FREQUENCY OF OCCURRENCE OF PHONEMES IN KANNADA

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

(2011-2012)

Sanction Number: SH/CDN/ARF/4.22/2011-12

Total Grants: ₹ 3,20,000

Total Duration of the Project: 04.1.2012 – 03.1.2013

Principal Investigator

Dr. N. Sreedevi

Reader in Speech Sciences

Research Officer

Vikas M D

Department of Speech Language Sciences

All India Institute of Speech and Hearing

Manasagangothri, Mysore - 570006

ACKNOWLEDGEMENTS

Our sincere gratitude to our Director, Prof. S R Savithri, All India Institute of Speech and Hearing, Mysore, for sanctioning the project and for all her support, advice and encouragement throughout the project.

It is our pleasure to extend our gratitude to Prof. Y V Geetha, Head, Department of Speech Language Sciences for lending us the departmental facilities to carry out the project work. We thank Dr. M S Vasanthalakshmi, Lecturer in Biostatistics, AIISH, Mysore, for helping us with statistics and Ms. Irfana, M for helping to incorporate the corrections.

We express our deep sense of gratitude to all the participants and their family for their co-operation during the study.

Dr. N. Sreedevi
Principal Investigator

Mr Vikas M D
Research Officer

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

INTRODUCTION

Language is the expression of thought by means of speech sounds. It is a system of communication by sound that is through the organs of speech and hearing using vocal symbols possessing arbitrary conventional meanings (Pei & Gaynor, 1954). Language is a systematic verbal symbolism which makes use of verbal elements such as sounds, words and phrases which are arranged in certain ways to make sentences.

Phonology is the study of speech sounds of a given language and their function within the sound system of that language. According to Bloomfield (1914), phonology is the organization of sounds into patterns and phonemes are the minimal unit of distinctive sound feature. In Webster's third new international (Webster & Gove, 1961), the phoneme is defined as " the smallest unit of speech distinguishing one unit from another, in all the variations it displays in the speech of one person or in one dialect as a result of modifying influences such as neighboring sounds or stress''. The term phoneme was first used in late 1870s notably by Kruszewski. Most contemporary linguists view phoneme as the minimal bundle of relevant sound features. Trager and Smith (1951) proposed 45 phonemes for English such as /p/, /b/, /t/, /d/, /k/, /g/ etc. The phonemes are distinctive in all the languages but use of phonemes differs.

Phonetic transcription is the use of phonetic symbols to represent speech sounds. Ideally, each sound in a spoken utterance is represented by a written phonetic symbol, so as to furnish a record sufficient to render possible the accurate reconstruction of the utterance. The transcription system will in general reflect the phonetic analysis imposed by the

transcriber on the material. For most phoneticians, the symbol set of choice is the alphabet of the International Phonetic Association, known as the International Phonetic Alphabet (IPA).

The spoken data or utterances are transcribed in to written form by this IPA for several applications, one of it is for calculating the frequency of occurrence of phonemes of a particular language.

Kannada is one of the major languages of India spoken by over 40 million people, particularly in the state of Karnataka located in southern part of India. Kannada is one of the 15 languages listed in the Eighth Schedule of the Indian Constitution (Sridhar, 2007). Kannada is one of the four major literary languages of the Dravidian family, the other three being Telugu, Tamil and Malayalam. Kannada is one of the traditional languages among Dravidian family with a fine grammatical tradition. Kannada has a very complex range of regional, social and stylistic variations: the Mysore/Bangalore dialect, the coastal dialect (Mangalore), the northern dialect (Dharwar) and Kalaburgi dialect (Upadhyaya, 1976). The Kannada lexicon has been enriched by uninhibited borrowing from several sources, majorly from Sanskrit, Hindi-Urdu, and English.

Kannada is a diglossic language. Nayak (as cited in Sridhar, 2007) describes the formal variety of Kannada which differs from spoken variety in several aspects including phonology, morphology, lexicon and syntax. The Kannada script which is evolved from the 5^{th} century Kadamba script is used to write the Kannada language. The history of Kannada can be conventionally divided in three periods; initial old Kannada (halegannada) flourished during the 6th Century in Ganga dynasty from 450-1200 A.D., middle Kannada (nadugannada) from 1200-1700 A.D., and modern Kannada from 1700 to present. Panini's

grammer, non Paninian's schools of Sanskrit grammer (Katantra & Sakatayana schools) and Prakrit grammer are the sources of influence on Kannada grammer.

