FREQUENCY OF OCCURRENCE OF PHONEMES IN CALICUT AND ERANAKULAM DIALECTS OF MALAYALAM

[1]Sreedevi, N., & [2]Irfana, M.

Abstract

Malayalam is a Dravidian language primarily spoken in the southwest of India. The knowledge of the frequency of occurrence of phonemes in a language is essential to understand the language structure. This information is useful in the field of audiology and speech language pathology for preparation of assessment tools and treatment of communication disordered population. The present study aimed to compare the frequency of occurrence of various phonemes in Calicut and Eranakulam dialects of Malayalam using conversation samples. Participants were fluent native speakers of Calicut or Eranakulam dialects of Malayalam in the age range of 30 to 55 years. Fourteen conversation samples were recorded with seven groups from each dialect separately for 25-30 minutes. A minimum of 4 to 5 participants were present for each recording. The obtained samples were transcribed using both broad and narrow IPA transcription. Inter judge and intra judge reliability of phonetic transcription was calculated for 10% of the recorded samples. Recorded data was analyzed using the SALT software for obtaining frequency of phonemes. Mean and standard deviation of frequency of phonemes of all the fourteen samples were obtained. The analysis showed that /a/ was the most frequently occurring vowel and /k/ was the most frequently occurring consonant in both dialects. In Calicut dialect, the other most frequently occurring vowels were /I, ə, a:, e/ and consonants were /n, ţ, t, l/. In Eranakulam dialect /a, I, ə, a:/ were most frequently occurring vowels and consonants were /k, n, ţ, p, t, l, m/. The results obtained will help audiologists and speech language pathologists in developing and updating the existing test material for evaluating various communication disorders and also for selection of treatment targets in such population. The study also has implications in linguistics analysis, speech synthesis experiments and development of automatic speech recognition systems.

Key words: *Dravidian languages, Malayalam, SALT software, Place of articulation, Manner of articulation*

Introduction

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 variations of a single language, they differ in terms of vocabulary, grammar, pragmatics and pronunciation (Siegel, 2010). The frequency of phonemes can vary from language to language and dialect to dialect.

There are about 600 consonants used in different languages around the world and some occur more frequently than others. The most common are the voiceless stops and about 98 % of the world's languages have the three voiceless stops /p/, /t/, /k/. As in 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). Ladefoged (2000) reported /a/ to be the most common vowel followed by /e/, /i/, /o/, /u/ in English.

Since 1930s, research has been ongoing on frequency of occurrence of phonemes. As the phonological organization changes across language use and dialects, several studies were carried out in different languages. The studies of phonemes were mostly on the written corpus since 1970s, where the studies mainly concentrated on the frequencies based on sources like newspapers, journals and scripts of plays. Mines, Hanson and Shoup (1978) studied in English based on spoken corpus of 1,03,887 phoneme. Vowel /a/ was the most frequently occurring phoneme followed by / n, t, i, s, r, l, d, ɛ /. The three nasals /m/, /n/ and /ŋ/ accounted for 18.45% of all consonants produced in initial, medial and final positions. Liquids /l/ and /r/ had 6% occurrence in adult American English speech. Stops showed 29.21% of occurrence. Delattre (as cited in Edwards, 2003) included 2000 syllables for his study in English from narration and dramatization and found that the most frequent vowels were /ə/, /I/, /æ/ and the consonants were /t/, /n/, /r/ and /l/.

