate the frequency of occurrence of each phoneme from the database obtained
3. To obtain the order of frequency of phonemes in the three dialects separately
4. To compare of phoneme frequency across dialects
5. To obtain the mean frequency of occurrence of phonemes in Malayalam language
6. To compare the frequency of occurrence of phonemes in Malayalam with previous studies in other languages

Implications of the study

This study has wide application in the development of phonetically balanced word lists which are routinely used in audiological tests. Presently used PB word lists are based on the frequency of occurrence of phonemes established almost four decades earlier. It is well known that the frequency of occurrence of speech sounds in a language can vary over a long period of time. Hence from the data base generated in the present study, the existing test material can be updated. This information also aids in hearing aid fitting as well as in assessing the severity of hearing disability and guides in setting speech therapy goals for communication disordered population as working on the most frequently occurring phonemes in the language improves speech intelligibility considerably.

CHAPTER II

REVIEW OF LITERATURE

In a language, a phoneme is the functional unit of speech that is used to distinguish meaning of a word. Different languages have different phonological systems and the same sound or the same groups of sounds do not necessarily have the same function in one language as in another. Although the dialects are the variation of a single language, it differs in terms of vocabulary, grammar, pragmatics and pronunciation (Siegel, 2010). There are about 600 consonants used in different languages around the world and some occur more frequently than others. Like consonants, an account of the number of vowels across the world may be difficult as one vowel glide into another easily. The well known forms of English have 14 vowels (Californian English) to 20 vowels (BBC English). The frequency can vary from language to language and dialect to dialect. High occurrence of a phoneme in a language means that the speakers of this language articulate this particular sound more frequently than other sounds, and are thus more familiar with the articulation of this sound.

The frequency of occurrence of phonemes is in research since 1930s. As the phonological structure varies with the language use and the dialects, several studies were carried out in different languages. Studies on frequency of occurrence of phonemes majorly concentrated to develop speech materials for audiologists, speech language pathologists for various clinical purposes (Palai & O'Hanlon, 2004) and a data base for speech identification devices. But few other studies aimed to develop a phonemic and syllabic frequency inventory for those particular languages (Sandoval, Toledano, Torre, Garrote & Guirao, 2008; Tamaoka & Makioka, 2004).

All these studies on frequency of occurrence of phonemes will be discussed under several headings as follows

- a) Phoneme frequency from written data in Non-Indian languages
- b) Phoneme frequency from spoken data in Non-Indian languages.
- c) Phoneme frequency: written vs spoken data.
- d) Frequency of occurrence of phonemes in Indian languages.

a) Phoneme frequency from written data in Non-Indian Languages

Initial studies of frequency of occurrence of phonemes were mainly focused on written text. Authors considered most popular medium like news paper, script of play, novels etc where they analyzed the inventory of phoneme, their distributions and combinations. Dewey (1923) is one of the earliest researchers in this field where he considered 100,000 words from written material such as newspaper editorials, news articles, fiction, letters, scientific articles and magazines in English. Results showed that consonant /t/ occurred 7.85 % and vowel /ə/ occurred 7.76 % of total phoneme count. /t/ and /ə/ were the most frequent consonant and vowel respectively. /z/ was the less frequent in the entire data.

After 1950's there were many studies in this line. Yegerlehner and Voegelin (1957) compared and gave a summary of frequencies and inventories of phonemes from nine different languages. The research was carried out by different authors on several tribal languages. The nine different languages were Maori (spoken in Malayo Polynesian), Hidasta (North Plain, Siouan), Winnebago (Eastern Woodlands, Siouan), Shawnee (Eastern Woodlands, Algonquian), Choctaw (South East, Muskogean), Havasupai (South West, Yuman), Navaho (South West,

Athapaskan), Chontal (Mexico, Mayan), and Tarascan (Mexico, Unclassified). The frequency of occurrence of phonemes in these languages is arranged in descending order as depicted in Table 2.1.

Table 2.1

Summary of the frequency of occurrence of phonemes in various languages (Yegerlehner & Voegelin, 1957)

Languages	Frequency of occurrence of phonemes-in decreasing order
Maori	/a/, /i/, /e/, /t/, /k/, /o/
Hidasta	/a/, /n/, /i/, /h/, /k/
Winnebago	/e/, /a/, /i/, /g/, /n/
Shawnee	/i/, /a/, /e/, /w/, /k/, /l/, /t/
Choctaw	/a/, /i/, /t/, /m/, /h/, /k/, /n/, /l/
Havasupai	/k/, /a/, /i/, /m/, /y/, /h/
Navaho	/a/, /i/, /d/, /n/
Chontal	/u/, /n/, /a/, /h/, /t/
Tarascan	/a/, /i/, /n/, /k/, /s/, /u/, /e/

Yadav (1976) attempted to study the aspirated consonants of Maithili, the second most important language of Nepal. Words from dictionary (Jha, 1952) were considered for the study which includes aspirated consonants occurring in all positions. Results showed that /b, k, s, p, m/ were the five frequently occurring phonemes and /l^h/ was the least occurring phoneme in their written corpus. Similarly the five most frequent aspirated phoneme of Maithili are /b^h, k^h, p^h, c^h, g^h/ in that order.

Mitton (1992) studied the frequency of occurrence of phonemes in English. He considered written data from Advanced Learner's Dictionary. The total number of words

included 70,646 which were selected based on random sampling method and consisted of 4,92,745 phoneme occurrences. Results showed that /i/, /ə/, /æ/, /e/ were the most frequently occurring vowels where only /i/ occurred more than 10 % and /t/, /s/, /n/, /l/, /r/, /k/ and /d/ were the most frequently occurring consonants. Phonemes /w/, /i/, /a/, /u/, /ɔ/, /h/, /ʃ/, /dʒ/, /ʒ/ occurred for less than 1% of the total data.

Zurinkas (2002) considered a large sample data, consisting of about 1,63,24,176 words in English. The samples were taken from newspaper (London Times). The analysis was carried out for calculating the frequency of words and phonemes including vowels and consonants. The most frequently occurring phonemes were found to be /n/, /t/, /i/ which occurred for 7% of the total data. Following these phonemes vowel /u/ occurred for 6.7%, and consonants /s/, /d/, /e/ occurred for 4% of the total data. Among vowels /i/ and /u/ were found to be the most occurring followed by /e/. Among consonants /n/ and /t/ occurred most frequently followed by /s/ and /d/.

Tamaoka and Makioka (2004) studied frequency counts for units of Japanese phonemes, morae, and syllables, using the word printed frequency index constructed from 1985 to 1998 editions of the newspaper. Results showed that among the five vowels, /a/ (23.42%), /i/ (21.54%), /u/ (23.47%), and /o/ (20.63%) showed similar frequency rates, whereas /e/ (10.94%) was less frequent. Among the 12 consonants, /k/ (17.24%), /t/ (15.53%), and /r/ (13.11%) were used often, whereas /p/ (0.60%) and /b/ (2.43%) appeared far less frequent. Among the five long vowels, /ar/ (34.4%) was used most frequently, whereas /ur/ (12.11%) was not used so often. The special sound /n/ appeared very frequently in Japanese. The /sj/ with vowels appeared most frequently. Among the CSV moraic units, /sjo/ appeared most frequently, whereas /pja/ was the least, being counted only 95 times. Among various syllabic combinations including /n/, the

combination of /k/+V+/n/ showed the highest frequency counts. The special sound /Q/ was also found to appear very frequently.

Tambovtsev (2007) discussed the distances between Basque and other world languages based on the frequency of occurrence of phonemes in text material. In Basque, consonants (51.37%) occurred more predominantly than vowels (48.63%). Voiced consonants accounted for 12.36% of total phonemes. And front sounds showed predominant occurrence compared to back sounds. Palatal phonemes occurred less frequently in the entire corpus. Percentages of occurrence of each class of phonemes with respect to total corpus and to consonants alone are shown in Table 2.2.

Table 2.2

Frequency of occurrence of phonemes in Basque (Tambovtsev, 2007)

Phonemes	Frequency % to all phonemes	Frequency % to consonants
Labial	5.72	11.13
Front	35.89	69.87
Palatal	0.47	0.92
Velar	9.29	18.08
Sonorant	20.23	39.38
Occlusive	21.92	42.67
Fricative	9.22	17.95

Renwick (2011) studied the relative frequencies of phonemes in Romanian. The relative occurrences of vowels are given more importance in the study. Vowels with different combinations of consonants forming segments were analyzed. The data consisted of about 88,580 words consisting of 7,88,119 phonemes where 419,149 were consonants. The most

occurring vowels were /i/ of about 25% of total vowel occurrences, /e/ and /a/ of about 20% of total vowel occurrences. Among the single phonemic vowels of Romanian, /əə/ and /i/ are the least frequent of all the vowels in this word list, /əə/ accounts for 5.3%, while /i/ constitutes only 1.7% of the total vowel count. The three diphthongs /jɛ/, /ja/ and /wa/ also have very low frequency. Among consonants /r/ is the most occurring for of about 17% of the total occurrences of consonants. Top five occurring consonants in descending order are /r/, /t/, /n/, /l/, /s/. These five consonants make up for about 67% of the total consonants in Romanian. Vowel /i/ appeared frequently with the contexts of /r/ and /n/.

b) Phoneme frequency from spoken data in Non-Indian Languages

Until the 1970s, the studies of phonemes were mainly on the written corpus where the frequencies were mostly calculated from sources like newspapers, journals and scripts of plays. However some of the studies investigated the frequency of occurrence of phonemes in conversations also. Tobians (1959) studied the frequency of occurrence of phonemes in spoken English and compared with another English study by French, Carter, and Koenig (1930). Both the studies shows the predominant occurrence of alveolar /t/ among consonants and /i/ among vowels. There was significant difference between studies in the occurrence of /r/. Similarly Denes (1959) studied spoken English where 72,210 phonemes comprised the entire data. As seen in other studies, consonants occurred highly than vowels. With respect to vowels, front vowels showed predominant occurrence following central and back vowels. Among consonants, plosives were higher than fricatives and nasals. Also voiced consonants were predominant than their voiceless counter parts. Percentage of occurrence of phonemes is in descending order depicted in Table 2.3

Table 2.3

Percentage of occurrence of phonemes in English (Denes 1959)

Vowels	% occurrence	Consonants	% occurrence
ə	9.0445	t	8.4033
i	8.2537	n	7.0849
ai	2.8473	s	5.0893
e	2.8126	d	4.1767
i:	1.7878	l	3.6892
ou	1.7477	m	3.2890
^	1.6701	ø	2.9927
o	1.5330	k	2.8985
a	1.5261	r	2.7697
ei	1.4956	w	2.5661
u:	1.4222	z	2.4927
o:	1.2007	b	2.0842
a:	0.7755	v	1.8515
au	0.7741	p	1.7698
u	0.7672	f	1.7283
ə:	0.6661	h	1.6729
eə	0.4335	j	1.5303
iə	0.2867	ŋ	1.2436
uə	0.1426	g	1.1619
oi	0.0872	ʃ	0.7021
Total	39.2742	θ	0.6955
		dz	0.5138
		tʃ	0.3684
		ʒ	0.0512
		Total	60.7256

Delattre (1965) studied 2000 English syllables from narration and dramatization samples. /ə/ was the most frequently occurring vowel with 20.12% of occurrence. Among consonants, /t/ occurred 12.77% of the time. Consonant /z/ was the least occurring phoneme in the entire data. Another study by Malecot (1974) discussed the computer assisted analysis of a tape recorded corpus of upper class French conversations. He considered 2,01,400 phonemes consisting 91,272 vowels and 1,03,610 consonants. Results showed more distribution of consonants than vowels

and oral vowels were predominant than nasal vowels. Front vowels were higher in occurrence more than back vowels. Labials and apicals comprised 73% of all the consonant articulation. 68% of all occurrences of consonants were voiced and examination of voiced and voiceless cognate's pairs showed a great deal of inconsistency among fricatives. Occurrence of clusters were varied and showed less occurrence in word final compared to word initial position.

Mines, Hanson and Shoup (1978) used a data base containing 103,887 phoneme occurrences taken from casual conversational American English obtained through interviews of sixteen adult males and ten adult females. The speech was transcribed using a quasi-phonemic system, known as ARPAbet, plus selected phoneme alternates and was analyzed with computer assistance to obtain the rank order of phonemes according to frequency of occurrence. The frequency of occurrence of phonemes were listed in descending order as / a, n, t, i, s, r, l, d, ε /, which accounted for 47% of the total data. Vowels and vowel alternates account for 38.1% of the total data, consonants and consonant alternates account for 58.5%, retroflex vowels and retroflex vowel alternates, 2.2%, and syllabics and syllabic alternates, 1.2%. The three nasals /m/, /n/ and /ŋ/ account for 18.45% of all consonants produced in initial, medial and final positions. Liquids /l/ and /r/ were occurring with nearly the same frequency in American English just over 6% of all consonant sounds in adult speech. Stops showed 29.21% of occurrence. Authors reported that the high frequency of occurrence of some phonemes is due to the fact that they occur in one or more very frequently occurring words, such as /a/ in *a*, *uh* and *the*. And the frequency of occurrence of phonemes was found to be related only slightly to age, sex, and level of education or early place of residence of the speakers.

Guirao and Jurado (1990) also studied the occurrence of phonemes in American Spanish. Speech samples included for the study were five modern plays of American Spanish. 1,63,861 phoneme occurrences were analyzed using specific computer softwares. The phoneme frequency distributions were analyzed for single phonemes, syllable position (initial and final) and manner of articulation. Vowels /e/, /a/ and /o/ accounted for 57% of phoneme occurrences, with consonants /s/, /n/, /r/, /l/, /t/, /k/, /d/ accounting for the other 35%. Front half closed vowel /e/, the most frequent of all items, and open central /a/ and back half closed /o/ accounted for 40.0% of all phoneme occurrences. The onset of syllables is marked by stops /t, k, d/ ranging between 8.0 and 10.0%, fricative /s/ is 10.0%. Liquids /l, r/ ranged between 5.0 and 4.0% of the time. Stop /k/ was predominant in the data with 12.7% of occurrence, followed by /a, e/ and fricative /s/, each close to 10%. Stops /p, d, b, t/, lateral /l/ and nasals /m, n/ each with values between approximately 6.0 and 7.5% were obtained. The coda portion of syllables and words were mainly by vowels /e, a, o/. These sounds distributed more or less evenly between 20.0 and 22.0% at the release of the syllable and between 17.0 and 22.0% at the end of words. Fricative /s/ follows with an average of 13% in word and 11% in syllable final positions.

Palai and O'Hanlon (2004) studied the phoneme frequency in Setswana language. The recorded conversations of 118 speakers from the radio programs were analyzed for the frequency of phoneme occurrences. The conversations selected for the study were of two or more people. 49,358 words were analyzed for the total number of phonemes, phonemes at initial position and phonemes at final position. The occurrence of consonant phonemes was 54% at initial position, 42% in medial and 4% in final position. The order of the most occurring phonemes was /l/, /n/, /g/, /r/, /b/, /m/ and /k/. Consonants /n/, /g/ and /b/ most frequently occurred in words at initial

position, /l/, /r/ and /m/ most frequently occurred at medial position of the words and /ŋ/ being the only consonant phoneme occurring at final position. Occurrence of /ŋ/ in medial position (7.2%) was almost twice its occurrence in initial position (4.1%). The second most frequently occurring phoneme was /n/ at 8.9%, with more than half the total number occurring in the initial position (total = 8.9%, 7.1% in the initial position). The third most frequently occurring phoneme was velar /x/ at 8.6%. The percentage of occurrence of phonemes gradually reduced from 8.6% for /x/ to 7.5% for /k/, after which it dropped drastically to 4.8% for ejective alveolar affricate /ts/ and then again reduced gradually to 0.004% for aspirated alveolar affricate /ts^h/.

Thomas (2005) studied five languages and compared the frequency of occurrence of phonemes. They considered Cantonese- 3,191,3; Mandarin- 10,53,863; Italian- 10,45,1644; German- 32,95,546 and American English- 20,94,251 phonemes as data base. Frequency of occurrence of consonants were predominant than vowels in all five languages. Front vowels occurred more frequently than central and back vowels. Unrounded vowels showed high occurrence than rounded vowels in all the five languages. Though with respect to place of articulation, alveolars showed predominant occurrence in all the languages, manner of articulation showed variation across languages; plosives were more in Cantonese, German and American English. And nasals were more in Mandarin and affricates highly occurred in Italian. Voiceless stops were frequent in Cantonese and Mandarin. Mean percentage of occurrence of vowels and consonants in these five languages are depicted in Table 2.4 and Table 2.5 respectively.

Table 2.4

Mean percentage of occurrence of vowels in five different languages (Thomas, 2005)

Vowels	Cantonese	Mandarin	Italian	German	American English
Front	41.96	54.57	52.95	48.39	52.71
Central	21.17	23.88	21.86	18.86	22.98
Back	36.87	21.55	25.19	32.75	24.31
Close	22.7	44.40	26.75	30.59	31.57
Close-Mid	15.19	31.73	0.00	26.38	32.14
Open-Mid	32.79	0.00	45.63	27.44	20.24
Open	29.24	23.88	21.86	15.59	16.04
Rounded	34.79	24.91	25.19	36.01	11.25
Unrounded	44.04	75.09	74.81	45.13	65.77
Neutral	21.17	0.00	0.00	18.86	22.98

Table 2.5

Mean percentage of occurrence of consonants in five different languages (Thomas, 2005)

Consonants		Cantonese	Mandarin	Italian	German	American English
Place of articulation	Bilabials	8.81	17.20	11.52	10.00	15.08
	Labiodentals	2.46	2.44	4.70	7.53	6.35
	Alveolars	43.77	25.25	71.52	55.23	57.06
	Palatals	10.60	14.42	0.84	2.49	1.08
	Velars	24.12	12.99	7.62	9.65	7.38
	Glottal	10.25	17.37	13.63	1.69	3.64
Manner of articulation	Fricatives	19.67	10.33	6.11	23.97	30.27
	Affricates	13.41	19.55	36.65	2.57	1.71
	Plosives	32.85	22.74	18.84	32.95	31.17
	Nasals	9.31	23.88	13.14	22.82	18.48
	Laterals	10.62	4.63	11.04	6.09	6.67
	Approximants	14.15	4.75	0.59	0.45	11.70
Voicing features	Voiced	34.08	33.83	59.78	63.55	63.16
	Voiceless	65.92	66.17	40.22	36.45	36.84

c) Phoneme Frequency: Written vs spoken data

There are studies conducted separately for written and spoken forms for a particular language. Some researchers have extended their interest to compare variations in speaking and written modalities due to the disparity with respect to frequency of occurrence of phonemes.