Similar to other Dravidian languages, Kannada also follows the canonical word order of SOV (Subject - Object - Verb). The word order is fairly free, since noun phrase are marked for case and verbs for agreement with subject in number, gender and person. Kannada is a highly inflected language with three genders (masculine, feminine, and neuter or common) and two numbers (singular and plural). Phonologically, there is a series of retroflex consonants, a series of voiced and voiceless aspirates (borrowed from Indo-Aryan) and frequent use of vowel deletion rules. The Kannada lexicon has been enriched by uninhibited borrowing, from several sources, principally Sanskrit, Hindi-Urdu and English (Sridhar, 2007).

Kannada language consists of 49 characters in its alphasyllabary and is phonemic. As different characters can be combined to form compound characters (ottaksharas), the number of written symbols however is far more than the 49 characters. The characters are divided into three groups: swaras (vowels), vyanjanas (consonants) and yogavaahakas (part vowel, part consonants). Two types of consonants have been identified in Kannada script; the structured consonants and the unstructured consonants. Table 1.1 shows the structured consonants of Kannada.

Table 1.1

Structured consonants of the Kannada script.

	Voiceless	**voiceless aspirate**	**Voiced**	**voiced aspitare**	**Nasal**
Velars	ಕ (ka)	ಖ (k^ha)	ಗ (ga)	ಘ (g^ha)	ಙ (nga)
Palatals	ಚ (ʧa)	ಛ ($ʧ^h$a)	ಜ (ʤa)	ಝ ($ʤ^h$a)	ಞ (ña)
Retroflex	ಟ (ʈa)	ಠ ($ʈ^h$a)	ಡ (ɖa)	ಢ ($ɖ^h$a)	ಣ (ɳa)
Dentals	ತ (ta)	ಥ (t^ha)	ದ (da)	ಧ (d^ha)	ನ (na)
Labials	ಪ (pa)	ಫ (p^ha)	ಬ (ba)	ಭ (b^ha)	ಮ (ma)

Unstructured consonants: the unstructured consonants are consonants that do not fall into any of the structured consonant characteristics. Each sound has a distinct letter and its own pronunciation. Table 1.2 shows the unstructured consonants of the Kannada language.

Table 1.2

Unstructured consonants of the Kannada script

ಯ (ja)	**ರ** (ra)	**ಲ** (la)	**ವ** (va)	**ಶ** (ʃa)
ಷ (ṣa)	**ಸ** (sa)	**ಹ** (ha)	**ಳ** (ɭa)	**ಕ್ಷ** (ksha)

Table 1.3

Thirteen vowels (swaras) of Kannada script

ಅ (a)	**ಆ(a:)**	**ಇ (I)**	**ಈ(I:)**	**ಉ (u)**	**ಊ(u:)**	**ಋ**
ಎ (e)	**ಏ (e:)**	**ಐ (aI)**	**ಒ (o)**	**ಓ (o:)**	**ಔ (au)**	

The yogavaahakas include two letters, the anusvara and the visarga. The anusvara is the diacritic used to mark a type of nasalization. Depending on the location of the anusvara the pronunciation varies in the word and the language within which it has been used. The anusvara in Kannada is ಅಂ (am). The visarga is a Sanskrit word which means sending forth or discharge. Visarga and anusvara appears between vowels and stop consonants. Visarga in Kannada is ಅಃ (aha).

Need for the study

In the last two decades, Kannada data has played a role in the development of linguistic theory in a number of areas including the definition of grammatical relations. The role of studying the phonemes has set up an importance in understanding the Dravidian languages. Determining frequency of occurrence of phonemes is foundation for linguistics and offers beneficial information to research and clinical fields.

India being a multicultural and multilinguistic background, there are dearth of studies about 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. Also knowledge about the most frequently occurring phonemes can help in targeting those phonemes in therapy for individuals with communication disorders. 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 (Egan, 1948; Campbell, 1965). Such information can also help in the development of different aids and devices like text to speech converters for individuals with communication disorders.

Earlier few studies on frequency of occurrence of phonemes in Kannada have considered only written source of materials (Ramakrishna, 1962; Ranganatha (1982); Jayaram, 1985). Manjula et al (2012) considered relatively a small corpus of phonemes including both written and spoken data in Kannada. Their spoken data included only abot 5000 phoneme occurrences. As Nayak (as cited in Sridhar, 2007) describes that the formal variety of Kannada differs from spoken variety in several aspects of phonology, the phoneme frequency data is essential in spoken form of Kannada also. Hence, the present study was planned to obtain the frequency of occurrence of phonemes in spoken Kannada from a large corpus.