Within the same language, there are observable differences in frequency of occurrence of

[1]Sreedevi, N., Reader in Speech Sciences, All India Institute of Speech and Hearing (AIISH), Mysore-06, E-mail: srij_01@yahoo.com, & [2]Irfana, M., JRF, AIISH, Mysore-06, E-mail: fanairfana@gmail.com.

phonemes between written and spoken data. Sandoval, Toledano, de la Torre, Garrote and Guirao (2008) investigated the syllabic and phonemic frequency in spoken and written contexts in Castilian Spanish and reported that /s/ occurred for 8% in spoken and 7% in written contexts. Similarly an Indian study by Bhagwat (1961) calculated the phonemic and morphemic frequencies in Marathi. In Hindi, Ghatage and Madhav (1964) considered the phonemic and morphemic frequencies based on written source of materials and found that vowels were predominant than consonants in Hindi. Ghatage (1994) studied Malayalam phoneme and morpheme frequencies using 1,00,000 words from various written materials. The results showed that vowels /a/ and /I/ were the most frequent phonemes in their corpus. Among consonants, palatal nasal /ŋ/ was most frequent followed by /k/ and /m/ respectively. Based on reports of Kelkar (1994), in Oriya /ə/ was the most frequent vowel followed by /a/ and /I/. Among consonants, /r/, /k/ and /t/ were the most frequently occurring phonemes.

In Kannada, the study by Ramakrishna (1962) was based on the written corpus. The results showed that long vowels and aspirated phonemes occurred less frequently. Vowel /a/ was the predominant vowel and consonants such as /r/, dentals /d/ and /t/ are the highly used consonants in conversational Kannada. Jayaram (1985) also studied phonemes in Kannada from news paper sources. Results revealed that /a, I, u, e/ were the frequently occurring vowels and /n, r, d, t, l/ were the frequently occurring consonants. Recently, Sreedevi, Smitha and Vikas (2012) found that in conversational or spoken Kannada, the most frequently occurring phoneme was vowel /a/ followed by /n/, /I/. /e/, /r/, /a:/, /d/, /l/, /u/, /g/ and /k/. Phonemes /h/, /s/, /p/, /ʧ/, /ʤ/, /ʃ/ occurred less frequently and vowels constituted 44.3% and consonants, 55.3% of the conversational data in Kannada.

In Telugu, Kalyani and Sunitha (2009) found 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 followed by bilabials and velars. Kumar and Mohanty (2012) also found similar results and reported that aspirated stops were less in occurrence in Telugu spoken mode of communication.

There is a wide application for the statistics of phoneme occurrence in a language. These phoneme frequency data is of use in automatic generation of speech for reading machines for the blind. Frequency of phonemes in a language provides a database for phoneme identification process in an automatic speech recognition system.

The frequency of occurrence of phonemes plays important role in the development of linguistic theories in a number of areas including the semantic relations, phonotactics 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 require most 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 goals for various communication disorders; knowledge about the most frequent phonemes help in targeting those phonemes in speech therapy for the hearing impaired which results in better intelligibility of their speech.

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. A database of different dialects with its phoneme frequency will help us to know a language better. The present study investigates the frequency of occurrence of phonemes in Calicut and Eranakulam dialects of Malayalam.

The earlier study on phoneme frequency in Malayalam by Ghatage and Madhav (1994) was based on various written materials. Their findings possibly might have undergone changes as there are many new words, modified and loan words used in daily conversation with time. Sreedevi and Irfana (inpress) studied conversational sample of Calicut dialect of Malayalam and results showed /a/ as the most frequently occurring vowel and /k/ as the most frequently occurring consonant. Frequency of occurrence of phonemes in each dialect varies based on its phonological system. Therefore the present study is aimed to compare the frequency of occurrence of phonemes in conversational speech samples of Calicut dialect (D1) and Eranakulam dialect (D2) of Malayalam.

Method

Participants: Fluent adult native speakers of Ernakulam and Calicut dialects of Malayalam were identified from individual homes and offices and were randomly selected for the study. The participants were in the age range of 30 to 55

years with a minimum of 10 to 12 years of schooling in Malayalam medium of instruction. The participants were devoid of any diseases or disorders.

Instrumentation: For recording the conversation samples a digital recorder (Olympus WS 100) was used.