Crystal (1995) compared the frequency of occurrence of phonemes in English from both written and spoken data. The written data considered by Crystal was from Mitton (1992) which consisted data from dictionary whereas spoken data was taken from a study by Fry (cited in Crystal, 1995). Crystal compared the frequency of occurrence of phonemes of both of these

studies and found there were considerable variations when comparing the spoken and written data. In written data vowel /i/ (10.5%) was found to be the most frequently occurring phoneme whereas in spoken data it was vowel /ə/ (8.3%). Also the vowel /u/ which has less than 1% occurrence in written data was found to occur more frequently in spoken data (1.2%). When written and spoken data were compared for consonants there were variations that are similar to vowels. Consonant /n/ highly occurred in spoken data whereas /t/ was high in written data. Occurrence of /p/ was very less in spoken compared to written mode, which occurred 3.1% of the time in written corpus.

There exist observable differences in frequency of occurrence of phonemes between written and spoken data in the same language. Sandoval, Toledano, de la Torre, Garrote and Guirao (2008) studied and compared the syllabic and phonemic frequency in spoken and written context in Castilian Spanish. Vowels /a/, /e/ and /o/ were the most frequent phonemes in both spoken and written contexts. Vowel /a/ occurred for about 12.2% in spoken and 12.8% in written data and /e/ occurred for about 15% in spoken and 12% in written context. Vowel /o/ occurred about 10% and 9% in spoken and written contexts respectively. Among consonants /s/ occurred about for 8% and 7% in spoken and written contexts respectively. Followed by /s/ was /n/ which occurred for about 7% in spoken and written contexts. /l/ and /d/ had similar frequency of occurrence of about 4% in spoken and 5% in written data. Overall the order of some phonemes and their percentage of occurrence were different across modality. Number of occurrences and mean percentage of occurrence of vowels and consonants in both spoken and written modes are depicted in Table 2.6.

Table 2.6

Number of occurrences and mean percentage of occurrence of vowels and consonants in spoken and written modes in Castilian Spanish (Sandoval et al, 2008)

Phoneme	Spoken		Written	
a	152664	12.27	323783	12.89
b	31126	2.50	64170	2.55
ɐ	18940	1.52	50301	2.00
ɸ	3744	0.30	4463	0.18
d	54284	4.36	136187	5.42
e	188196	15.12	320140	12.74
f	6217	0.50	23042	0.92
g	11359	0.91	26138	1.04
i	89799	7.22	190756	7.59
x	7681	0.62	19362	0.77
k	55863	4.49	95427	3.80
l	56107	4.51	137148	5.46
m	39278	3.15	69445	2.76
n	87775	7.05	178012	7.09
ɲ	2427	0.19	7729	0.31
o	129308	10.38	234238	9.32
P	34135	2.74	68687	2.73
r	5236	0.42	25016	0.99
ɾ	63702	5.12	155632	6.19
s	100881	8.11	184085	7.33
t	56287	4.52	108398	4.31
u	39146	3.14	76390	3.04
ʎ	10356	0.83	13307	0.53

d) Frequency of occurrence of phonemes in Indian Languages

India as a multilingual country, it is essential to obtain the frequency of phonemes in a language to understand the languages better. One of the initial studies on Indian languages in this context is by Bhagwat (1961) who compared the phonemic and morphemic frequencies in Marathi. Ghatage and Madhav (1964) calculated the phonemic and morphemic frequencies in Hindi using written source of materials. Results showed that vowels occurred more frequently than consonants.

Pandit (1965) studied spoken and written Gujarati. Data base consisted of 28,500 and 21,500 words from written and spoken materials respectively. Results showed that occurrence of consonants were predominant than vowels. /a/ was the most frequently occurring vowel and /r/ was the most frequently consonant. Vowel /ɔ/ was the least occurring phoneme in the entire data. Kelkar (1994) studied phonemic and morphemic frequencies in Oriya and indicated that /ə/ was the most occurring vowel followed by /a/ and /I/. Among consonants /r/, /k/ and /t/ were the most found phonemes in Oriya.

Frequency of occurrence of phonemes in Kannada was first studied by Ramakrishna (1962) in written form. The results revealed that long vowels and aspirated phonemes are used relatively less frequently and vowel /a/ is the highest occurring vowel and consonants like /r/, dentals /d/ and /t/ are the highly used consonants in Kannada. Research on relative frequency of phonemes and morphemes in Kannada was carried out by Ranganatha (1982). He considered 1,00,000 words from different written sources. The data were collected under three main heads namely: fiction (novels, short stories, and dramas), non-fiction (biographical, scientific) and newspaper and periodicals (dailies, weeklies etc). 96,234 words, out of which 6,01,985

phonemes accounted for the total data. Results of his study indicated that vowels /a/, /i/ and /u/ were the most occurring phonemes. /a/ occurred for 17.7%, /i/ occurred 7.8% and /u/ 6.6% of the total data. /r/ was the most occurring consonant occurring for 4.9% of the total data followed by /d/ (4.8%) and /n/ (3.7%). Vowels constituted 48.6% of the total data and consonants constituted 51.3%. Diphthongs /ai/ and /au/ constituted 0.1% of the total data. The rank order of frequency of occurrence of phonemes in Kannada were /a/, /i/, /u/, /r/, /d/, /a:/, /e/, /n/, /t/, /k/, /g/ and /v/ according to Ranganatha (1982).

Jayaram (1985) carried out a study for calculating the frequency of occurrence of phonemes in Kannada which considered written data from many newspapers, books, magazines. These sources of data were published during the period of 1974-1976. The corpus consisted of 7,60,792 phoneme occurrences from 1,00,000 words. He listed out the frequency of occurrence of phonemes and the order was /a, I, n, r, u, d, ə, e, t, l/. The short vowels /a/, /i/, /u/ were more frequent than their longer counterparts. The unaspirated consonants /t/ and /d/ were occurring more frequently than aspirated /t^h/, /d^h/. The most frequently occurring phonemes accounted for 67% of the total data. All the consonants constituted for 54.3% of the total data. /a/ was the most frequently occurring vowel followed by /i/, /u/ and /e/. Vowels constituted 43.5% of the total data.

Manjula, Geetha, Sharath and Antony (2012) calculated the frequency of occurrence of phonemes in Kannada to construct phonemically balanced bisyllabic word list for audiological evaluations. They considered a corpus of 15,000 phoneme occurrences including both written as well as spoken data. Conversations lasting for ten minutes and lectures which were audio recorded accounted for the spoken data which included about 5,000 phoneme counts, and written sources like newspapers, journals etc accounted for about 10,000 phoneme counts. Results

indicated that the most frequently occurring vowel was /a/ followed by /i/. Among consonants /n/, /r/, /t/, /l/ were the most frequently occurring phonemes.

Recently, Sreedevi, Smitha and Vikas (2012) considered conversational Kannada where adult native speakers of Mysore dialect of Kannada participated. A total of 74 subjects including 28 females and 46 males were considered for the study. A total of 2,68,625 phonemes were obtained from the recorded speech sample. In that 44.37% phonemes accounted as vowels and 55.3% were consonants. Vowel /a/ was the most frequently occurring phoneme followed by /n/, /l/, /e/, /r/, /a:/, /d/, /l/, /u/, /g/ and /k/. Diphthongs and consonants such as /h/, /s/, /p/, /ʃ/, /dʒ/, /ʒ/ were less frequent in conversational Kannada. Frequency of occurrence of phonemes of Kannada compared with the present study is depicted in Appendix 1.

In Telugu, Kalyani and Sunitha (2009) studied 12,378 syllables from conversation sample. Results showed that consonants (51.21%) were predominant in spoken data than vowels (44.98%). Open vowels were higher in frequency and among consonants, alveolars showed significant occurrence of 45.62% followed by bilabials and velars. Glottal fricative showed fewer occurrence of 0.82%. Frequency of occurrence of vowels were classified as CF- closed front, HCF- high closed front, CB- closed back, HCB- high closed back and shown in Figure 2.1. Frequency of occurrence of consonants in the same study with respect to places of articulation is depicted in Figure 2.2.

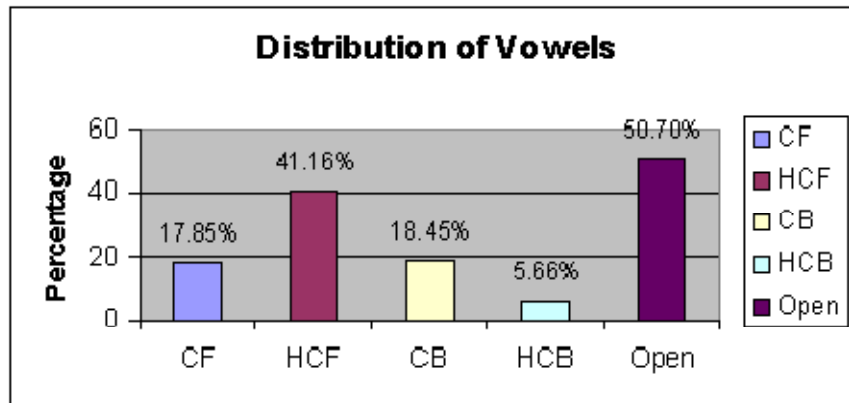


Figure 2.1. Occurrence of vowels in Telugu (Kalyani & Sunitha, 2009)

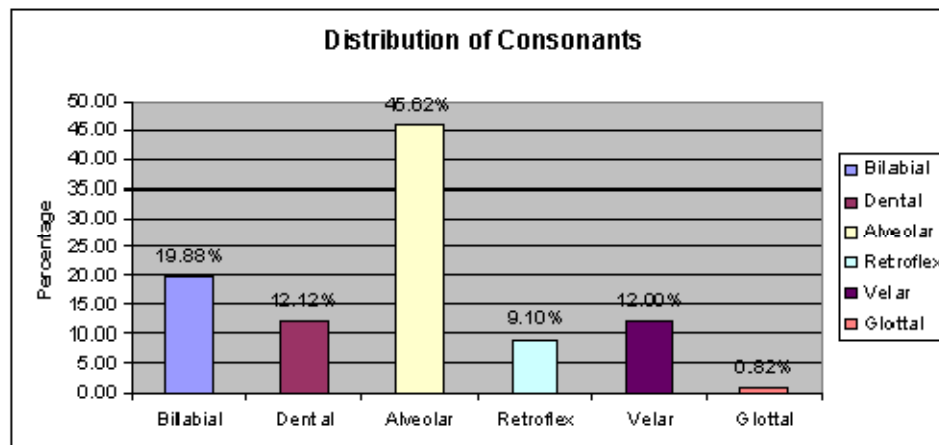


Figure 2.2. Occurrence of consonants in Telugu (Kalyani & Sunitha, 2009)

Another study in Telugu by Kumar and Mohanty (2012) collected 500 words from different sources like periodicals, newspapers, magazines, journals, general books, phonetic books and spontaneous speech. Similar results found as reported by Kalyani and Sunitha (2009). Also results showed reduced aspirated stops in Telugu spoken mode of communication.

Ramakrishna, Nair, Chiplunka, Atal, Ramachandran and Subramanian (1962) studied the relative frequencies of the different speech sounds in some of the Indian languages including Hindi, Marathi, Tamil, Malayalam, Telugu and Kannada. The data was obtained from a sample of 20,000 speech sounds in each language. Consonants predominantly occurred than vowels in all the six languages. Vowel /a/ was the most occurring phoneme with more than 10% occurrence in all language groups. The most frequently occurring consonant was /k/ in Hindi and Tamil (Hindi- 5.96%; Tamil- 9.19%), /n/ in Malayalam, Telugu and Kannada (Malayalam- /n/- 4.14%; Telugu- /n/- 6.35%; Kannada- /n/- 4.9%) and /r/ in Marathi (4.48%). Occurrence of voiceless sounds was more than their voiced counterparts in all the languages. Occurrence of aspirated sounds was minimal. Ghatage (1994) studied phonemic and morphemic frequencies in Malayalam using 1,00,000 words from various written materials such as novels, short stories, poems, essays, periodicals. The results indicated /a/ and /I/ were the most frequent vowels. Among consonants, bilabial nasal /m/ was most occurring followed by /n/ and /k/ respectively. Phonemes such as /z, f/, diphthongs and all the aspirated phonemes occurred less than 1%. Aspirated /j^h/ was the least occurring phoneme in the entire data

Sreedevi and Irfana (2013) studied Calicut dialect of Malayalam which included seven conversational groups. The data was collected through conversations in controlled natural environments for about 25 to 30 minutes of duration. After transcription of each session, the obtained data, 83,561 phonemes were analyzed and results showed that consonants (57.26%) constituted larger part of the total corpus than vowels (42.54%). The most frequently occurring vowels were /a, I, ə, a:, e/ and consonants in descending order were /k, n , ʈ, t, l/. Aspirated sounds and long vowels were rarely present in the data.

Hence the review of literature on frequency of occurrence of phonemes indicates that it has been studied in various languages across the world. Data from both written as well as spoken sources are available in several languages. The phoneme frequency varies across languages and also when the data is compared across written and spoken sources. A common finding from most of the studies across the Indian and non-Indian languages is that consonants were high in occurrence compared to vowels. However the order of the consonants and vowels vary widely across languages. Hence the present study aimed to obtain the frequency of occurrence of phonemes in three major regional dialects of Malayalam.

CHAPTER III

METHOD

The present study is planned to obtain the frequency of occurrence of phonemes in conversational speech samples in the three major regional dialects of Malayalam: Calicut (D1), Eranakulam (D2) and Thiruvananthapuram (D3).

1. To obtain a database with the occurrence of various phonemes in the three major dialects of Malayalam
2. To calculate the frequency of occurrence of each phoneme from the database obtained
3. To compare of phoneme frequency across dialects
4. To obtain the mean frequency of occurrence of phonemes in Malayalam language
5. To compare the frequency of occurrence of phonemes in Malayalam with previous studies in other languages

Participants: Fluent adult native speakers of Malayalam in the age range of 30 to 55 years with a minimum of 10 to 12 years of schooling in Malayalam medium of instruction were selected for the study. Three dialects of Malayalam including Calicut, Eranakulam and Thiruvananthapuram were considered for the study. Data consisted of seven recordings each from the region of the three dialects. This amounted to a total of 21 conversational recordings. Each recording session had 4 to 5 participants engaged in conversation. The study included a total of 101 participants out of which 47 were males and 54 were females. The participants did not have a clinical background of any known diseases or disorders. All participants were informally assessed and

were ensured to have normal speech, language, hearing abilities. Table 3.1 shows the details of the participants included in each recording session.

Table 3.1

Number of participants in each recording session

Recording Sessions	R1	R2	R3	R4	R5	R6	R7
No.of Participants (D1)	5	5	5	5	5	5	6
Males	0	1	2	3	1	2	4
Females	5	4	3	2	4	3	2
No.of Participants (D2)	4	5	4	5	5	5	5
Males	2	2	3	2	4	3	1
Females	2	3	1	3	1	2	4
No.of Participants (D3)	5	4	5	5	4	4	5
Males	2	2	3	1	3	2	4
Females	3	2	2	4	1	2	1

*R indicates the number of the recording session

Instrumentation: A digital recorder (Olympus WS 100) was used for recording the conversation samples. Dell (Vostro1014) laptop with headphones (i-ball 1600 MV) was used for the transcription and analysis.

Procedure: The demographic details, education and the general medical history of all the participants were noted prior to the conversation recording. The data was collected through conversations in controlled natural environments for about 25 to 30 minutes of duration. All the participants were made to sit comfortably facing each other and asked them to speak as natural as

possible. The digital recorder was kept at equidistance from all the speakers. Participants themselves initiated conversation without any specific topic being assigned to them. This was to avoid the high occurrence of certain phonemes and also the purpose of the study was to obtain the frequency of phonemes on conversational speech. Hence a uniform/common conversational topic was not compulsory for the sample recorded. They were instructed to avoid words from other languages and to speak naturally in Malayalam only. They were not restricted from using commonly used loan English words (E.g.: Bus, ticket, phone, car etc). Appendix 2 shows these loan English words considered in the analysis of the study. Table 3.2 shows the topic of conversation for each recording session.

Table 3.2

Topic of conversation in each recording session for the three dialects

	Recording Sessions	Topic of Conversation
D1	Recording 1	General talk about home and surroundings.
	Recording 2	Current education system and political issues.
	Recording 3	Current educational system and general talk about office works.
	Recording 4	General conversation and government rules.
	Recording 5	Corruption and price hike.
	Recording 6	Education of children at present.
	Recording 7	Political scenario in Kerala and women's right.
D2	Recording 1	Political scenario in Kerala and need for education.
	Recording 2	Current youth issues, on mobile phone and internet usage.
	Recording 3	Political issues and foot ball.
	Recording 4	Fancy stores and general conversation.
	Recording 5	Cricket and youth and current state of government offices.
	Recording 6	Problems faced by women.
	Recording 7	Movies and other entertainments and public issues.
D3	Recording 1	Marriages in Kerala and general talk about life style changes.
	Recording 2	General conversation.
	Recording 3	Fashion and youth and usage of mobile phones.
	Recording 4	Political parties and issues regarding problems in the nation.
	Recording 5	Hospital and travelling facilities in Kerala.
	Recording 6	General talk on home and school.
	Recording 7	Current political issues and general conversation.

Data Analysis: The conversation samples were transcribed using International Phonetic Alphabet for Regional languages (Malayalam) by Asher and Kumari (1997) which is provided in APPENDIX 3. The repetitive words and exclamatory remarks were excluded from analysis. Loan words that are commonly used (e.g: bus, car, bat, ticket etc) in day to day conversation were considered in the analysis.