Aim of the Study

To identify the frequency of occurrence of phonemes in the spoken dialect of Mysore Kannada.

Objectives

1. To obtain a database for the frequency of occurrence of various phonemes in Kannada using conversation samples.
2. From the database obtained, to calculate the frequency and percentage of occurrence of each phoneme in spoken Kannada.

Review of Literature

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. Most of the studies carried out were in languages like English, Spanish, French, Japanese and African languages. These studies have used written or spoken samples as data for calculating phoneme frequency. The studies in these languages focus on several aspects like frequency of occurrence of phonemes, morphemes, syllables, words etc. Phonetic transcription is used to represent the speech sounds, irrespective of written or spoken samples to calculate the phoneme frequency. Different transcription systems may be appropriate for different purposes (Trager & Smith, 1951). A transcription may be impressionistic (narrow) or systematic (broad), depending on whether the symbols are simple/comparative and phonemic or allophonic. Majority of the studies which are quoted in the literature have used broad transcription for calculating phoneme frequency.

Studies on frequency of occurrence of phonemes majorly concentrated to develop speech materials for audiologists, as well as for speech language pathologists for various clinical purposes (Palai & O'Hanlon, 2004). 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.

c) Phoneme frequency: written vs spoken data.

d) Phoneme frequency: comparison of vowel and consonantal data.

e) Frequency of occurrence of phonemes in Indian languages.

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

The earlier studies on frequency of occurrence of phonemes concentrated data from written sources like journals, magazines, periodicals etc. Yegerlehner and Voegelin (1957) 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). Conversation samples were taken as data for calculating the frequencies of phonemes. The data of languages which do not have a script was translated to a script of standard language which is widely used in those countries. The frequency of occurrence of phonemes was arranged in decreasing order as follows

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/ and /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/.

Guirao and Barzone de Manrique (1972) analyzed phoneme frequency in American Spanish. With computer assistance, a total of 14,577 words were included from various sources like witten dialogues, newspaper notes and written passages of modern play. A total of 62,980 phonemes were obtained which consisted 46.2% of vowels and 52.8% of consonants. Vowel /e/ occurred most frequently followed by /a/, /o/, /s/, /n/ and /i/.

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.

Table 2.1 shows the percentage of frequency of occurrence of phonemes in English according to Mitton (1992). /i/, /ə/, /æ/, /e/ were the most frequently occurring vowels and /t/, /s/, /n/, /l/, /r/, /k/ and /d/ were the most frequently occurring consonants. Phonemes /w/, /i/, /a/, /u/, /ɔ/, /h/, /ʧ/, /ʤ/, /ʒ/ occurred in less than 1% of the total data.

Table 2.1

The frequency of occurrence of phonemes in English (Mitton, 1992)

Phonemes	**/i/**	**/t/**	**/s/**	**/n/**	**/ə/**	**/l/**	**/r/**	**/k/**	**/d/**
Percentage of occurrence	10.5%	6.9%	6.8%	6.4%	6.2%	5.5%	4.6%	4.5%	4.3%

Zurinskas (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 transcription of words in IPA was according to the US accent. 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 in 7% of the total data. Following these phonemes vowel /u/ occurred in 6.7%, and consonants /s/ /d/ /e/ occurred in 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/.

Very few studies have been quoted in literature about the frequency of occurrence of syllables. One of the studies is by Tamaoka and Makioka (2004) where they studied the frequency of occurrence of phonemes, morae and syllables in Japanese. As Japanese is a mora timed language, the frequency of occurrence of morae was also considered for this research. The data considered were a corpus of words from a Japanese newspaper. The results indicated that vowels /u/ and /a/ had similar frequency rates of 23% followed by /i/ and /o/ whereas among consonants /k/ (17%) and /t/ (15%) were most frequently occurring

consonants followed by /r/ (13%). The special sound /N/ appeared very frequently (7.5%) and the syllable combination /k/+V+/N/ (19.91%) appeared most frequently with the nasal /N/.