Procedure: The data was collected through conversations in controlled natural environments for about 25 to 30 minutes of duration. There were fourteen recordings with seven conversational groups from each dialect. A minimum of 4 to 5 participated in each recording. The digital recorder was kept at equidistance from all the speakers. Participants initiated the conversation without any specific topic being assigned to them. They were instructed to avoid words from other languages and to speak naturally only in Malayalam. They were not restricted from using commonly used loan English words (E.g.: News paper, table, school etc). Each of the seven recording sessions in both dialects involved different participants.

Data Analysis: The conversation samples were transcribed using International Phonetic Alphabet for Regional languages (Malayalam) by Asher & Kumari (1997) which is provided in Appendix 1. The transcribed data was analyzed using the software Systematic Analysis of Language Transcripts (SALT version 9) for the frequency count. A database of Malayalam phonemes was prepared and saved in the editable standard wordlists of SALT software. The SALT software compares the database and provides the phoneme count based on the loaded phoneme file.

Inter judge and intra judge reliability: 10% sample of each conversation recording was subjected to inter judge and intra judge reliability. Three graduate speech language pathologists served as judges for inter judge reliability measures. For intra judge reliability, the 10% of each conversational recording were transcribed and re-analyzed by one of the authors. Cronbach alpha test was applied and a reliability index (α) of 0.86 was obtained for inter judge and 0.91 was obtained for intra judge reliability.

Results

The aim of the present study was to compare the frequency of occurrence of phonemes in conversational speech in Eranakulam and Calicut dialects of Malayalam. Following the SALT analysis, it was found that more number of phonemes were recorded from Calicut dialect of Malayalam (83, 561) and the number of phonemes recorded from Eranakulam dialect were 74,144. Each recording session elicited more than ten thousand phonemes except two conversational groups (G2 &G3) from the dialect of Eranakulam and the grand total was 1,57,705 phonemes from the fourteen recordings.

The transcribed data consisted a total of 1, 57,705 phonemes including both consonants and vowels. Based on the descriptive statistical analysis the mean percentage of occurrence of consonants was more than vowels in both dialects. Mean occurrence of consonants was more in Calicut dialect and vowels were higher in Ernakulam dialect. Figure 1 shows the mean percentage of occurrence of consonants and vowels in both dialects. Occurrences of diphthongs were scanty in both dialects.

- *Malayalam is a Dravidian language primarily spoken in the southwest of India. According to Summer (2009), it is the official language of Kerala state and Lakshadweep union territory. Within India alone there are over 35 million speakers of Malayalam, not including the other nearly 500,000 speakers outside India. Malayalam 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.*
- *Calicut is a northern district of Kerala where more of Muslim community is lodged. The dialect of Malayalam spoken by mappila Muslims shows deep influence of Arabic language. And the dialect is partially different from mainstream Malayalam.*
- *Eranakulam is a middle district of Kerala where more of middle class and upper middle class community is accommodated, which has more influence of English.*
- *SALT, the Systematic Analysis of Language Transcripts, is a freely downloadable computer program to analyze and interpret the language samples from one or more speakers during a conversation. It is providing options to analyze lexical, syntactic, semantic, pragmatic, rate, fluency, and error categories in different languages. It also gives information about the frequency of words, morphemes, phonemes, grammatical categories, etc.*

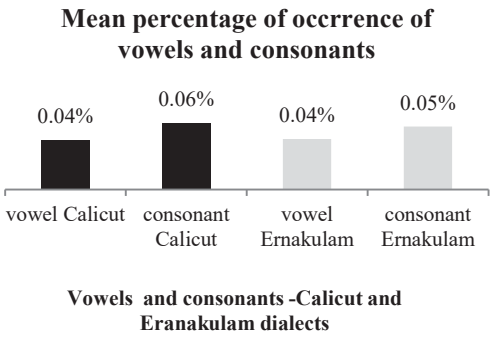

Figure 1: Mean percentage of occurrence of consonants and vowels.