The raw data obtained was analyzed using freely downloadable software SALT* RV version 9 for the frequency count. Using SALT, an individual's language sample may also be compared to a reference database of language measures. It helps in managing the steps of word count, phoneme count, which can be preloaded with the editable database. So a database of Malayalam phonemes was prepared and saved in the editable standard wordlists of SALT software. The database consisted of all the phonemes available in Malayalam adapted from International Phonetic Alphabet for Regional languages (Malayalam) by Asher and Kumari (1997). The conversation data was analyzed through the following steps.

1. Transcribed word data copied and pasted in to another word file to put space between each phoneme.
2. Edited file is loaded into SALT software.
3. The SALT software identifies and compares the loaded phoneme file with editable standard wordlists and provides the phoneme count.
4. The obtained frequency count is copied in to excel sheet for further analysis based on places of articulation, manner of articulation and voicing features.

Inter-judge Reliability: Three post graduate speech language pathologists who are trained in transcription served as judges for determining inter judge reliability measures. Judges were instructed about the phonetic coding procedure and the material to be transcribed before the actual transcription procedure. 10% of each sample was transcribed by each of the three judges. The recorded samples were played to the judges individually in approximately 2-223 sittings based on the judge's convenience. They were restricted from discussion about the transcription of the sample before or after the task. The frequency of occurrence of phonemes from the transcribed data of the three judges was subjected to statistical analysis. Cronbach alpha reliability index (alpha) indicated a correction of 0.89 for inter judge reliability.

Intra-judge Reliability: For intra judge reliability, a randomly selected 10% of each recording session was separately transcribed and analyzed by one of the researchers (a speech language pathologist). Cronbach alpha score of 0.92 was obtained for intra judge reliability.

* *SALT (Systematic Analysis of Language Transcripts), is a computer program designed to help in analyzing and interpreting language samples from one or more speakers during a communicative interaction. It can be used to analyze samples from everyday speech like conversation and narration. The SALT program provides clinicians and researchers with the means to transcribe language samples into a common format and to compute a series of general analyses of lexical, syntactic, semantic, pragmatics, rate, fluency, and error categories. This software analyses the transcribed sample for different parameters such as MLU (Mean Length of Utterance), NDW (Number of Different Words), TTR (Type Token Ratio) etc. It also gives information about the frequency of words, morphemes, grammatical categories, etc.*

CHAPTER IV

RESULTS

The present study aimed to obtain the frequency of occurrence of phonemes in conversational speech samples in three major regional dialects of Malayalam including Calicut (D1), Eranakulam (D2) and Thiruvananthapuram (D3) with the following objectives:

1. To obtain a database of various phonemes in the three major dialects of Malayalam
2. To calculate the frequency of occurrence of each phoneme from the database obtained
3. To obtain the order of frequency of phonemes in the three dialects separately
4. To compare the phoneme frequency across dialects
5. To obtain the mean frequency of occurrence of phonemes in Malayalam language including all the three dialects
6. To compare the frequency of occurrence of phonemes in Malayalam with earlier studies in other languages

Data included audio recorded samples of 101 fluent adult native speakers of Malayalam in the age range of 30 to 55 years. Three major dialects of Malayalam including Calicut, Eranakulam and Thiruvananthapuram were considered for the study. A total of 21 conversational recordings were considered including seven recordings from each dialect. Each recording session had 4-5 different participants engaged in conversation. The recorded data was transcribed phonetically using the International Phonetic Alphabet (IPA) for regional languages (Malayalam) (Asher & Kumari, 1997). The obtained data was analyzed in the SALT software (RV version 9)

for the frequency count. Further analysis was carried out keeping in view the above mentioned objectives. The results are discussed under the following sections:

1. Frequency of occurrence of phonemes in Calicut dialect of Malayalam
2. Frequency of occurrence of phonemes in Ernakulam dialect of Malayalam
3. Frequency of occurrence of phonemes in Thiruvananthapuram dialect of Malayalam
4. Comparison of frequency of occurrence of phonemes across the three dialects of Malayalam
5. Overall frequency of occurrence of phonemes in Malayalam

1. Frequency of occurrence of phonemes in Calicut dialect of Malayalam

Total number of phonemes recorded in each of the seven conversational groups of Calicut dialect of Malayalam are shown in Fig 4.1; maximum number of phonemes were recorded in the sixth session (R6), which amounted to 14,358 phonemes and minimum number of phonemes were recorded from fifth recording (R5-10,199 phonemes). Each recording session elicited more than ten thousand phonemes and the grand total was 83,561 phonemes from the 7 recordings.

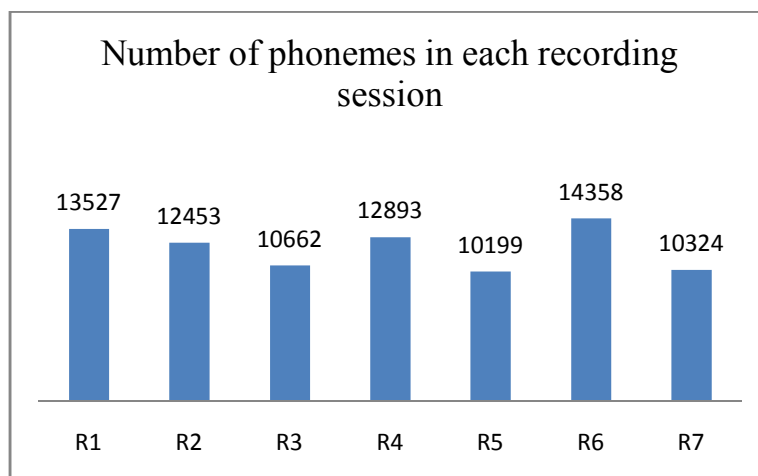


Fig 4.1. Number of phonemes obtained in each recording session

The transcribed data consisted a total of 83,561 phonemes including both consonants and vowels. Based on descriptive statistical analysis, the mean percentage of occurrence of consonants was more than vowels. Mean occurrence of consonants was 57.26% and vowels was 42.54% of the total corpus as shown in Fig 4.2. Occurrence of diphthongs were fewer in all the seven recording sessions.

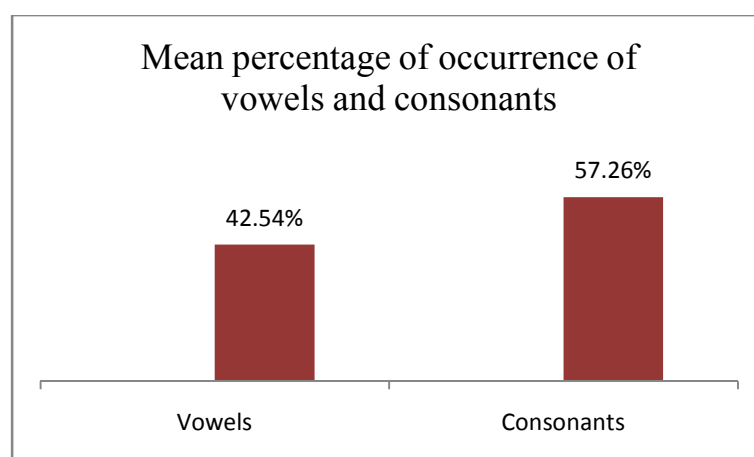


Fig 4.2. Mean percentage of occurrence of vowels and consonants in Calicut dialect.

The mean percentage of occurrence of each phoneme in the corpus is depicted in Table 4.1. Vowel /a/ occurred most frequently in all the recordings. The mean percentage of occurrence of /a/ was more than 10% in all the seven recordings. Some phonemes showed highly reduced frequency of occurrence such as aspirated sounds. /ɟ^h/ was not present in the entire data.

Table 4.1

Mean percentage and standard deviation of occurrence of vowels and consonants in spoken Calicut dialect of Malayalam in alphabetical order

Vowels	Mean % (SD)	Consonant	Mean % (SD)	Consonant	Mean % (SD)	Consonant	Mean % (SD)
/a/	13.83 (1.3)	/k/	7.06 (0.73)	/t̪/	4.20 (0.44)	/ŋ̪/	1.53 (0.7)
/a:/	4.31 (1.74)	/k ^h /	0.02 (0.02)	/t̪ ^h /	0.12 (0.06)	/m/	3.73 (0.5)
/i/	6.07 (0.24)	/g/	0.21 (0.08)	/d̪/	0.37 (0.20)	/j/	2.68 (0.31)
/i:/	0.71 (0.17)	/g ^h /	0.01 (0.01)	/d̪ ^h /	0.38 (0.18)	/r/	2.65 (0.34)
/u/	3.46 (0.6)	/c/	1.59 (0.35)	/p/	3.24 (0.39)	/l/	4.04 (0.39)
/u:/	0.64 (0.2)	/c ^h /	0.002 (0.00)	/p ^h /	0.024 (0.01)	/v/	2.19 (0.14)
/e/	3.75 (0.45)	/ɟ/	0.19 (0.11)	/b/	0.46 (0.10)	/s/	1.35 (0.21)
/e:/	0.88 (0.42)	/ɟ ^h /	-	/b ^h /	0.17 (0.08)	/ʂ/	0.30 (0.06)
/ai/	0.15 (0.07)	/t̪/	4.44 (0.5)	/ŋ̪/	1.48 (0.32)	/ʃ/	0.44 (0.14)
/o/	1.23 (0.49)	/t̪ ^h /	0.10 (0.10)	/p̪/	0.78 (0.31)	/h/	0.15 (0.05)
/o:/	1.67 (0.39)	/d̪/	0.09 (0.06)	/ŋ̪/	2.49 (0.2)	/l̪/	1.89 (0.25)
/au/	0.027 (0.02)	/d̪ ^h /	0.007 (0.00)	/n/	5.32 (1.07)	/z̪/	0.24 (0.12)
/ə/	5.58 (1.24)	/r/	2.01 (0.20)	/T/	0.53 (0.13)	/f/	0.035 (0.06)

Occurrence of short vowels were considerably higher than long vowels. Short open vowel /a/ (13.83%) was highly predominant than short closed front vowels (6.78%) and closed back vowels (4.00%). Fig 4.3 shows the mean percentage of occurrence of long and short vowels.

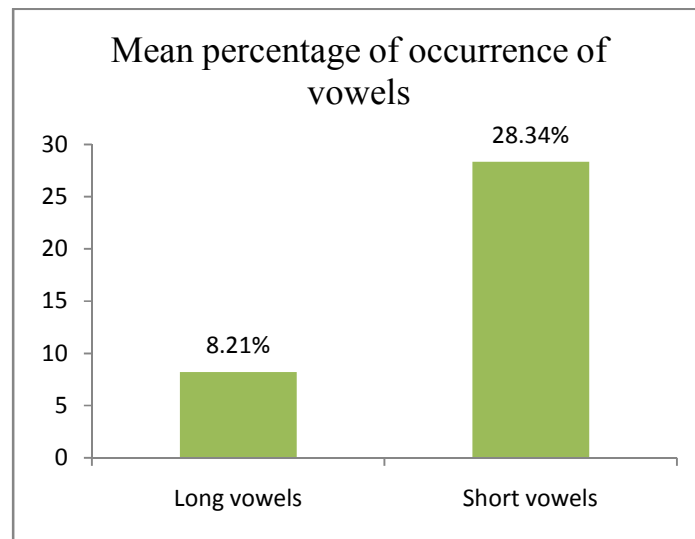


Fig 4.3. Mean percentage of occurrence of long and short vowels in Calicut dialect.

Considering single consonants, all the places of articulation except glottal /h/ showed significant occurrence in the conversational corpus and voiceless velar /k/ showed highest percentage of occurrence, followed by alveolar nasal /n/ and retroflex /ʈ/ respectively. According to consonant classes based on place of articulation, dentals showed predominant occurrence followed by labials, retroflex and velars. Fig 4.4 depicts the mean percentage of occurrence of consonants based on places of articulation.

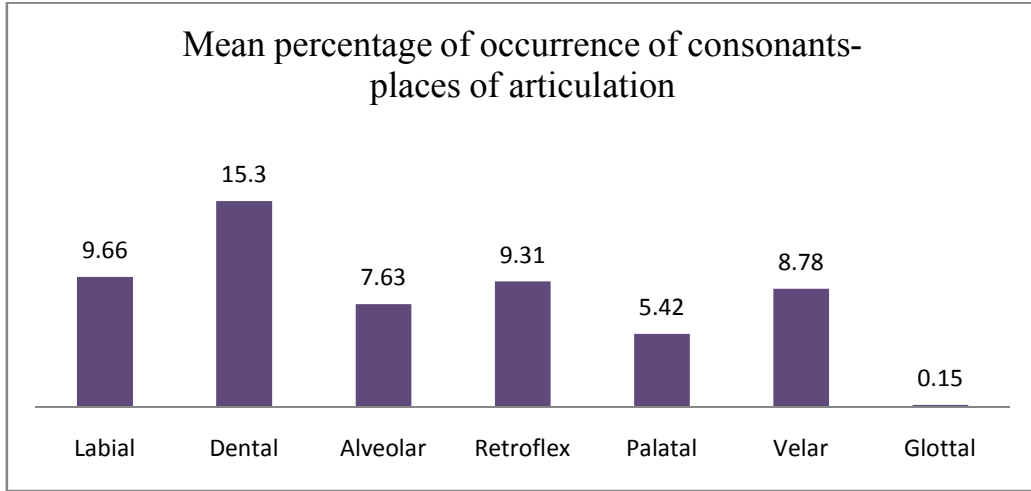


Fig 4. 4. Mean percentage of occurrence of consonants based on places of articulation in Calicut dialect.

Based on manner of articulation, stops showed predominant occurrence followed by nasals, laterals, glides and trills. Fricatives were the lowest occurring manner of articulation. Fig 4.5 shows the mean percentage of occurrence of consonants based on manner of articulation.

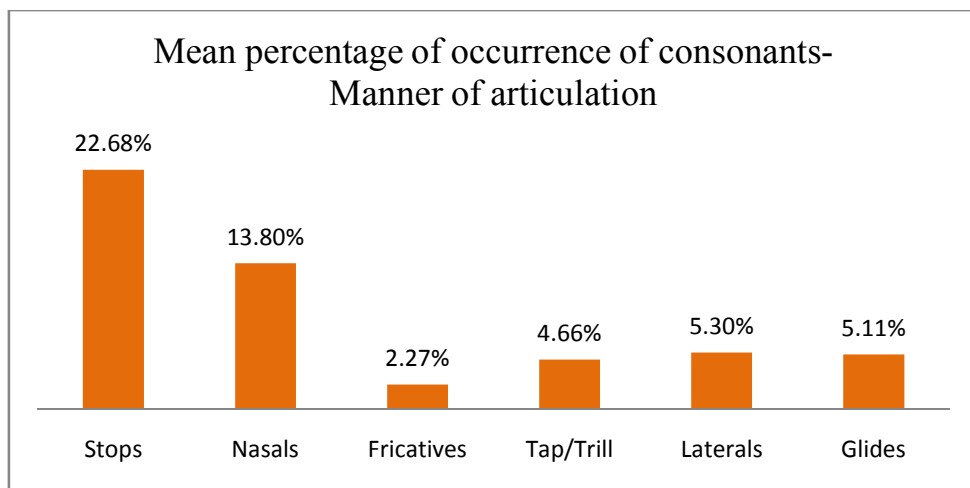


Fig 4.5. Mean percentage of occurrence of consonants based on manner of articulation in Calicut dialect.

Among stops, unvoiced unaspirated stops were significantly higher in occurrence compared to their voiced unaspirated counterparts. Occurrence of unvoiced velar consonant /k/ was highest and occurred for 7.06% of the total phonemes. Next in order was unvoiced retroflex stop /ʈ/ (4.44%) followed by unvoiced dental stop /ʈ/ (4.20%). Unvoiced palatal affricate /c/ was comparatively reduced in the sample. Respective voiced counterparts of the above mentioned consonants were highly reduced in all the seven recordings. Occurrence of aspirated phonemes also significantly reduced compared to their unaspirated counterparts. Unvoiced unaspirated stops accounted for 20.53% and voiced unaspirated stops occurred for only a mere 1.32% in the conversational corpus. Voiced aspirated stops (0.58%) showed relatively higher occurrence than unvoiced aspirated stops (0.3%). Fig 4.6 indicates the mean percentage of occurrence of stops in the total corpus obtained.

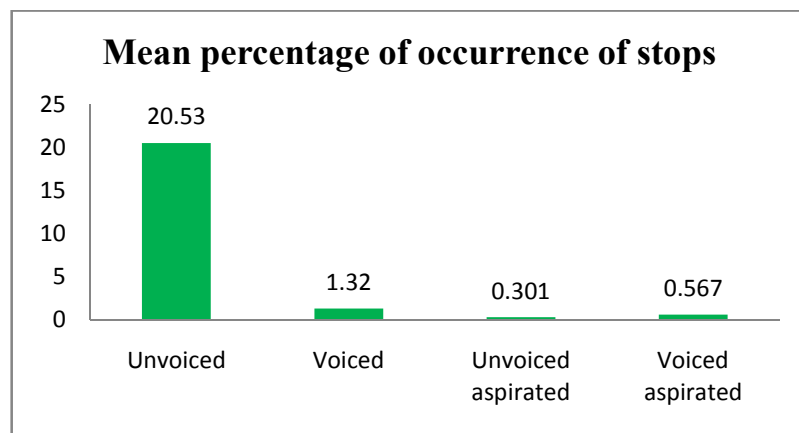


Fig 4.6. Mean percentage of occurrence of stops in Calicut dialect

All nasal phonemes showed their presence in the corpus with significant percentage of occurrence. Alveolar /n/ occurred for 5.32% followed by bilabial /m/ with 3.73% of occurrence. Palatal /ɲ/ was less found than other nasal sounds. Malayalam is known to have the maximum places of articulation among nasals and Calicut dialect showed all these six places of nasals.

Among fricatives /s, ʃ, ʒ, h/, unvoiced alveolar fricative /s/ had the highest occurrence of 1.35% of the data. Glottal fricative /h/ occurred for 0.15% of the total corpus and it was the least seen among fricatives.