Renwick (2011) studied the relative frequencies of phonemes in Romanian. The relative occurrence of high vowel /ɨ/ which was a recent addition to Romanian language was given more importance in the study. The vowels with different combinations of consonants, forming segments were analyzed. The data consisted of about 88,580 words consisting of 7, 88, 119 phonemes. The most occurring vowels were /i/ in about 25% of the total vowel occurrences, /e/ and /a/ in 20% of total vowel occurrences. Among consonants /r/ is most occurring consisting of about 16% of total occurrences of consonants. Top five occurring consonants in descending order are /r/, /t/, /n/, /l/, /s/. These five consonant make up for about 67% of total consonants in Romanian. The study mainly compares the occurrence of different consonants preceded by frequently occurring vowels. Consonants /l/ and /n/ were preceded by /i/ in about 10%, /r/ and /n/ were preceded by /e/ 12% and 10% respectively.

b) Phoneme frequency from spoken data

Till 1970s the study of phonemes were mainly concentrating on the written context where the frequencies were mostly calculated from sources like newspapers, journals, script of plays etc. The first study which used spoken material as data was by Mines, Hanson and Shoup (1978) which was through an interview method. Mines et al (1978) studied the frequency of occurrence of phonemes in conversational English. Sixteen adult males and ten adult females were involved in a casual conversation in an interview and database of about

1,03,887 phoneme occurrences were obtained. The frequency of occurrences of phonemes were listed in descending order as /a, n, t, i, s, r, i, l, d, ɛ/. These phonemes accounted for 47% of the data. The rest of the phonemes contributed for 53% of the data. Authors compared their study with other similar studies and found some minimal variations with the occurrences of phonemes.

Similar study was carried out in American Spanish by Guirao and Jurado (1990). They considered speech samples from five modern plays of American Spanish. A corpus of 1,63,861 phoneme occurrences was analyzed using specific computer software. The phoneme frequency distributions were analyzed for different categories like single phonemes, syllable position (initial and final) and manner of articulation. Results of their study indicated that vowels /e/ (15%), /a/ (13.3%) and /o/ (10.7%) were the most occurring vowels which accounted for 39% of the phoneme occurrences. Among consonants /s/, /n/, /r/, /t/, /k/, /d/ occurred most frequently accounting for the other 38%. The fricative sound /s/ was the most occurring consonant accounting for about 10% of the total phoneme occurrences. Consonants /f/, /x/, /ʧ/, /ʒ/, /rr/ occurred for less than 1% of total data.

Palai and O'Hanlon (2004) studied the phoneme frequency in Setswana language (the national language of Botswana, South Africa). The taped conversation samples of 118 speakers from radio programs were analyzed for the frequency of phoneme occurrences. The data involved conversations between two or more people consisting of 50,569 words. Among these 49, 358 (97.6%) were Setswana and 1, 211 (2.4%) non-Setswana words. These words were analyzed for the total number of phonemes, phonemes at initial position and phonemes at final position. A total of 82,461 Setswana consonant phonemes accounted the data of

Setswana language. The occurrence of consonant phonemes was 54% at initial position, 41.8% in medial and 4% in final position. The order of the most frequently occurring phonemes are /l/ (11.2%), /n/ (8%), /g/ (8.5%), /r/ (8.5%), /b/ /m/ /k/ (7%). /n/, /g/, /b/ at initial position of words, /l/, /r/, /m/ being most frequently occurring phonemes at medial position and /ŋ/ being the only consonant phoneme occurring at final position. But the study did not account the vowels for the frequency count.

c) Phoneme Frequency: Written vs spoken data

Although there are no major differences in language aspects of written and spoken basis of a particular language, the variations in speaking, stress and prosody influences certain aspects of speech production. Hence these can influence the phoneme frequency data when comparing written and spoken data.

The studies on frequency of occurrence of phonemes from different languages either considered samples from the written form or from the oral form like conversations, recorded plays etc. When the results of these studies are compared there do exist some differences among the written mode and spoken mode. Hence Crystal (1995) compared 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 text by Fry (cited in Crystal, 1995) was considered for spoken material. Crystal compared the frequency of occurrence of phonemes of both of these studies and found that there were considerable variations when comparing the spoken and written data.