The mean percentage of occurrence of each phoneme is depicted in Table 1. Overall, the mean percentage of occurrences of phonemes was similar in Calicut and Eranakulam dialect. Vowel /a/ occurred most frequently in all the recordings and it occurred for more than 10% in each of the fourteen conversational recordings. Some phonemes such as /ɖʰ, ɟh ,ch/ showed highly reduced frequency of occurrence.

Table 1: Mean percentage and standard deviation of occurrence of vowels and consonants in Calicut and Ernakulam dialects of Malayalam.

Phoneme	Mean % (SD)-D1	Mean % (SD)-D2	Phoneme	Mean % (SD)-D1	Mean % (SD)-D2	Phoneme	Mean % (SD)-D1	Mean % (SD)-D2
/a/	13.83 (1.3)	13.22 (0.59)	/c/	1.59 (0.35)	1.85 (0.39)	/ɲ/	0.78 (0.31)	1.18 (0.22)
/a:/	4.31 (1.74)	5.01 (0.53)	ʰ /c/	0.002 -	-	/ɳ/	2.49 (0.2)	2.41 (0.49)
/i/	6.07 (0.24)	6.53 (0.42)	/ɟ/	0.19 (0.11)	0.20 (0.12)	/n/	5.32 (1.07)	5.06 (0.70)
/i:/	0.71 (0.17)	0.60 (0.18)	ʰ /ɟ/	-	-	/n̪/	1.53 (0.7)	1.81 (0.74)
/u/	3.46 (0.6)	2.81 (0.25)	/ʈ/	4.44 (0.5)	4.35 (1.79)	/m/	3.73 (0.5)	3.75 (0.22)
/u:/	0.64 (0.2)	0.64 (0.8)	/ʈʰ/	0.10 (0.10)	0.10 (0.05)	/j/	2.68 (0.31)	2.66 (0.23)
/e/	3.75 (0.45)	4.40 (0.42)	/ɖ/	0.09 (0.06)	0.11 (0.07)	/ɾ/	2.65 (0.34)	2.59 (0.33)
/e:/	0.88 (0.42)	1.18 (0.16)	/ɖʰ/	0.01 -	-	/l/	4.04 (0.39)	3.79 (0.32)
/ai/	0.15 (0.07)	0.09 (0.03)	/t̪/	4.20 (0.44)	3.88 (0.44)	/v/	2.19 (0.14)	2.61 (0.32)
/o/	1.23 (0.49)	1.52 (0.21)	/t̪ʰ/	0.12 (0.06)	0.03 (0.03)	/s/	1.35 (0.21)	1.11 (0.13)
/o:/	1.67 (0.39)	1.76 (0.26)	/d̪/	0.37 (0.20)	0.27 (0.12)	/ʂ/	0.30 (0.06)	0.27 (0.05)
/au/	0.03 (0.02)	0.04 (0.03)	/d̪ʰ/	0.38 (0.18)	0.07 (0.04)	/ʃ/	0.44 (0.14)	0.36 (0.08)
/ə/	5.58 (1.24)	5.88 (0.63)	/p/	3.24 (0.39)	3.89 (0.40)	/h/	0.15 (0.05)	0.11 (0.04)
/k/	7.06 (0.73)	5.40 (2.18)	/pʰ/	0.024 (0.01)	0.01 (0.03)	/ɭ/	1.89 (0.25)	1.76 (0.26)
ʰ /k/	0.02 (0.02)	0.01 (0.01)	/b/	0.46 (0.10)	0.36 (0.18)	/ɻ/	0.24 (0.12)	0.42 (0.07)
/g/	0.21 (0.08)	0.14 (0.06)	/bʰ/	0.17 (0.08)	0.11 (0.04)	/f/	0.035 (0.06)	0.07 (0.04)
ʰ /g/	0.01 (0.01)	0.01 (0.01)	/ɳ/	1.48 (0.32)	1.56 (0.38)	/r/	2.01 (0.20)	1.79 (0.12)
						/T/	0.53 (0.13)	0.65 (0.11)

*D1-Calicut dialect and D2-Eranakulam dialect

In Calicut dialect, following /a/, the other most frequently occurring vowels were /I, ə, a:, e/ and consonants were /n̪, ʈ, t, l/. In Eranakulam dialect /a, I, ə, a: / were most frequently occurring vowels and consonants were /k, n̪, ʈ, p, t, l, m/. 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 both dialects. Figure 2 shows the mean percentage of occurrence of long and short vowels.