Liquids including both tap/trills and laterals occurred in the conversational corpus with considerable percentage of occurrence. Lateral retroflex /ɭ/ occurred for 1.89% and lateral alveolar /l/ occurred for 4.04%. Tap /ɾ/ and trill /r/ had similar percentage of occurrences which were 2.01% and 2.65% respectively. Among the approximants, /z/ showed 0.24% of occurrence. Labial /v/ and palatal /j/ occurred for less than 3% of the total phoneme occurrences.

As per the results, the mean percentage of occurrence of the ten prominent phonemes is shown in Fig 4.7. The most frequently occurring vowels were /a, I, ə, a:, e/ and consonants were /k, n, ʈ, t, l/. As discussed above, low-central vowel /a/ occurred more frequently followed by velar stop /k/. Velar /k/ was produced for 7.06% of the total corpus of all seven recordings. Followed by /a/ and /k/, high-front vowel /i/, mid- central vowel /ə/ and alveolar nasal /n/ showed prominent occurrence. Appendix 4 presents the occurrence of all phonemes in the Calicut dialect (D1) in descending order.

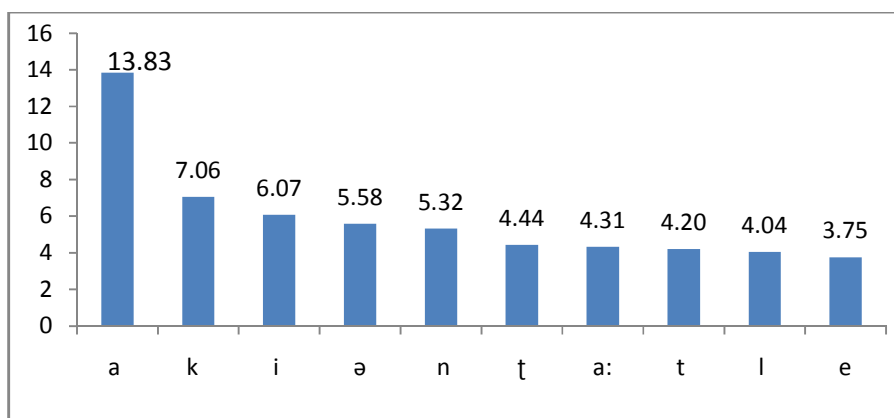


Fig 4.7. Mean percentage of frequency of occurrence of ten most frequently occurring phonemes in Calicut dialect

2. Frequency of occurrence of phonemes in Eranakulam dialect of Malayalam

Based on the SALT analysis, the total number of phonemes obtained were 74,114 from the seven recordings in Eranakulam dialect. Maximum number of phonemes were recorded in the first session (R1), which amounted to 11,608 phonemes and minimum phonemes were recorded from the second (R2- 9,739 phonemes). Except R1 and R2 the remaining five recordings elicited more than ten thousand phonemes. The numbers of phonemes obtained in each recording session are depicted in Fig 4.8.

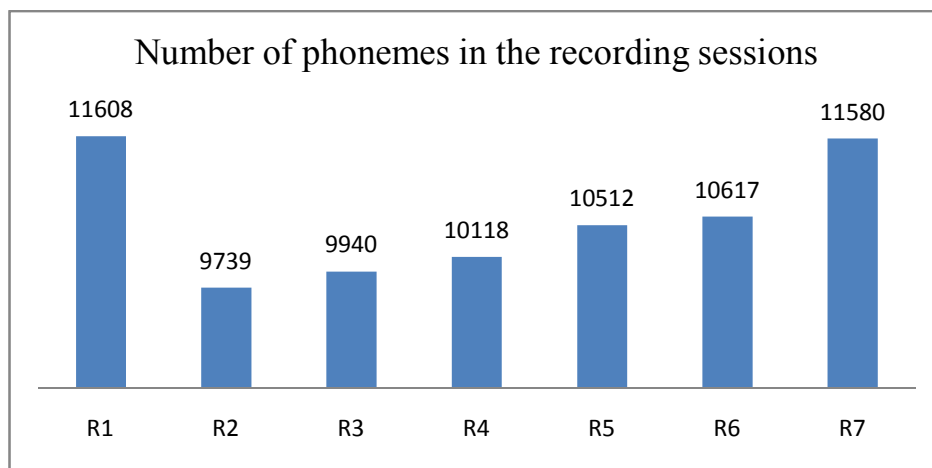


Fig 4.8. Number of phonemes obtained in each recording session in Eranakulam dialect.

The transcribed data consisted a total of 74,114 phonemes including both consonants and vowels. Based on descriptive statistical analysis the mean percentage of occurrence of consonants were more than vowels. Mean occurrence of consonants was 54.1 % and vowels, was 43.48% of the total corpus as shown in Fig 4.9. Occurrence of diphthongs including /ai/ and /au/ were 1.62 % and it was fewer in all the seven recording sessions.

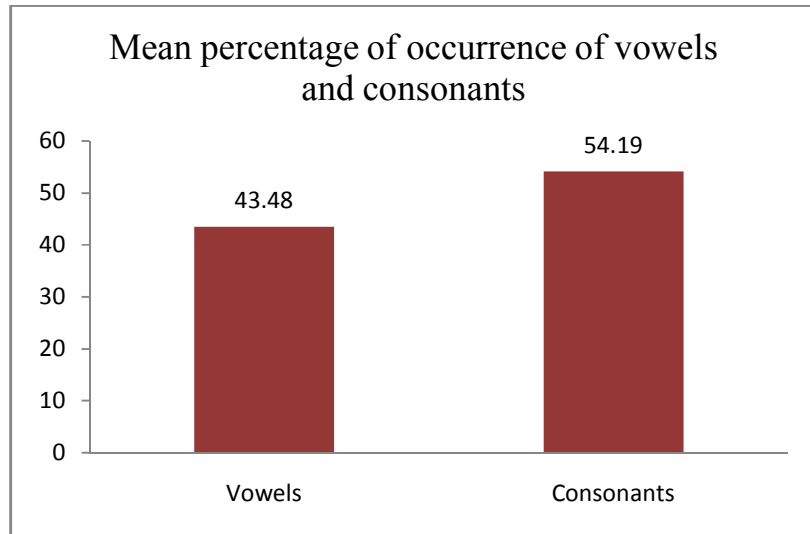


Fig 4.9. Mean percentage of occurrence of vowels and consonants in Eranakulam dialect.

The mean percentage of occurrence of each phoneme is depicted in Table 4.2. Vowel /a/ occurred most frequently in all the recordings. Similar to the Calicut dialect the mean percentage of occurrence of /a/ was more than 10% in all the seven groups. Some phonemes such as /p^h/, /t^h/, /d^h/ and /f/ showed highly reduced frequency of occurrence. Aspirated consonants /c^h/, /j^h/ and /d^h/ were not present in the entire corpus.

Table 4.2

Mean percentage and standard deviation of occurrence of vowels and consonants in spoken Eranakulam dialect of Malayalam in alphabetical order

Vowels	Mean % (SD)	Consonant	Mean % (SD)	Consonant	Mean % (SD)	Consonant	Mean % (SD)
/a/	13.22 (0.59)	/k/	5.40 (2.18)	/t̪/	3.88 (0.44)	/ŋ̪/	1.81 (0.74)
/a:/	5.01 (0.53)	/k ^h /	0.007 (0.013)	/t̪ ^h /	0.03 (0.03)	/m/	3.75 (0.22)
/i/	6.53 (0.42)	/g/	0.14 (0.063)	/d̪/	0.27 (0.12)	/j/	2.66 (0.23)
/i:/	0.60 (0.18)	/g ^h /	0.007 (0.01)	/d̪ ^h /	0.07 (0.04)	/r/	2.59 (0.33)
/u/	2.81 (0.25)	/c/	1.85 (0.39)	/p/	3.89 (0.40)	/l/	3.79 (0.32)
/u:/	0.64 (0.8)	/c ^h /	-	/p ^h /	0.01 (0.03)	/v/	2.61 (0.32)
/e/	4.40 (0.42)	/ɟ/	0.20 (0.12)	/b/	0.36 (0.18)	/s/	1.11 (0.13)
/e:/	1.18 (0.16)	/ɟ ^h /	-	/b ^h /	0.11 (0.04)	/ʃ/	0.27 (0.05)
/ai/	0.09 (0.03)	/t̪/	4.35 (1.79)	/ŋ̪/	1.56 (0.38)	/j̪/	0.36 (0.08)
/o/	1.52 (0.21)	/t̪ ^h /	0.10 (0.05)	/p̪/	1.18 (0.22)	/h/	0.11 (0.04)
/o:/	1.76 (0.26)	/d̪/	0.11 (0.07)	/ŋ̪/	2.41 (0.49)	/l̪/	1.76 (0.26)
/au/	0.04 (0.03)	/d̪ ^h /	-	/n̪/	5.06 (0.70)	/z̪/	0.42 (0.07)
/ə/	5.88 (0.63)	/r/	1.79 (0.12)	/T/	0.65 (0.11)	/f/	0.07 (0.04)

As in D1, occurrences of short vowels were considerably more than long vowels. Open vowel /a/ (13.22%) was highly predominant than closed front vowels (7.13%) and closed back vowels (3.45%). Fig 4.10 shows the mean percentage of occurrence of long and short vowels.

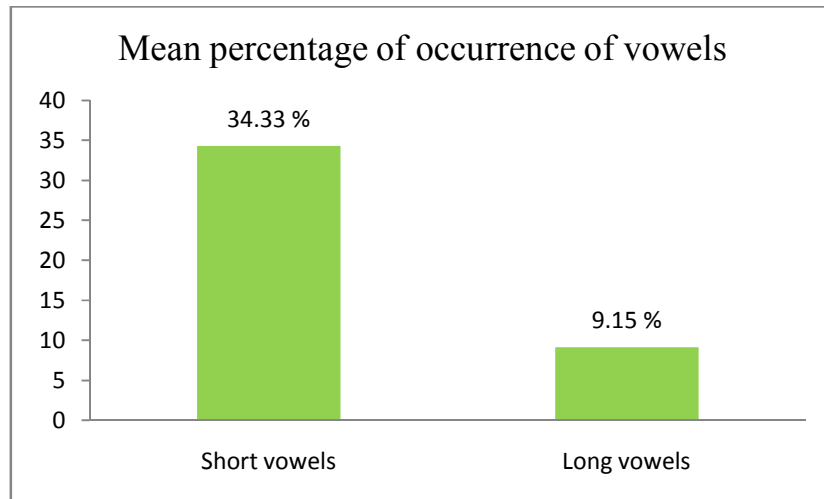


Fig 4.10. Mean percentage of occurrence of long and short vowels in Eranakulam dialect.

Among single consonants, voiceless velar /k/ showed the highest percentage of occurrence and this was followed by alveolar nasal /n/ and retroflex /ʈ/ respectively. Based on place of articulation dentals occurred predominantly followed by labials, retroflex and velars. Glottal /h/ occurred the least compared to all other places of articulation as seen in D1. Fig 4.11 depicts the mean percentage of occurrence of consonants based on places of articulation.

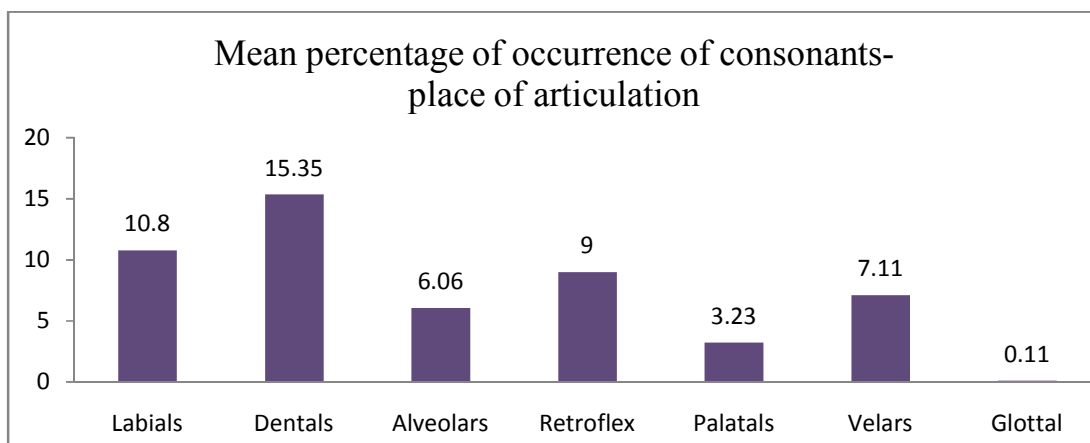


Fig 4.11. Mean percentage of occurrence of consonants based on places of articulation in Eranakulam dialect.

Based on manner of articulation, stops showed more predominant occurrence followed by nasals, glides, laterals and tap/ trills. Fricatives were the lowest occurring manner of articulation as in D1. Fig 4.12 shows the mean percentage of occurrence of consonants based on manner of articulation.

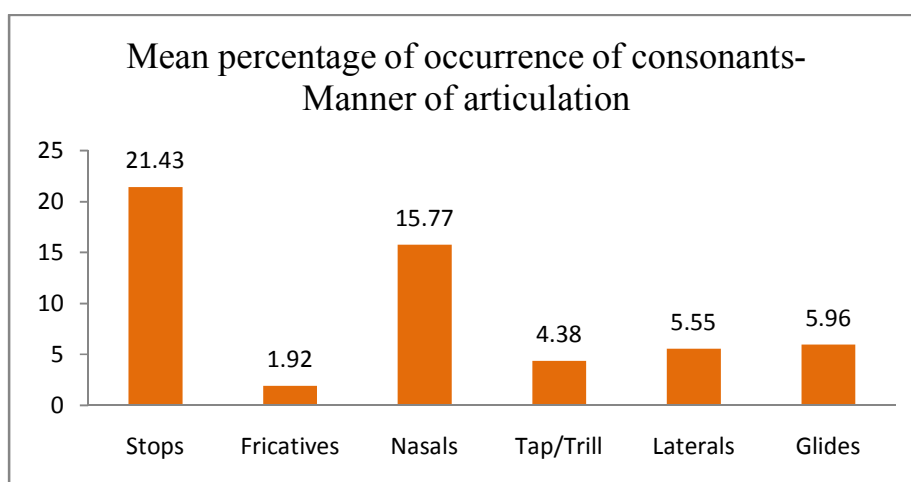


Fig 4.12. Mean percentage of occurrence of consonants based on manner of articulation in Eranakulam dialect.

Among stops, unvoiced unaspirated stops were significantly higher in occurrence compared to their voiced unaspirated counterparts. Occurrence of unvoiced velar consonant /k/ was the highest and occurred for 5.40% of the total phonemes. Occurrence of unvoiced retroflex stop /ʈ/ was 4.35% followed by unvoiced bilabial stop /p/. Unvoiced palatal affricate /c/ was comparatively less in the sample. Respective voiced counterparts of the above mentioned consonants were highly reduced in all the seven recordings of D2. Occurrence of aspirated phonemes also was significantly reduced compared to their unaspirated counterparts. Unvoiced unaspirated stops accounted for 19.37% and voiced unaspirated stops occurred for 1.08% in the conversational corpus. Voiced aspirated stops (0.18%) showed relatively higher occurrence than unvoiced aspirated stops (0.14%). Overall the results obtained in D2 were very similar to the findings in D1. Fig 4.13 indicates the mean percentage of occurrence of stops in the total corpus obtained in D2.

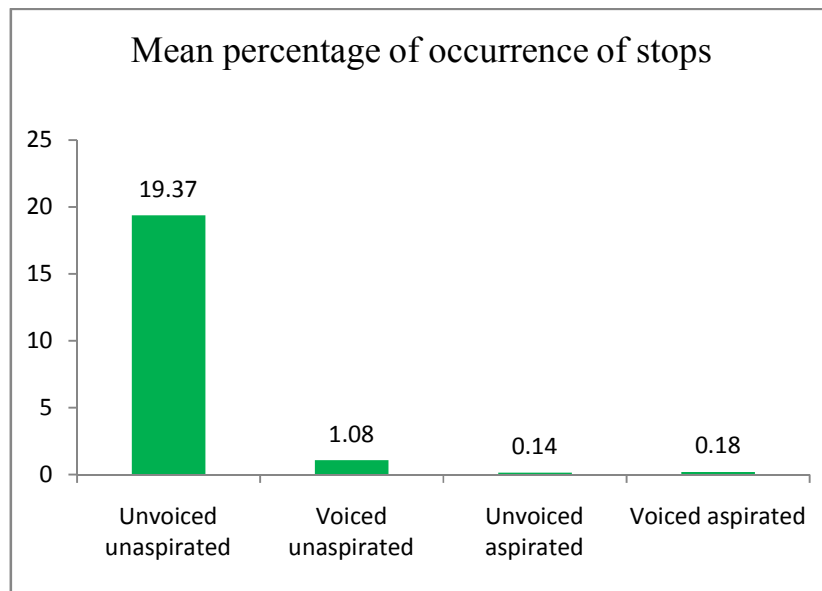


Fig 4.13. Mean percentage of occurrence of stops in Eranakulam dialect.

Among fricatives /s, ʃ, ʒ, h/, unvoiced alveolar fricative /s/ occurred for 1.11% of the data. Glottal fricative /h/ occurred for 0.11% of the total corpus and /f/ was the least among fricatives (0.07%). All nasal phonemes showed their presence in the corpus with significant percentage of occurrence. Alveolar /n/ occurred for 5.06% followed by bilabial /m/ with 3.75% of occurrence. Palatal /ɲ/ was less found than other nasal sounds (1.18%).

Among liquids, lateral retroflex /ɭ/ occurred for 1.76% and lateral alveolar /l/ occurred for 3.79%. Tap /r/ and trill /r/ had occurrences of 2.59% and 1.79% respectively. In the category of approximants, /z/ showed 0.42% of occurrence. Labial glide /v/ and palatal glide /j/ occurred 2.61% and 2.66% of the total phoneme occurrences. Overall mean frequency of occurrence of consonants based on manner of articulation was similar to the results in D1.

As per the results, the mean percentage of occurrence of the ten most prominent phonemes is shown in Fig 4.14. The most frequently occurring vowels were /a, I, ə, a: / and consonants were /k, n, t̪, p, t, l, m/. As discussed above, low-central vowel /a/ occurred more frequently followed by velar stop /k/. Velar /k/ was produced for 5.40% of the total corpus of all seven recordings. Followed by /a/ and /k/, high-front vowel /i/, mid- central vowel /ə/ and alveolar nasal /n/ showed prominent occurrence. Appendix 5 presents the occurrence of Eranakulam dialect phonemes in descending order.

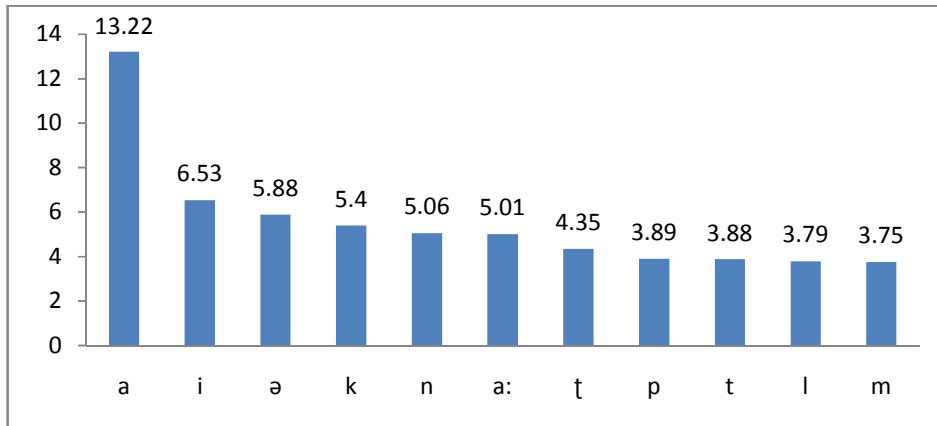


Fig 4.14. Mean percentage of frequency of occurrence of ten most frequently occurring phonemes in Calicut dialect in Ernakulam dialect.

3. Frequency of occurrence of phonemes in Thiruvananthapuram dialect of Malayalam

From the SALT analysis of the Thiruvananthapuram dialect (D3) of Malayalam, it was seen that maximum number of phonemes were recorded in the first session (R1), which amounted to 15,500 phonemes and minimum phonemes were recorded from second (R2-10,406 phonemes). Each recording session elicited more than ten thousand phonemes and the grand total was 84,563 phonemes from the 7 recordings. Fig 4.15 shows the number of phonemes obtained in each recording session

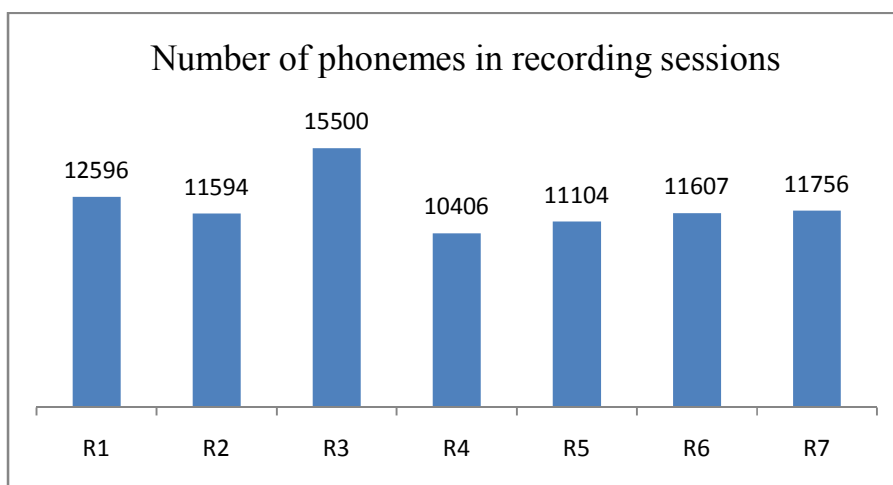


Fig 4.15. Number of phonemes obtained in each recording session in Thiruvananthapuram dialect.

The transcribed data consisted a total of 84,563 phonemes including both consonants and vowels. Like other two dialects, the mean percentage of occurrence of consonants was more than vowels. Mean occurrences of consonants were 56.33 % and vowels were 43.67 % of the total corpus as shown in Fig 4.16. Occurrence of diphthongs was 0.75 % and it was fewer in all the seven recording sessions. /ai/ occurred highly than /au/ diphthong.

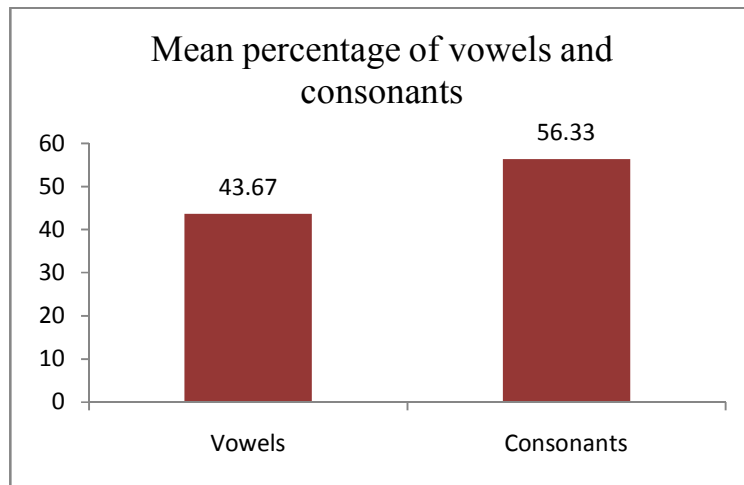


Fig 4.16. Mean percentage of occurrence of vowels and consonants in Thiruvananthapuram dialect.

The mean percentage of occurrence of each phoneme is depicted in Table 4.3. Vowel /a/ occurred most frequently in all the recordings. The mean percentage of occurrence of /a/ was more than 12 % in all seven groups. Some phonemes such as /ɖ^h/, /t^h/, /b^h/ and /f/ showed highly reduced frequency of occurrence. Aspirated phonemes /p^h/, /j^h/ and /q^h/ were not present in the entire corpus of D3.

Table 4.3

Mean percentage and standard deviation of occurrence of vowels and consonants in spoken Thiruvananthapuram dialect of Malayalam

Vowels	Mean % (SD)	Consonant	Mean % (SD)	Consonant	Mean % (SD)	Consonant	Mean % (SD)
/a/	12.96 (0.30)	/k/	5.89 (0.51)	/t̪/	4.25 (0.34)	/ŋ̪/	2.14 (0.20)
/a:/	4.48 (0.26)	^h /k/	0.002 (0.005)	/t̪ ^h /	0.01 (0.007)	/m/	3.98 (0.44)
/i/	6.57 (0.54)	/g/	0.07 (0.015)	/d̪/	0.26 (0.08)	/j/	2.80 (0.02)
/i:/	0.61 (0.08)	^h /g/	0.012 (0.01)	/d̪ ^h /	0.02 (0.01)	/r/	2.48 (0.22)
/u/	2.78 (0.37)	/c/	1.19 (0.43)	/p/	3.80 (0.38)	/l/	3.39 (0.36)
/u:/	0.80 (0.16)	^h /c/	0.01 (0.02)	/p ^h /	-	/v/	2.77 (0.14)
/e/	4.51 (0.31)	/ɟ/	0.119 (0.03)	/b/	0.46 (0.06)	/s/	0.94 (0.25)
/e:/	1.18 (0.10)	^h /ɟ/	-	/b ^h /	0.06 (0.02)	/ʂ/	0.20 (0.06)
/ai/	0.72 (0.70)	/t̪/	4.91 (0.42)	/ŋ̪/	1.53 (0.20)	/j̪/	0.26 (0.05)
/o/	1.76 (0.42)	/t̪ ^h /	0.04 (0.04)	/ɲ/	1.05 (0.15)	/h/	0.06 (0.03)
/o:/	1.16 (0.92)	/d̪/	0.08 (0.06)	/ŋ̪/	2.34 (0.30)	/l̪/	1.68 (0.23)
/au/	0.02 (0.03)	/d̪ ^h /	-	/n/	5.34 (0.21)	/z̪/	0.48 (0.07)
/ə/	6.06 (0.39)	/r/	1.20 (0.24)	/T/	0.72 (0.17)	/f/	0.06 (0.01)

As in D1 and D2, occurrence of short vowels were considerably higher than long vowels.

Open vowel /a/ (12.96%) was highly predominant than closed front vowels (12.8%) and closed

back vowels (6.5%). Fig 4.17 shows the mean percentage of occurrence of long and short vowels in D3.

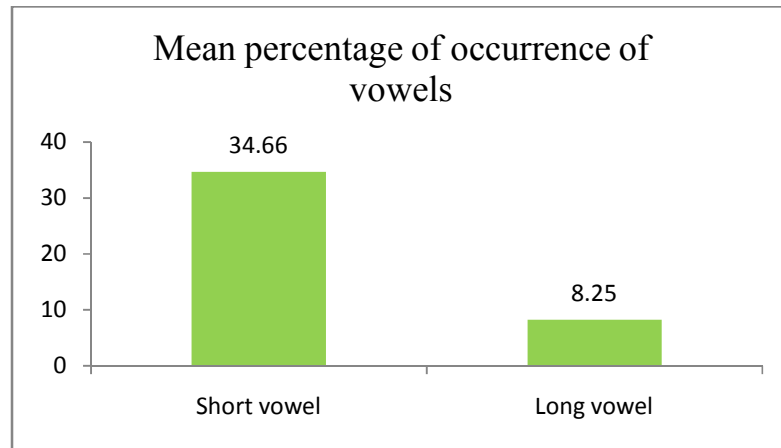


Fig 4.17. Mean percentage of occurrence of long and short vowels in Thiruvananthapuram dialect.

Considering single consonants, all the places of articulation except glottal showed significant occurrence in the conversational corpus where as voiceless velar /k/ showed highest percentage of occurrence which was followed by alveolar nasal /n/ and retroflex /ʈ/ respectively. Glottal /h/ occurred least frequently compared to all other places of articulation. Dentals showed predominant occurrence followed by labials, retroflex and velars which is similar to other two dialects. Fig 4.18 depicts the mean percentage of occurrence of consonants based on places of articulation.

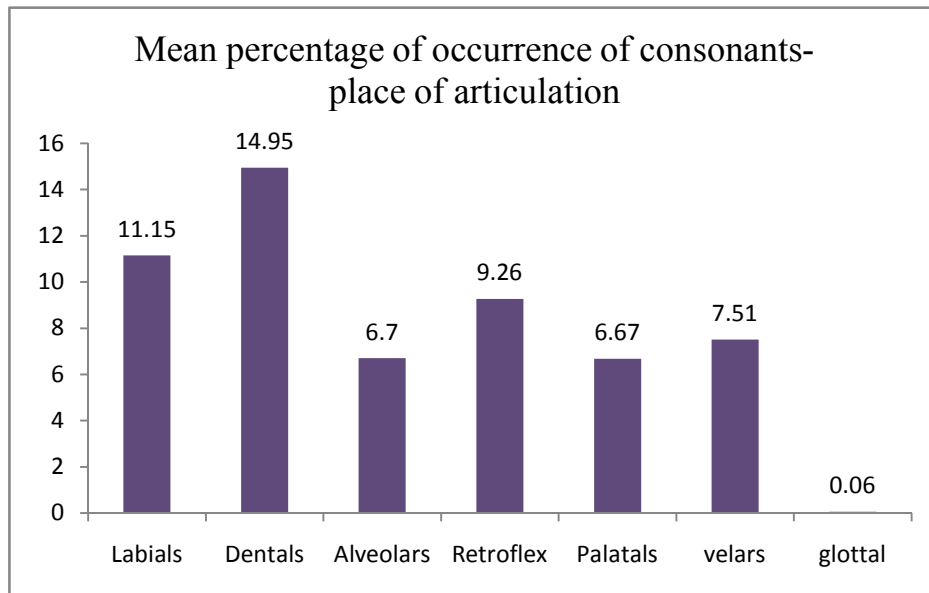


Fig 4.18. Mean percentage of occurrence of consonants based on places of articulation in Thiruvananthapuram dialect.

Based on manner of articulation, stops showed predominant occurrence followed by nasals, laterals, glides and trills. Stops and nasals occurred more than 15% of time where laterals, glides and tap/ trills occurred in the range of 4-7% of data. Fricatives were the lowest occurring manner of articulation. Fig 4.19 shows the mean percentage of occurrence of consonants based on manner of articulation.

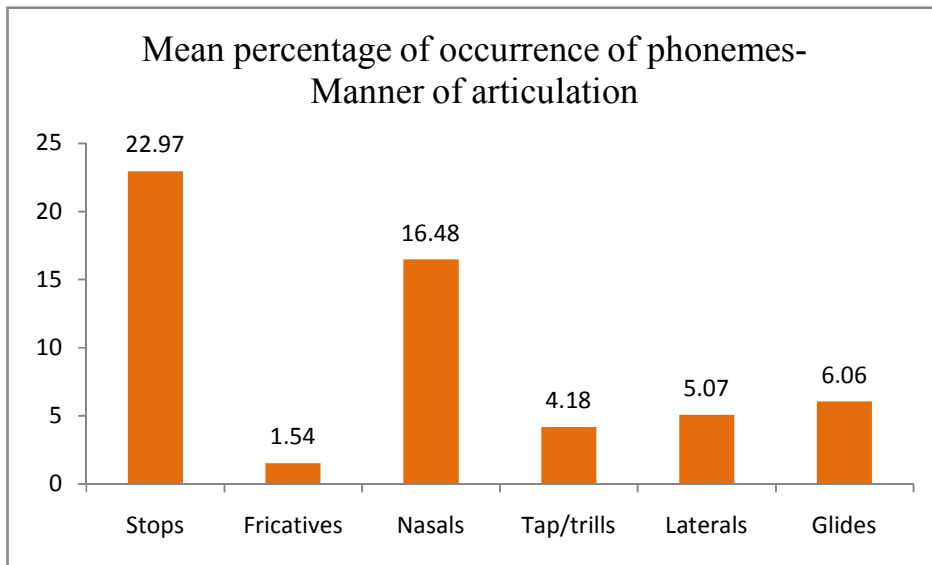


Fig 4.19. Mean percentage of occurrence of consonants based on manner of articulation in Thiruvananthapuram dialect.

As seen in the other two dialects, unvoiced unaspirated stops were significantly higher in occurrence compared to their voiced unaspirated counterparts. Occurrence of unvoiced velar consonant /k/ was highest and occurred 5.89% of the total phonemes. Occurrence of unvoiced retroflex stop /ʈ/ was 4.91% followed by unvoiced bilabial stop /p/. Unvoiced palatal affricate /c/ was comparatively less in the sample with the frequency of 1.19%. Respective voiced counterparts of the above mentioned consonants were highly reduced in all the seven recordings. Occurrence of aspirated phonemes also was significantly reduced compared to their unaspirated counterparts. Unvoiced unaspirated stops accounted for 21.06% and voiced unaspirated stops occurred for 1.01% in the conversational corpus. Voiced aspirated stops (0.1%) showed relatively higher occurrence than unvoiced aspirated stops (0.07%). Fig 4.20 indicates the mean percentage of occurrence of stops in the total corpus obtained.

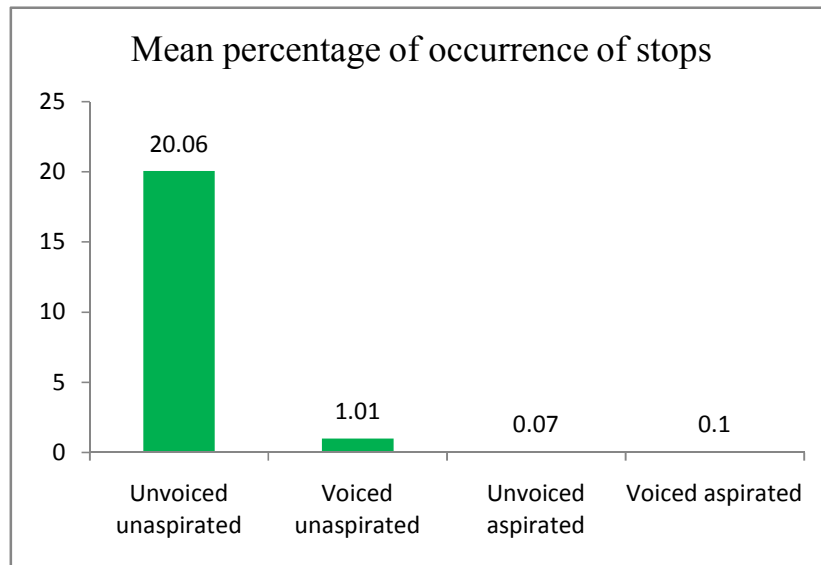


Fig 4.20. Mean percentage of occurrence of stops in Thiruvananthapuram dialect.

Unvoiced alveolar fricative /s/ occurred for 0.94% of the D3 data. Glottal fricative /h/ and labiodentals /f/ occurred for 0.06 % of the total corpus. All nasal phonemes showed their presence in the corpus with significant percentage of occurrence. Alveolar /n/ occurred for 5.35% followed by bilabial /m/ with 3.75% of occurrence. Palatal /ɲ/ was less found than other nasal sounds (1.05%).

Liquids including both tap/trill and laterals occurred in the conversational corpus. As seen in the other two dialects, occurrence of lateral retroflex /ɭ/ was less compared to lateral alveolar /l/. Tap /ɾ/ and trill /ɽ/ had occurrences of 1.20% and 2.40% respectively. In the category of approximants, /z/ showed 0.48% of occurrence. Labial glide /v/ and palatal glide /j/ occurred less than 3% of the total phoneme occurrences.

As per the results, the mean percentages of frequency of occurrence of the ten prominent phonemes in D3 are shown in Fig 4.21. The most frequently occurring vowels were /a, i, ə, e, a: / and consonants were /k, n, ʈ, t, m/. As discussed above, low-central vowel /a/ occurred more

frequently followed by /i/ (6.57%) and /ə/ (6.06%). Velar /k/ was produced 5.89% of the total corpus including all seven recordings. Appendix 6 presents the most frequently occurring phonemes in D3 in descending order.