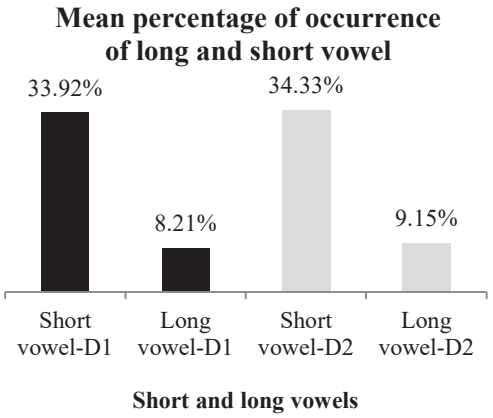

Figure 2: Mean percentage of occurrence of long and short vowels.

Considering the consonants, voiceless velar stop /k/ showed highest percentage of occurrence and this was followed by alveolar nasal /n/ in both dialects. Glottal fricative /h/ occurred the least compared to all other consonants. Figure 3 depicts the mean percentage of occurrence of consonants based on places of articulation.

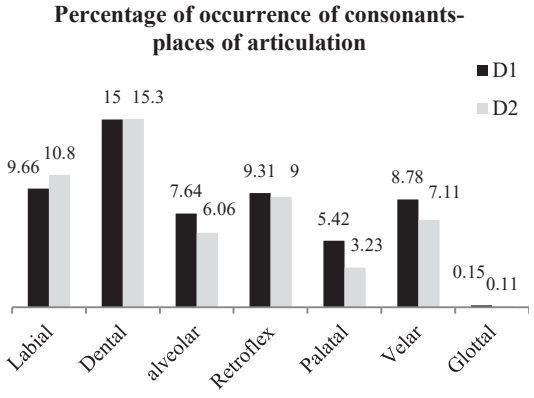

Figure 3: Mean percentage of occurrence of consonants based on place of articulation in both dialects.

Based on manner of articulation, stops showed predominant occurrence followed by nasals, glides, laterals and taps/trills in both groups. Fricatives were the lowest occurring manner of articulation. Occurrences of stops were more and nasals were less in D1compared to D2. Figure 4 shows the mean percentage of occurrence of consonants based on manner of articulation.

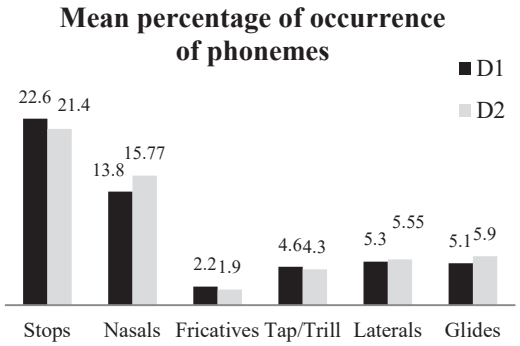

Figure 4: Mean percentage of occurrence of consonants based on manner of articulation in both dialects.