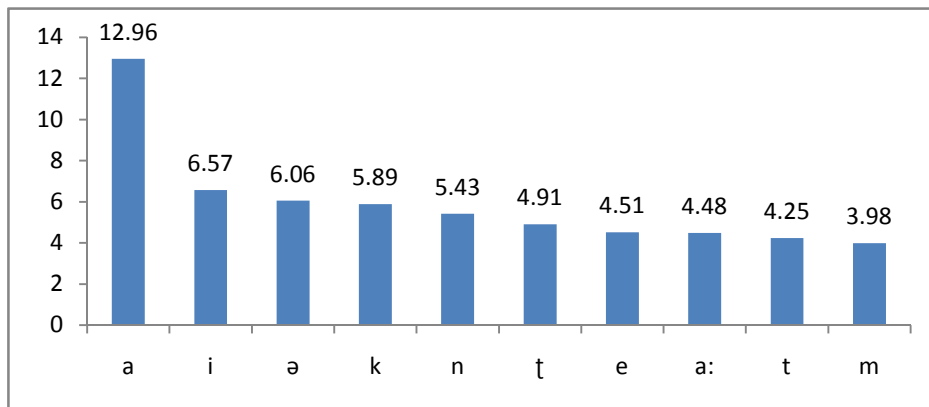


Fig 4.21. Mean percentage of frequency of occurrence of ten most frequently occurring phonemes in Thiruvananthapuram dialect.

4. Comparison of frequency occurrence of phonemes across three dialects of Malayalam

The frequency of occurrence of phonemes across the three major dialects of Malayalam was compared. A total of 2,42,268 phonemes were recorded and transcribed from twenty-one conversational recordings. More number of phonemes were recorded from Thiruvananthapuram dialect of Malayalam (84,561) followed by Calicut dialect (83,561) and the number of phonemes recorded from Ernakulam dialect were 74,144.

Based on descriptive statistical analysis the mean percentage of occurrence of consonants was more than vowels in all the three dialects. Mean occurrence of consonants were higher in Calicut dialect followed by Thiruvananthapuram and Ernakulam dialects. Occurrence

of vowels is higher in Thiruvananthapuram followed by Ernakulam and Calicut dialects. Occurrence of diphthongs was limited in all the three dialects. Fig 4.22 shows the mean percentage of occurrence of consonants and vowels in these three major dialects of Malayalam.

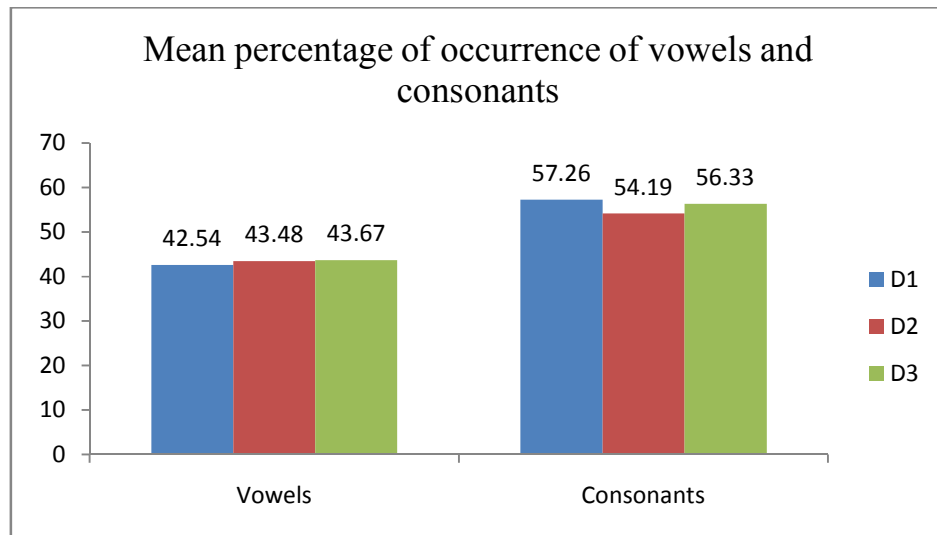


Fig 4.22. Comparison of mean percentage of occurrence of consonants and vowels across dialects

In Calicut dialect, the other most frequently occurring vowels were /a, i, ə, a:, e/ and consonants were /k, n, ʈ, t, l/. In Ernakulam dialect /a, i, ə, a: / were most frequently occurring vowels and consonants were /k, ŋ, ʈ, p, t, l, m/. In Thiruvananthapuram dialect, the frequently occurring vowels were /a, i, ə, e, a: / and consonants were /k, n, ʈ, t, m/. As seen, vowels /a, i, ə / and consonants /k, n, ʈ/ were predominant in all the three dialects. Occurrences of short vowels were considerably more than long vowels and open vowel /a/ was highly predominant than closed front vowels and back vowels in all dialects. Fig 4.23 shows the mean percentage of occurrence of long and short vowels.

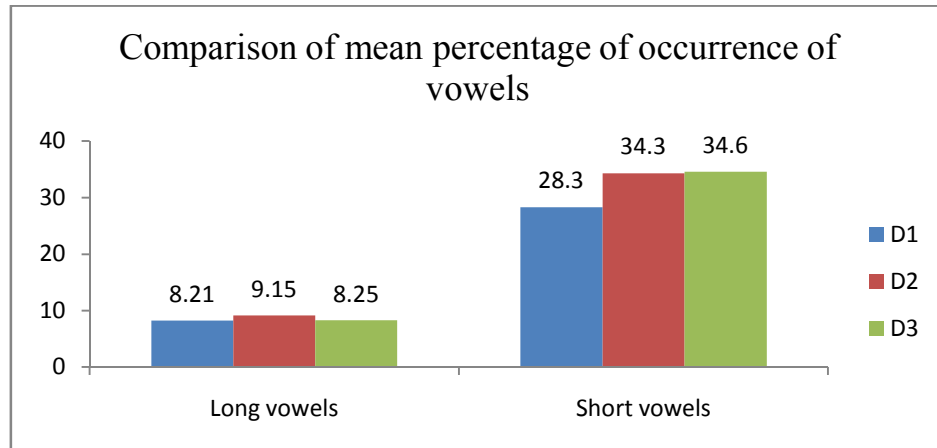


Fig 4.23. Comparison of mean percentage of occurrence of long and short vowels across dialects.

Considering the consonants, based on place of articulation, voiceless velar /k/ showed highest percentage of occurrence, followed by alveolar nasal /n/ in all the three dialects. Glottal /h/ occurred least compared to all other places of articulation. Fig 4.24 depicts the comparison of mean percentage of occurrence of consonants based on places of articulation. Dentals showed predominant occurrence followed by labials and retroflex sounds. Even though all the three dialects showed similar order of frequency of occurrence, there were small variations in mean percentage of occurrence across dialects.

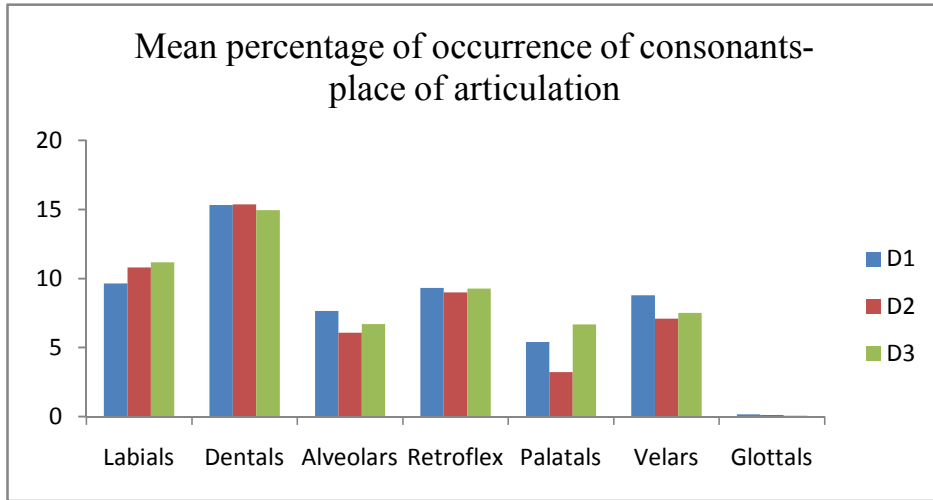


Fig 4.24. Comparison of the mean percentage of occurrence of consonants based on places of articulation across dialects.

Based on manner of articulation, stops showed the maximum occurrence followed by nasals in all the three dialects. Stops and nasals were higher in D3 compared to D1 and D2. Occurrence of stops were more and nasals were less in D1 compared to D2. Occurrence of glides, laterals and taps/trills were almost similar in occurrence in all the three dialects. Fricatives were the lowest occurring phoneme and it was the least in D3. Fig 4.25 indicates the comparison of mean percentage of occurrence of consonants based on manner of articulation in Calicut, Eranakulam and Thiruvananthapuram dialects of Malayalam.

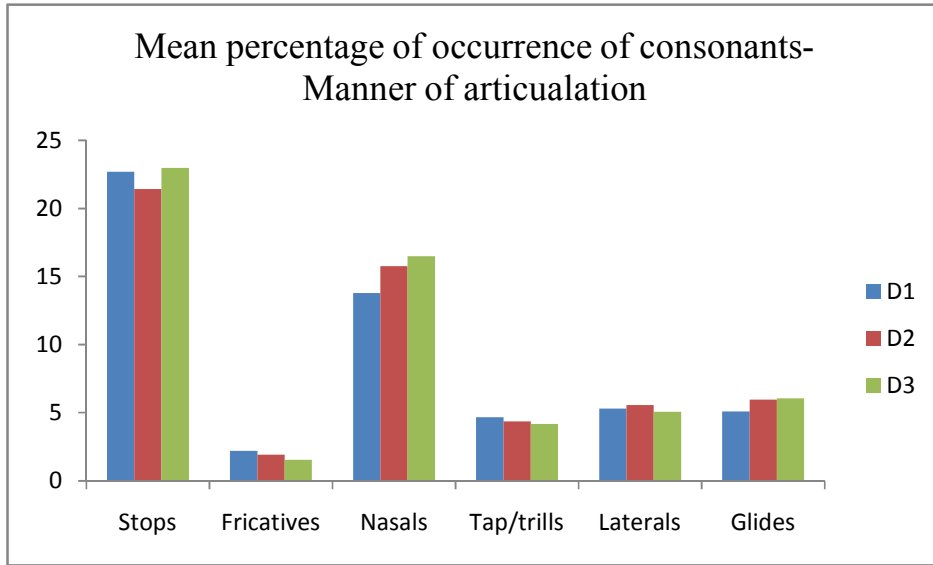


Fig 4.25. Comparison of mean percentage of occurrence of consonants based on manner of articulation.

Among stops, unvoiced aspirated stops were significantly higher in occurrence compared to their voiced aspirated counterparts. Occurrence of unvoiced velar consonant /k/ was highest and occurred for 7.06% 5.40% and 5.89% in D1, D2 and D3 respectively. Occurrence of unvoiced retroflex stop /ʈ/ was 4.44%, 4.35% and 4.91% in D1, D2 and D3 respectively, followed by unvoiced dental /ʈ/ in D1 and D3, unvoiced bilabial stop /p/ in D2. Occurrence of unvoiced palatal affricate /c/ was comparatively reduced in all three dialects. Voiced counterparts of the above mentioned consonants such as /k, ʈ, ʈ, c/ were highly reduced in all the three dialects. Occurrence of aspirated phonemes were also significantly reduced compared to their unaspirated counterparts. Unvoiced unaspirated stops accounted for 20.52% 19.37% and 20.06% in D1, D2 and D3 respectively and voiced unaspirated stops occurred for only 1.32%, 1.08% and 1.01% in the conversational corpus of D1, D2 and D3 respectively. Voiced aspirated stops showed relatively higher occurrence than unvoiced aspirated stops in all the dialects. Fig 4.26 indicates the comparison of mean percentage of occurrence of stops across dialects.

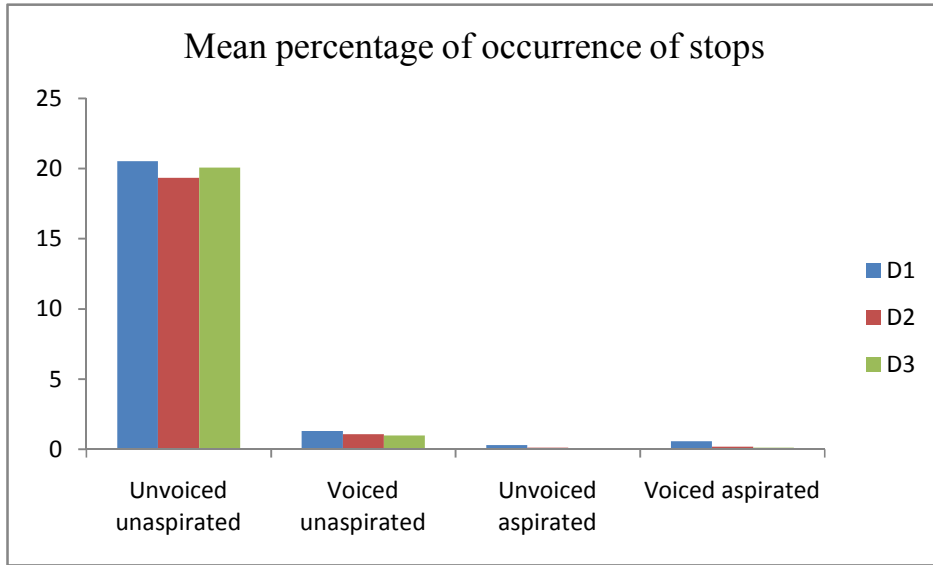


Fig 4.26. Comparison of mean percentage of occurrence of stops across dialects.

5. Overall frequency of occurrence of phonemes in Malayalam

In this section, the overall frequency of occurrence of phonemes in Malayalam is discussed considering the data from all the three dialects. A grand total of 2,42, 268 phonemes were recorded from the 21 conversational recordings including 101 participants. As seen in other section, it is very much clear that frequency of occurrence was predominant than vowels in Malayalam. Fig 4.27 shows the mean percentage of occurrence of vowels and consonants.

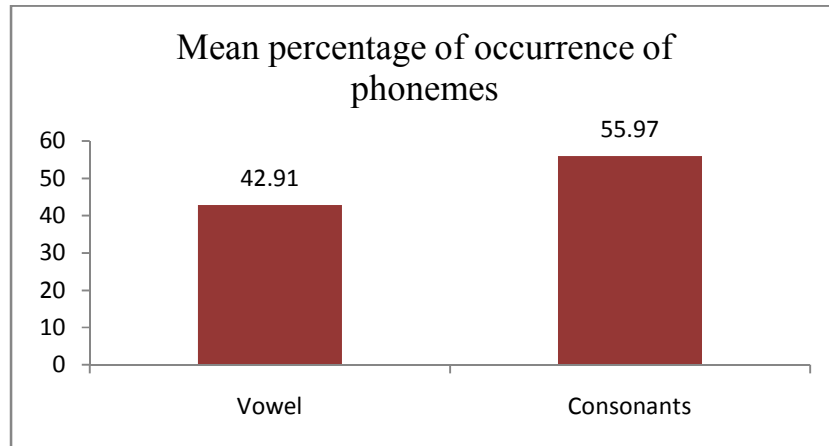


Fig 4.27. Mean percentage of occurrence of vowels and consonants in Malayalam.

The mean percentage of occurrence of each phoneme in the Malayalam corpus is depicted in Table 4.4. Vowel /a/ occurred most frequently in all the dialects. Fig 4.28 shows the mean percentage of occurrence of short and long vowels where short vowels occurred significantly high compared to long vowels.

Table 4.4

Overall mean percentage and standard deviation of occurrence of vowels and consonants in spoken Malayalam

Vowels	Mean % (SD)	Consonant	Mean % (SD)	Consonant	Mean % (SD)	Consonant	Mean % (SD)
/a/	13.34 (0.44)	/k/	6.12 (0.85)	/t̪/	4.11 (0.19)	/ŋ̪/	1.83 (0.30)
/a:/	4.60 (0.36)	/k ^h /	0.01 (0.01)	/t̪ ^h /	0.05 (0.05)	/m/	3.82 (0.13)
/i/	6.39 (0.28)	/g/	0.14 (0.07)	/d̪/	0.30 (0.06)	/j/	2.72 (0.07)
/i:/	0.64 (0.05)	/g ^h /	0.01 (0.00)	/d̪ ^h /	0.16 (0.19)	/r/	1.83 (0.15)
/u/	3.01 (0.38)	/c/	1.88 (0.30)	/p/	3.64 (0.35)	/l/	3.74 (0.32)
/u:/	0.70 (0.09)	/c ^h /	0.004 (0.006)	/p ^h /	0.01 (0.01)	/v/	2.52 (0.30)
/e/	4.22 (0.40)	/ɟ/	0.17 (0.04)	/b/	0.43 (0.05)	/s/	1.13 (0.20)
/e:/	1.08 (0.17)	/ɟ ^h /	-	/b ^h /	0.12 (0.05)	/ʃ/	0.26 (0.05)
/ai/	0.32 (0.34)	/t̪/	4.57 (0.29)	/ɽ/	1.52 (0.03)	/ʒ/	0.35 (0.08)
/o/	1.50 (0.26)	/t̪ ^h /	0.08 (0.03)	/p̪/	1.00 (0.20)	/h/	0.11 (0.04)
/o:/	1.53 (0.32)	/d̪/	0.09 (0.01)	/ŋ̪/	2.41 (0.07)	/l̪/	1.77 (0.10)
/au/	0.03 (0.00)	/d̪ ^h /	0.002 (0.004)	/n/	5.27 (0.18)	/z̪/	0.38 (0.12)
/ə/	5.84 (0.24)	/r/	2.57 (0.08)	/T/	0.64 (0.09)	/f/	0.05 (0.12)

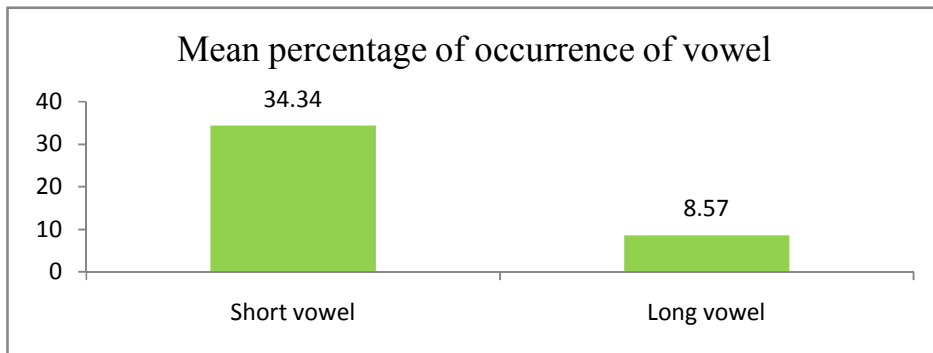


Fig 4.28. Mean percentage of occurrence of short and long vowels in Malayalam.

As seen in Fig 4.29, dentals and labials highly occurred in Malayalam conversation corpus. Mean percentage of occurrence of retroflex, velars, dentals and palatals were ranging from 5-10% of the total data. Glottal /h/ occurred for less than 1%. Fig 4.29 shows the mean percentage of occurrence of consonants based on various places of articulation.

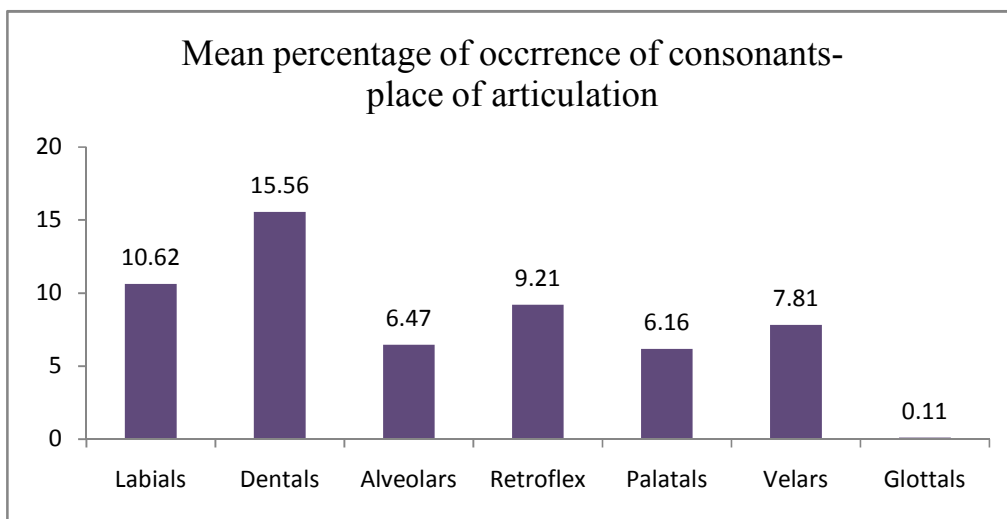


Fig 4.29. Mean percentage of occurrence of consonants based on places of articulation in Malayalam.

Based on manner of articulation, stops were the highest in Malayalam (22.59%) followed by nasals (15.87%). Fricatives were very less (1.92%) in the entire Malayalam corpus. Fig 4.30 depicts the mean percentage of occurrence of consonants based on manner of articulation.

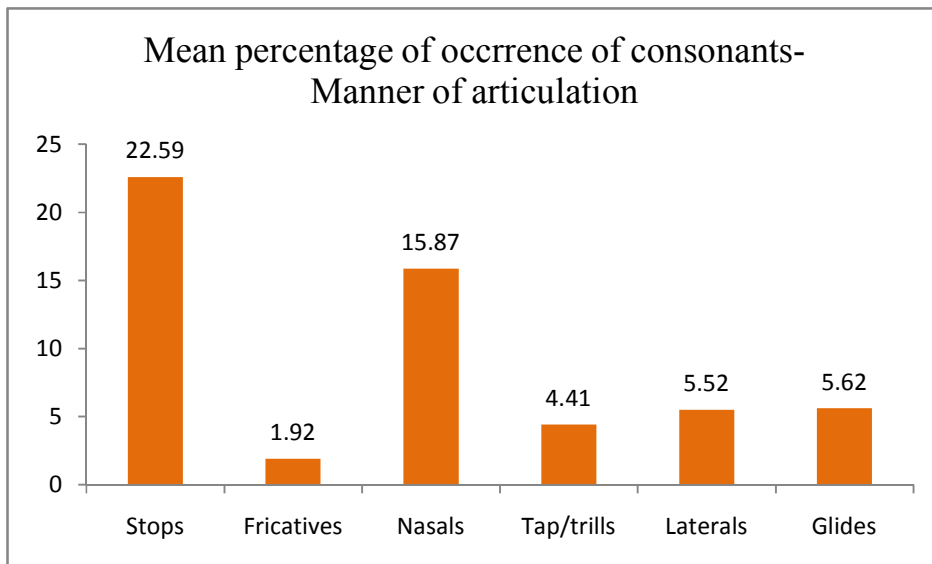


Fig 4.30. Mean percentage of occurrence of consonants based on manner of articulation in Malayalam.

Among stops, unvoiced unaspirated stops were significantly higher in occurrence compared to their voiced unaspirated counterparts. Occurrence of unvoiced velar consonant /k/ was highest and occurred 6.12% of the total phonemes. Occurrence of unvoiced retroflex stop /ʈ/ was 4.57% followed by unvoiced dental stop /t̪/, bilabial stop /p/ and palatal affricate /c/. Occurrence of voiced counterparts and aspirated phonemes were significantly reduced in overall Malayalam corpus. Fig 4.31 indicates the mean percentage of occurrence of stops in the total corpus obtained.

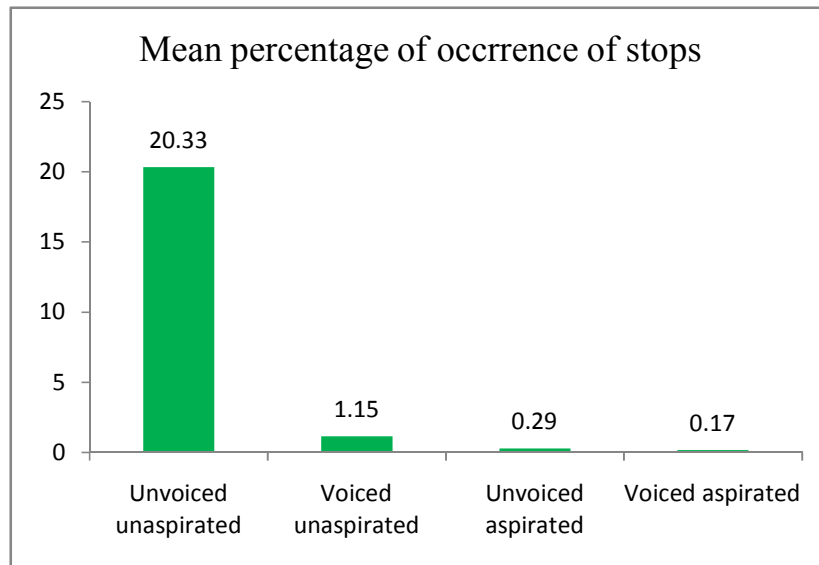


Fig 4.31. Mean percentage of occurrence of stops in the total corpus obtained in Malayalam.

Among fricatives /s, ʃ, ʂ, h/, unvoiced alveolar fricative /s/ occurred for 1.13% of the entire data and this was the most predominant fricative. All nasal phonemes showed their presence in the corpus with significant percentage of occurrence. Alveolar /n/ occurred for 5.27% followed by bilabial /m/ with 3.82% of occurrence. Palatal /ɲ/ (1%) was less found than other nasal sounds.

Liquids including both tap/trill and laterals occurred less frequently in the conversational corpus. Occurrence of lateral retroflex /ɭ/ was less compared to lateral alveolar /l/. Tap /ɾ/ and trill /ɽ/ had occurrences of 1.83% and 2.57% respectively. In the category of approximants, /z̥/ (example: word /pəz̥əm/) showed 0.38% of occurrence. Labial glide /v/ and palatal glide /j/ occurred for 2-3% of the total phoneme occurrences.

As per the results, the mean percentages of frequency of occurrence of the ten prominent phonemes in Malayalam are shown in Fig 4.32. The most frequently occurring vowels were /a, i,

ə, a:, e / and consonants were /k, n, t̪, ʈ, m/. Appendix 7 presents the most frequently occurring phonemes in Malayalam in descending order.

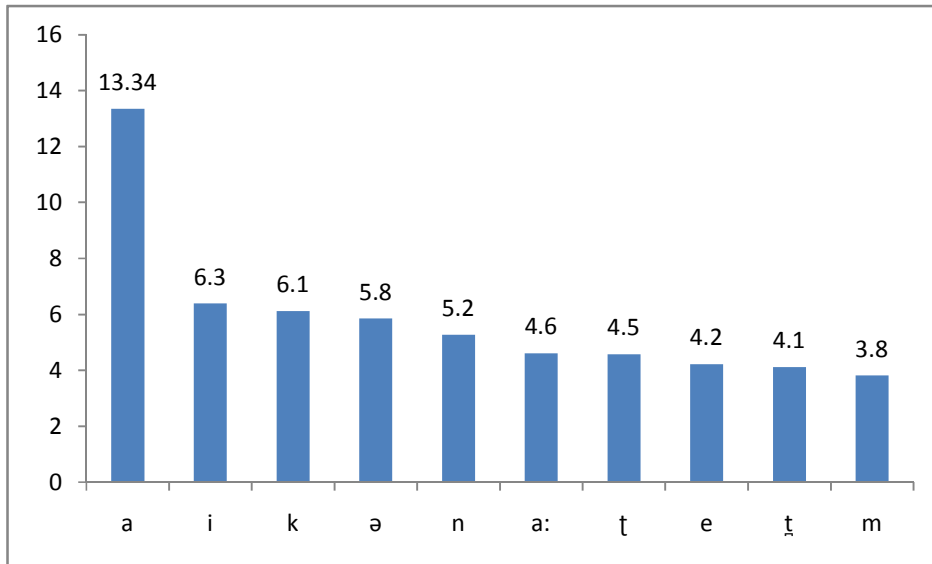


Fig 4.32. Mean percentages of frequency of occurrence of the ten most frequently occurring phonemes in Malayalam.

CHAPTER V

DISCUSSION

This study aimed to investigate the frequency of occurrence of phonemes in conversational speech samples in three major dialects of Malayalam: Calicut, Ernakulam and Thiruvananthapuram dialects. A grand total of 2,42,268 were collected from the 21 conversational recordings including 101 participants. The results revealed some salient findings.

Based on descriptive statistical analysis, the mean frequency of occurrence of consonants were predominant than vowels which is concurrent with some of the previous studies in spoken English (Denes, 1959; Delattre, 1965; Mines et al, 1978), written English (Renwick, 2011) and written Basque language (Tambovtsev, 2007). Spoken Cantonese, Mandarin, Italian, German and American English also showed similar trend (Thomas, 2005). Some of the Indian languages including written Hindi (Ghatage & Madhav, 1964), spoken and written Gujarati (Pandit, 1965), Kannada (Ranganatha, 1982; Jayaram, 1985; Sreedevi et al, 2012), Telugu (Kalyani & Sunitha, 2009; Kumar & Mohanty, 2012) and spoken Calicut dialect of Malayalam (Sreedevi & Irfana, 2013) also showed similar findings. As per these studies, consonants were predominant across languages including written and spoken mode of communication. As we know conversation means conveying information through multiple path. When all the information such as lexical, syntactical, prosody and semantic information are removed listeners could only rely on the acoustic cues of consonants than vowels (Duez, 2001).

Within the vowel category, central vowel /a/ was highly predominant and is concomitant with several earlier studies such as Ghatage and Madhav (1964), Sreedevi and Irfana (2013) in Malayalam, Ramakrishna (1962), Jayaram (1985), Manjula et al (2012) and Sreedevi et al

(2012) in Kannada, Pandit (1965) in Gujarati, Mines et al (1978), Ladefoged (2000) in English, Tamaoka and Makioka (2004) in Japanese, Sandoval et al (2008) in Castilian Spanish. But this finding is incongruent with some other studies where front vowels showed predominant occurrence followed by back and central vowels (Denes, 1959; Malecot, 1974; Mitton, 1992; Crystal, 1995; Zurinkas, 2002; Thomas, 2005; Tambovtsev, 2007; Kalyani & Sunitha, 2009; Renwick, 2011).

As seen in the present study, diphthongs (Ranganatha, 1982; Renwick, 2011) and long vowels (Ramakrishna, 1962; Sreedevi & Irfana, 2013) were highly reduced in daily conversational speech. In the present study, considering the singleton consonants, velar stop /k/ occurred with high percentage of occurrence which is concurrent with some previous studies (Tamaoka & Makioka, 2004 in Japanese, Ramakrishna et al, 1962; Sreedevi & Irfana, 2013 in Malayalam) and in disagreement with other languages where alveolar /t/ (Dewey, 1923, Tobian, 1959; Mitton, 1992,) and alveolar /n/ (Mines et al, 1978; Jayaram, 1985; Ghatage, 1994, Crystal, 2005; Zurinkas, 2002; Manjula et al, 2012, Sreedevi et al, 2012) showed highest occurrence.

In this study, among various manners of articulation, stops occurred the highest which is in concurrence with several earlier studies in English (Denes, 1959; Mines et al, 1978; Guirao & Jurado, 1990), spoken Cantonese, German and American English (Thomas, 2005) and Indian languages (Jayaram, 1985; Sreedevi et al, 2012; Kalyani & Sunitha, 2009; Kumar & Mahanty, 2012). Considering the place, present study evidenced high occurrence of dentals in the entire corpus and it is in agreement with previous studies in Kannada (Jayaram, 1985; Sreedevi et al, 2012). However, other studies have reported different places of articulation; alveolars showed higher occurrence in Telugu (Kalyani & Sunitha, 2009; Kumar & Mahanty, 2012), Cantonese, Mandarin, Italian, German and American English (Thomas, 2005). Labials showed predominant

occurrence in French (Malecot, 1974). Some of the interesting phonemes in Malayalam such as /z, t/ occurred less than 1% which is agreement with one of the previous study in Malayalam (Ghatage, 1994).

The present study observed significantly high occurrence of unvoiced stops compared to their voiced counterparts which is in congruent with earlier studies in Malayalam (Ramakrishna et al, 1962) and in spoken Cantonese, Mandarin (Thomas, 2005). The aspirated sounds were negligibly present, which is in agreement with all the other Indian language studies. This is possibly because the aspiration feature is rarely used in colloquial conversations.

CHAPTER VI

SUMMARY AND CONCLUSIONS

The present study planned to obtain the frequency of occurrence of phonemes in conversational speech samples in three major regional dialects of Malayalam: Calicut (D1), Eranakulam (D2) and Thiruvananthapuram (D3). The frequency of occurrence of phoneme data is important to understand the language structure and also has wide applications in the area of audiology and speech language pathology especially in the assessment and rehabilitation of persons with communication disorders.

In the present study, participants included were fluent adult native speakers of Malayalam in the age range of 30 to 55 years with a minimum of 10 to 12 years of schooling with Malayalam medium of instruction. Data consisted of seven recordings each from the region of the three dialects. This amounted to a total of 21 conversational recordings. Each recording session had 4 to 5 participants engaged in conversation. Data consisted a total of 101 participants out of which 47 were males and 54 were females including the three different dialects. The participants did not have any clinical background of any known diseases or disorders. All participants were ensured to have normal speech, language, hearing abilities.

The data was collected through conversations in controlled natural environments for about 25 to 30 minutes of duration. The digital recorder (Olympus WS 100) was kept at equidistance from all the speakers. Participants themselves initiated the conversation topics without any specific topic being assigned to them to avoid the high occurrence of certain phonemes. All conversation samples including three major regional dialects were transcribed using International Phonetic Alphabet for Regional languages (Malayalam) by Asher and

Kumari (1997). The repetitive words and exclamatory remarks were excluded from analysis. Loan words from English that are commonly used (e.g: bus, car, bat, ticket etc) in day to day conversation were included in the analysis.

The raw data obtained was analyzed using freely downloadable SALT (Systematic Analysis of Language Transcripts) software for calculating the frequency count. The results obtained from the SALT software were subjected to statistical analysis. A total of 2,42,268 phoneme counts were obtained from the 21 recorded conversation samples including seven conversation recording each from Calicut (83,561), Eranakulam (74,144) and Thiruvananthapuram (84,561) dialects of Malayalam. Mean and standard deviation of the frequency of phonemes of all the 21 samples were obtained. Vowels consisted of 42.91% of the total data whereas consonants accounted for 55.97% of the total phoneme corpus. The results showed that the most frequently occurring vowels were /a, i, ə, a:, e / and consonants were /k, n, t, ʈ, m/. Vowel /a/ was the most frequently occurring phoneme among vowels and stop velar /k/ among consonants in all the three dialects. Aspirated affricatives /tʃ^h/ (Example: word /tʃ^ha:nsi/) was absent in the entire data. Diphthongs and aspirated consonants also showed reduced occurrence in the Malayalam conversational corpus.

Occurrence of short vowels was significantly higher than long vowels. Short open vowel /a/ was highly predominant than short closed front vowels and closed back vowels. Considering single consonants, voiceless velar /k/ showed highest percentage of occurrence, followed by alveolar nasal /n/ and retroflex /ʈ/ respectively. According to consonant classes based on place of articulation, dentals showed predominant occurrence followed by labials, retroflex and velars. Glottal /h/ showed fewer occurrences in all the three considered dialects. Based on manner of

articulation, stops showed predominant occurrence followed by nasals, laterals, glides and trills. Fricatives were the lowest occurring manner of articulation.

Among stops, unvoiced unaspirated stops were significantly higher in occurrence compared to their voiced unaspirated counterparts. All nasal phonemes showed their presence in the corpus with significant percentage of occurrence. Among fricatives /s, ʃ, ʒ, h/, unvoiced alveolar fricative /s/ was the highest in occurrence in all the three dialects. Liquids including both tap/trills and laterals occurred in the conversational corpus with considerable percentage of occurrence. Tap /ɾ/ and trill /ɽ/ had similar percentage of occurrences. Among the approximants, /z/ showed less than 1% of occurrence. Labial /v/ and palatal /j/ occurred for less than 3% of the total phoneme occurrences.