Among stops, unvoiced stops were significantly higher in occurrence compared to their voiced counterparts. Occurrence of unvoiced velar consonant /k/ was highest and constituted 7.06% and 5.40% of the total phonemes in D1 and D2 respectively. Occurrence of unvoiced retroflex stop /ʈ/ 4.44% was followed by unvoiced dental stop /t̪/ in D1. But in D2, the occurrence of unvoiced retroflex stop /ʈ/ was 4.35% followed by unvoiced bilabial stop /p/. Unvoiced palatal affricate /ʧ/ was comparatively reduced in both dialects. Respective voiced counterparts of the above mentioned consonants were highly reduced in all the fourteen recordings. Occurrences of aspirated phonemes were also significantly reduced compared to their unaspirated counterparts. Unvoiced unaspirated stops accounted for 20.52% and 19.37% in D1 and D2 respectively and voiced unaspirated stops occurred for only 1.32% and 1.08% in the conversational corpus of D1and D2. Voiced aspirated stops showed relatively higher occurrence than unvoiced aspirated stops in both groups. Figure 5 indicates the mean percentage of occurrence of stops in both dialects.

Among fricatives /s, ʃ, ʂ, h/, unvoiced alveolar fricative /s/ occurred predominantly for 1.35% and 1.11% in D1 and D2 data respectively. Glottal fricative /h/ was the least occurring among fricatives (0.15%) in D1 and /f/ was the least in D2 corpus. All six nasal phonemes of Malayalam were present in the corpus with significant percentage of occurrence. Alveolar /n/ occurred significantly (5.32% and 5.06%) followed by bilabial /m/ with 3.75% and 3.75% of occurrence in both dialects. Palatal /ɲ/ occurred less

frequently than other nasal sounds in both Calicut and Eranakulam dialects.

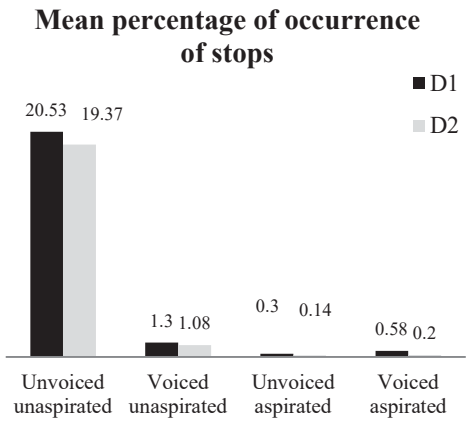

Figure 5: Mean percentage of occurrence of stops

Liquids including both tap/trills and laterals occurred in the conversational corpus with considerable percentage of occurrence. In the Calicut dialect, lateral retroflex /ɭ/ and lateral alveolar /l/ occurred 1.89% and 4.04% respectively. Tap /ɾ/ and trill /ɾ/ had similar percentage of occurrences which was 2.01% and 2.65%. In the Eranakulam dialect lateral retroflex /ɭ/ occurred 1.76% and lateral alveolar /l/ occurred 3.79%. Tap /ɾ/ and trill /ɾ/ had occurrences of 2.59% and 1.79% respectively. In the category of approximants, /ɻ/ showed 0.24% and 0.42% of occurrence in both dialects. Labial glide /v/ and palatal glide /j/ showed less than 2% of the total phoneme occurrences in the entire corpus of the two dialects.

In the present study considering the 10 most frequently occurring phonemes, Calicut dialect had 5 vowels and 5 consonants where as Eranakulam dialect had 4 vowels and 6 consonants. The order of most frequently occurring vowels were /a, I, ə, aː, e/ and consonants were /k, n, ʈ, t, l/ in Calicut dialect. In Eranakulam dialect /a, I, ə, aː / were the most frequently occurring vowels and consonants were /k, n, ʈ, p, t, l/.

Discussion

This research aimed to obtain the frequency of occurrence of phonemes in Calicut and Eranakulam dialects of Malayalam. Results showed some salient findings. Based on descriptive statistical analysis, the mean frequency of occurrence of consonants were predominant than vowels in both dialects. Some of the studies in Kannada (Jayaram, 1985; Sreedevi et al, 2012) and Telugu (Kalyani & Sunitha, 2009) also showed similar findings. Within the vowel category, /a/ was highly predominant and is concomitant with several earlier studies (Ghatage & Madhav, 1964; in Malayalam; Ramakrishna, 1962; Jayaram, 1985 and Sreedevi et al, 2012 in Kannada, Mines et al, 1978; Sandoval et al, 2008 and Ladefoged, 2000 in English).