In Calicut dialect, the other most frequently occurring phonemes were /a, k, i, ə, ŋ, t, a:, e, t, l /. In Eranakulam dialect /a, i, ə, k, ŋ, a:, t, p, t, l, m / were most frequently occurring phonemes. In Thiruvananthapuram dialect, the frequently occurring phonemes were /a, i, ə, k, ŋ, t, e, a:, t, m /. Even though all the three major dialects of Malayalam showed similar order of frequency of occurrence, there were small variations in mean percentage of occurrence across dialects. Overall, the order of most frequently occurring phonemes were /a, i, k, ə, ŋ, a:, t, e, t, m/.

The present study findings will assist audiologists and speech language pathologists for developing and updating the current test materials for assessing different communication disorders. The information obtained is useful for selecting treatment goals for this population. Machines that speak and recognize speech has been a dream for communication engineers. The realization of these machines, be it a phonetic type writer or a computer which responds to

spoken words, depends upon the availability of a great deal of information about the acoustical characteristics of sounds in each language spoken by its community. Hence for the same reason the knowledge about the frequency of phoneme provides a database for phoneme identification process of an automatic speech recognition system. These statistics are also applicable to the fields of linguistics and speech, to teach the language as a foreign language and to clinical services in speech pathology. Furthermore, the possibility of translation from one language into another by computer, points out the need for the study of languages. These data may also be of use in automatic generation of speech for reading machine for the blind. India is a country with large number of languages. And different languages have different phoneme systems. A database of individual languages with its phoneme frequency will help us to know a language better. Also it has application as a measure of conversation efficacy in speech to text and text to speech systems or softwares.

REFERENCES

- Asher, R. E., & Kumari, T. C. (1997). *Malayalam*. London: Routledge.
- Bhagwat, S. V. (1961). *Phonemic frequencies in Marathi and their relation to devising a speed script*. Pune: Deccan College.
- Campbell, R. A. (1965). Discrimination test word difficulty, *Journal of Speech and Hearing Research*, 8, 13-22.
- Chomsky, N. (1957). *Syntactic structures*. Mouton de Gruyter: Berlin.
- Christian, D., Wolfram, W., & Nube, N. (1988). *Variation and change in geographically isolated communities: Appalachian English and Ozark English*. Tuscaloosa: University of Alabama Press.
- Crystal, D. (1995). *The Cambridge Encyclopedia of the English Language*. Cambridge University Press.
- Delattre, P. (1965). *Comparing the phonetic features of English, French, German and Spanish: An interim report*. London, UK: Julius Groos Verlag Heidelberg.
- Denes, P.B. (1959). On the statistics of spoken English, *Journal of the acoustics society of America*, 35 (6), 892-904.
- Dewey, G. (1923). *Relative frequency of English speech sounds*. Cambridge, MA: Harvard University Press.
- Duez, D. (2001). Restoration of deleted and assimilated consonant sequences in conversational French speech: effects of preceding and following context. *Journal of the International Phonetic Association*, 31(1), 101-114.
- Egan, J. P. (1948). Articulation testing methods, *Laryngoscope*, 58, 955-991.
- Gelfand, S.A. (2009). *Essentials of audiology (3rd ed)*. Thieme Medical Publishers, Inc, New York.

- Ghatage, A. M. & Madhav. A. (1964). *Phonemic and Morphemic frequencies in Hindi*. Poona: Deccan College Postgraduate and Research Institute.
- Ghatage, A. M. (1994). *Phonemic and morphemic frequencies in Malayalam*. Mysore: Central Institute of Indian Languages.
- Guirao, M., & Jurado, M. A. (1990). Frequency of occurrence of phonemes in American Spanish. *Revue Quebecoise de Linguistique*, 19, 135-149.
- Ishtiaq, M. (2012). Language shifts among the scheduled tribes in India: A geographical study. Delhi: Motilal Banarsidass Publishers. 26-27.
<https://www.cia.gov/library/publications/the-world-factbook/geos/in.html#People>
- Jayaram, M. (1985). Sound and Syllable distribution in written Kannada and their application to Speech and Hearing. *Journal of All India Institute of Speech and Hearing*, 16 & 17, 19-30.
- Jiang, H. (2010). *Malayalam: a Grammatical Sketch and a Text*. Rice University. Online version: <http://www.owlnet.rice.edu>
- Kalyani, N., & Sunitha, K.V.N. (2009). Syllable analysis to build a dictation system in Telugu Language. *International Journal of Computer Science and Information Security*, 6, 193-205.
- Kelkar, A. R. (1994). *Phonemic and Morphophonemic frequency in Oriya*. Mysore: CIIL.
- Kumar, R. S. B., & Mohanty, P. (2012). Speech recognition performance of adults: a proposal for a battery for Telugu. *Theory and practice in language studies*, 2, 193-204.
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the World's languages*. Blackwell Publishers Ltd, Oxford:UK.
- Manjula, P., Geetha, C., Sharath Kumar, K.S., & Antony, J. (2012). *Development of PB word list in Kannada*. ARF Project. AIISH, Mysore.

- Mines, M. A., Hanson, B. F., & Shoup, J. E. (1978). Frequency of occurrence of phonemes in conversational English. *Language and Speech*, 21, 221-41.
- Mitton, R. (1992). *Computer-usable dictionary file based on the Oxford Advanced Learner's Dictionary of current English*. <http://www.oucs.ox.ac.uk/internal/annrep/annrep85-86.html> (accessed: 16 March 2012).
- Palai, E. B., & O'Hanlon, L. (2004). Word and phoneme frequency of occurrence in conversational Setswana: A clinical linguist application. *Southern African Linguistics and Applied Language Studies*, 22, 125-142.
- Pandit, P.B. (1965). *Phonemic and morphemic frequencies of the Gujarati language*. Deccan College Postgraduate and Research Institute.
- Parker, F., & Riley, K. (2005). *Linguistics for non-linguistics: A primer with exercises (4th ed)*. Boston: Allyn and Bacon.
- Ramakrishna, B. S. (1962). *Some aspects of the relative efficiencies of Indian languages: A study from information theory point of view*. Bangalore: Indian Institute of Science.
- Ranganatha, M. R. (1982). *Morphophonemic analysis of the Kannada language*. Mysore: CIIL.
- Renwick, M. E. L. (2011). Phoneme type frequency in Romanian. *In proceedings of the 34th Annual Penn working papers in Linguistics*, 17, 194-204.
- Sandoval, A. M., Toledano, D. T., de la Torre, R., Garrote, M., & Guirao, J. M. (2008). Developing a phonemic and syllabic frequency inventory for spontaneous spoken Castilian Spanish and their comparison to text-based inventories. *Proceedings of the Language Resources and Evaluation Conference*. Available at: www.lrec-conf.org/proceedings/lrec2008/pdf/283_paper.pdf (accessed: 14 March 2012).
- Sapir, E. (1921). *Language: An introduction to the study of speech*. New York: Harcourt and Brace.
- Siegel, P. (2010). *Second language acquisition (1st ed)*. Cambridge University Press. US.

Sreedevi, N., & Irfana, M. (2013). Frequency of occurrence of phonemes in Malayalam. *In Proceedings of 41st All India Dravidian Linguists Conference*, 27, June, Chennai.

Sreedevi, N., Smitha, N., & Vikas, M.D. (2012). *Frequency of phonemes in Kannada*. Completed ARF Project. AIISH, Mysore.

Systematic Analysis of Language Transcripts (SALT)
<http://www.saltsoftware.com/salt/downloads/servicepacks.cfm>

Tamaoka, K., & Makioka, S. (2004). Frequency of occurrence for units of phonemes, morae, and syllables appearing in a lexical corpus of a Japanese newspaper. *Behavioral Research in methods of Instrument computer*, 36 (3), 531-47.

Tambovtsev, Y. (2007). How can typological distances between Latin and some Indo-European language taxa improve its classification?. *The Prague Bulletin of Mathematical Linguistics*, 88, 73–90.

Thomas, W. C. (2005). *The effects of occurrence frequency of phonemes on second language acquisition: A quantitative comparison of Cantonese, Mandarin, Italian, German and American English* (Doctoral dissertation), Chinese university of Hong Kong.

Tobians, J. V. (1959). Relative occurrence of phonemes in American speech, *Journal of acoustic society of America*, 31.

Yadav, R. (1976). Generative phonology and the aspirated consonants of colloquial Maithili, *Contributions to Nepalese Studies*, 4 (1), 77-91.

Yegerlehner, J., & Voegelin, F. M. (1957). Frequencies and Inventories of Phonemes from nine languages. *International journal of American Linguistics*, 23, 85-93.

Zurinkas, T. E. (2002). Analysis of English in USA accent developing a Truespel Leengwu print. Retrieved from <http://lincs.ed.gov/nifl-assessment/2002/0093.html>.

=====

APPENDIX 1

Comparison of frequency of occurrence of phonemes in the present study (Malayalam) with Kannada (Sreedevi, et al. 2013)

Malayalam				Kannada			
Phoneme	%	Phoneme	%	Phoneme	%	Phoneme	%
/a/	13.34	/ɲ/	1.00	/a/	14.57	/j/	1.24
/i/	6.39	/u/	0.70	/n/	7.59	/p/	1.22
/k/	6.12	/i/	0.65	/i/	6.70	/o:/	1.20
/ə/	5.84	/T/	0.64	/a:/	5.66	/i:/	1.09
/n/	5.27	/b/	0.43	/r/	5.53	/u:/	0.55
/a:/	4.60	/z/	0.38	/d/	5.35	/ɖ/	0.53
/t/	4.57	/ʃ/	0.35	/e/	5.27	/ʒ/	0.42
/e/	4.22	/ai/	0.32	/l/	4.98	/ə/	0.35
/t̪/	4.11	/ɖ/	0.30	/t/	4.54	/tʃ/	0.34
/m/	3.82	/ʒ/	0.26	/k/	3.49	/au/	0.33
/l/	3.74	/ɻ/	0.17	/g/	3.30	/ŋ/	0.29
/p/	3.64	/ɖʰ/	0.16	/m/	2.76	/ai/	0.28
/u/	3.01	/g/	0.14	/v/	2.57	/ɖʰ/	0.20
/j/	2.72	/bʰ/	0.12	/d/	2.24	/bʰ/	0.09
/r/	2.57	/h/	0.11	/e:/	2.23	/kʰ/	0.05
/v/	2.52	/ɖ/	0.09	/b/	2.08	/tʰ/	0.05
/ŋ/	2.41	/tʰ/	0.08	/o/	1.85	/ŋ/	-
/c/	1.88	/t̪ʰ/	0.06	/h/	1.75	/pʰ/	-
/r/	1.83	/f/	0.05	/s/	1.75	/ɖʰ/,	-
/ɲ/	1.82	/au/	0.03	/t/	1.53	/ɖʒʰ/	-
/l/	1.77	/pʰ/	0.013	/ʃ/	1.33	/tʃʰ/	-
/o:/	1.53	/kʰ/	0.011				
/ŋ/	1.52	/gʰ/	0.01				
/o/	1.50	/cʰ/	0.004				
/s/	1.13	/ɖʰ/	0.002				
/e/	1.08	/ɻʰ/	-				

APPENDIX 2

Loan English words included in the analysis

Apple	Internet
Ball	Jeep
Bank	Key
Bat	Mobile
Bike	News
Book	Orange
Bucket	Paper
Bus	Paper
Car	Pass
Company	Pen
Cricket	Pencil
Cup	Phone
Degree	Pineapple
Doctor	Plus two
Door	Room
Dress	Seat
Engineer	Shirt
Foot ball	Shoe
Free	Stool
Fridge	Table
Gas	Time
Gate	Traffic
Hotel	TV

APPENDIX 3

CONSONANT SEGMENTS

International Phonetic Alphabet for Regional languages (Malayalam)

Asher & Kumari (1997)

Manner of Articulation	Place of Articulation						
	Labial	Dental	Alveolar	Retroflex	Palatal	Velar	Glott
Stops							
Voiceless	p	t̪	T	ʈ	c	K	
Voiceless	p ^h	t̪ ^h		ʈ ^h	c ^h	k ^h	
Voiced	B	d̪		ɖ	ɟ	G	
Voiceless	b ^h	d̪ ^h		ɖ ^h	ɟ ^h	g ^h	
Fricatives	f		s, ʃ	ʂ			H
Nasal	m	n̪	N	ɳ	ɲ	ŋ	
Liquid							
Tap /Trill			r, ɽ				
Lateral			l	ɭ			
Approximants					ɻ		
Glide	v				j		

APPENDIX 4

Mean percentage of occurrence of phonemes in Calicut dialect of Malayalam in descending order

Vowels	Mean %	Consonants	Mean %	Consonants	Mean %	Consonants	Mean %
/a/	13.83	/k/	7.06	/c/	1.59	/j/	0.19
/i/	6.07	/n/	5.32	/ɲ/	1.53	/b ^h /	0.17
/ə/	5.58	/t/	4.44	/ɳ/	1.48	/h/	0.15
/a:/	4.31	/ʈ/	4.20	/s/	1.35	/t ^h /	0.12
/e/	3.75	/l/	4.04	/p/	0.78	/t ^h /	0.10
/u/	3.46	/m/	3.73	/ɽ/	0.53	/d/	0.09
/o:/	1.67	/p/	3.24	/b/	0.46	/f/	0.035
/o/	1.23	/j/	2.68	/ʃ/	0.44	/p ^h /	0.024
/e:/	0.88	/r/	2.65	/d ^h /	0.38	/k ^h /	0.02
/i:/	0.71	/ɳ/	2.49	/d/	0.37	/g ^h /	0.01
/u:/	0.64	/v/	2.19	/ʂ/	0.30	/d ^h /	0.007
/ai/	0.15	/r/	2.01	/z/	0.24	/c ^h /	0.002
/au/	0.027	/ʎ/	1.89	/g/	0.21	/j ^h /	-

APPENDIX 5

Mean percentage of occurrence of phonemes in Eranakulam dialect of Malayalam in descending order

Vowels	Mean %	Consonants	Mean %	Consonants	Mean %	Consonants	Mean %
/a/	13.83	/k/	5.40	/r/	1.79	/d/	0.118
/i/	6.53	/n/	5.06	/l/	1.76	/b ^h /	0.117
/ə/	5.88	/t/	4.35	/ŋ/	1.56	/h/	0.11
/a:/	5.01	/p/	3.89	/ɲ/	1.18	/t ^h /	0.10
/e/	4.40	/ʈ/	3.88	/s/	1.11	/d ^h /	0.076
/u/	2.81	/l/	3.79	/T/	0.65	/f/	0.074
/o:/	1.76	/m/	3.75	/z/	0.42	/t ^h /	0.038
/o/	1.52	/j/	2.66	/b/	0.36	/p ^h /	0.014
/e:/	1.18	/v/	2.61	/ʃ/	0.36	/k ^h /	0.01
/u:/	0.64	/r/	2.59	/ʒ/	0.27	/g ^h /	0.007
/i:/	0.60	/ŋ/	2.41	/d/	0.27	/d ^h /	-
/ai/	0.09	/c/	1.85	/j/	0.20	/c ^h /	-
/au/	0.04	/ɳ/	1.81	/g/	0.14	/j ^h /	-

APPENDIX 6

Mean percentage of occurrence of phonemes in Thiruvananthapuram dialect of Malayalam in descending order

Vowels	Mean %	Consonants	Mean %	Consonants	Mean %	Consonants	Mean %
/a/	12.96	/k/	6.12	/ŋ̄/	1.83	/g/	0.14
/i/	6.57	/n/	5.27	/ʎ/	1.77	/b ^h /	0.12
/ə/	6.06	/t/	4.57	/ŋ̄/	1.52	/h/	0.11
/e/	4.51	/ṭ/	4.11	/s/	1.13	/ḍ/	0.09
/a:/	4.48	/m/	3.82	/ɲ/	1.00	/t ^h /	0.08
/u/	2.78	/l/	3.74	/T/	0.64	/ṭ ^h /	0.058
/o/	1.76	/p/	3.64	/b/	0.43	/f/	0.055
/e:/	1.18	/j/	2.72	/z/	0.38	/p ^h /	0.014
/o:/	1.16	/r/	2.57	/ʃ/	0.35	/k ^h /	0.011
/u:/	0.80	/v/	2.52	/ḍ/	0.30	/g ^h /	0.01
/ai/	0.72	/ŋ̄/	2.41	/ʂ/	0.26	/c ^h /	0.004
/i:/	0.61	/c/	1.88	/ɟ/	0.17	/ḍ ^h /	0.002
/au/	0.02	/r/	1.83	/ḍ /	0.16	/j ^h /	-

APPENDIX 7

Overall mean percentage of occurrence of phonemes in Malayalam in descending order (D1+D2+D3)

Vowels	Mean %	Consonants	Mean %	Consonants	Mean %	Consonants	Mean %
/a/	13.34	/k/	7.06	/c/	1.59	/j/	0.19
/i/	6.39	/n/	5.32	/ŋ/	1.53	/b ^h /	0.17
/ə/	5.84	/t/	4.44	/ŋ/	1.48	/h/	0.15
/a:/	4.60	/t̪/	4.20	/s/	1.35	/t̪ ^h /	0.12
/e/	4.22	/l/	4.04	/p/	0.78	/t̪ ^h /	0.10
/u/	3.01	/m/	3.73	/T/	0.53	/d/	0.09
/o:/	1.53	/p/	3.24	/b/	0.46	/f/	0.035
/o/	1.50	/j/	2.68	/j/	0.44	/p ^h /	0.024
/e:/	1.08	/r/	2.65	/d ^h /	0.38	/k ^h /	0.02
/u:/	0.70	/ɽ /	2.49	/d̪/	0.37	/g ^h /	0.01
/u:/	0.64	/v/	2.19	/ʃ/	0.30	/d̪ ^h /	0.007
/ai/	0.32	/r/	2.01	/z/	0.24	/c ^h /	0.002
/au/	0.03	/l̪/	1.89	/g/	0.21	/j ^h /	-