In the present study, among various manner of articulations, stops occurred the highest which is in concurrence with several previous studies in English (Mines et al, 1978) and Indian languages (Jayaram, 1985; Sreedevi et al, 2012; Kalyani & Sunitha, 2009; Kumar & Mahanty, 2012). Comparing the present study with the earlier studies, nasal dental /n/ was higher in most of the previous studies (Mines et al, 1978 -English; Ghatage & Madhav, 1994 -Malayalam; Jayaram, 1985 & Sreedevi et al, 2012- Kannada). Also the present study observed significantly high occurrence of unvoiced stops compared to their voiced counterparts. There are no earlier reports on such high disparity between unvoiced and voiced stops. Ghatage and Madhav have not commented on the voiced/unvoiced distinctions in Malayalam. The aspirated sounds were minimally present, which is in agreement with all the other studies in Indian languages. This is possibly because; the aspiration feature is rarely used in colloquial conversations though it is phonemic.

Conclusions

The frequency of occurrence of different phonemes in Calicut and Eranakulam dialects of Malayalam were determined using conversational samples collected from fourteen different recording sessions and the results showed that consonants constituted larger part of the total corpus than vowels. In Eranakulam dialect, /a, I, ə, a: / were most frequently occurring vowels and consonants were /k, n̪, ʈ, p, t, l/ and in Calicut dialect, the most frequently occurring vowels were /I, ə, aː, e/ and consonants were /n̪, ʈ, t, l/. Aspirated sounds and long vowels were rarely present in the data. The findings of the present study are applicable in diagnostic and therapeutic aspects of communication disordered population and also for use in automatic generation of speech reading machines for the blind.

References

Asher, R. E., & Kumari, T. C. (1997). Malayalam (pp. 405-435). 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 and word test difficulty. *Journal of Speech and Hearing Research, 8*, 13–22.

Edwards, H. T. (2003). *Applied Phonetics: The sounds of American English (3rd edition).* Canada: Delmar Learning.

Egan, J. P. (1948). Articulation testing methods. *Laryngoscope, 58*, 955–991.

Ghatage, & 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.

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*, 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,* Vol. 6.

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*, vol. 2, pp: 193-204.

Ladefoged, P. A. (2000). Vowels and Consonants (2nd ed.).Oxford, UK: Blackwell.

Mines, M. A., Hanson, B. F., & Shoup, J. E. (1978). Frequency of occurrence of phonemes in conversational English. *Language and Speech, 21*, 221-41.

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.

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 Catalian 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).

Siegel, P. (2010). Second language acquisition (1st ed.). Cambridge University Press. US.

Sreedevi, N., Smitha, N & Vikas, M, D. (2012). Frequency of occurrence of phonemes in Kannada. ARF project. AIISH.

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

Appendix I
CONSONANT SEGMENTS

International Phonetic Alphabet for Regional languages (Malayalam)
Asher & Kumari (1997)

Manner of Articulation	Place of Articulation						
	Labial	Dental	Alveolar	Retroflex	Palatal	Velar	Glottal
Stops							
Voiceless	p	t̪	T	ʈ	c	k	
Voiceless Aspirated	pʰ	t̪ʰ		ʈʰ	cʰ	kʰ	
Voiced	b	d̪		ɖ	ɟ	g	
Voiceless Aspirated	bʰ	d̪ʰ		ɖʰ	ɟʰ	gʰ	
Fricatives	f		s, ʃ	ʂ			h
Nasal	m	n̪	n	ɳ	ɲ	ŋ	
Liquid							
Tap /Trill			r, ɾ				
Lateral			l	ɭ			
Approximants					ɻ		
Glide	v				j		