Table 2.2

Frequency of occurrence of vowels (a), consonants (b) from written and spoken English in percentage

Vowels	/i/	/ə/	/æ/	/e/
Written Data	10.5%	6.2%	2.3%	2.3%
Spoken Data	8.3%	10.7%	1.4%	2.9%

(a)

Consonants	/t/	/s/	/n/	/d/	/l/	/r/	/k/	/p/
Written Data	6.9%	6.8%	6.4%	4.3%	5.5%	4.6%	4.5%	3.1%
Spoken Data	6.4%	4.8%	7.5%	5.1%	3.6%	3.5%	3%	1.7%

(b)

Table 2.2 (a) represents the frequency of occurrence of vowels from written and spoken data (Crystal, 1995). On comparison between the written and spoken data there were lot of variations in percentage of occurrence of vowels. In written data vowel /i/ was found to be the most frequently occurring phoneme whereas in spoken data it was vowel /ə/. Also vowel /u/ which has less than 1% occurrence in written data was found to occur more frequently in spoken data (1.2%). Table 2.2 (b) represents the frequency of occurrence of consonants from written and spoken data as found by Crystal (1995). When written and spoken data were compared for consonants there were variations that are similar to vowels. Consonants /p/, /k/, /v/, /d/, /w/, and /h/ occurred more frequently in spoken data compared to written sources.

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. The results of the study indicated that the phonemes /a/ and /e/ were the most occurring phonemes in both spoken and written context. /a/ occurred in 12.2% in spoken and 12.8% in written and /e/ occurred in 15% in spoken and 12% in written contexts. /o/ occurred for 10% and 9% in spoken and written contexts respectively. Among consonants /s/ occurred for 8% and 7% in spoken and written context respectively. Followed by /s/, it was /n/ which occurred for 7% in both spoken and written contexts. /l/ and /d/ had frequency of occurrences of about 4% in spoken and 5% in written form. Syllables which occurred frequently in spoken context are /a/ (4.9%), /ke/ /de/ (3.5%), /es/, /i/, /no/ (2%), whereas the syllables which occurred frequently in written context were /a/ (4.5%), /ke/ (3.5%), and /de/ (2.6%). Sandoval's study indicated that even though there is not much of a difference in written and spoken data of frequency of occurrence of phonemes or syllables, there are variations across the two modes.

d) Phoneme Frequency - Comparison of Vowel and Consonantal data

The percentage of occurrence of vowels and consonants in a particular language and across languages will give an idea about the use of phonemes and its frequency. Yegerlehner and Voegelin (1957) compared vowel to consonant ratios in nine different languages.

Table 2.3

Percentage of vowels and consonants in nine languages

Languages	Vowels (%)	Consonants (%)
Maori	58.7	41.3
Winnebago	51	49
Shawnee	46.5	53.5
Choctaw	49.2	50.8
Havasupai	47.5	52.5
Navaho	44	56
Chontal	44.2	55.8
Hidasta	47.4	52.6
Tarascan	50	50

Table 2.3 shows the percentage of vowel and consonant occurrences among nine different languages. It was noted that only Maori (spoken in Malayo Polynesian) and Winnebago (Eastern Woodlands, Siouan) had more vowel dominance and other six languages had consonant dominance over vowels. Tarascan (spoken in few regions of Mexico) is the only language where vowels and consonants occurred equally.

Guirao and Garcia Jurado (1990) indicated that vowels /e/, /a/ and /o/ were the most occurring vowels which accounted for 39% of phoneme occurrences. Among consonants /s/, /n/, /r/, /t/, /k/, /d/ occurred most frequently accounting for 38%. Vowels all together

constituted 48.4% of the total data out of which the three most occurring vowels constituted 39% whereas occurrence of consonants was 51.6% of the total data.

Delattre's (as cited in Edwards, 2003) study in English indicated that the vowel constituted 42.5% of the total data. Most frequently occurring vowels /ə/, /i/, /æ/ constituted 36.2% of vowel data. The consonant data constituted 57.3% of the total data. Consonants /t/, /n/, /r/ and /l/ constituted of 30.3% of the total consonantal data.

Table 2.4 represents the different studies on frequency of occurrence of phonemes in American Spanish. The results of Navarro Tomas (as cited in Guirao & Garcia Jurado, 1990) and Zipf and Rogers (as cited in Guirao & Garcia Jurado, 1990) indicated that the vowel data constituted 43.5% of the total data and consonant data was 56.5% of the total data. Most frequently occurring vowels /a/, /e/, /o/ constituted for 33.7% of vowel data in Navarro Tomas's study whereas 35.9% in Zipf and Roger's study. Phonemes /s/, /n/, /r/ and /l/ were the most frequently occurring consonants constituting 26.8% of consonantal data in Navarro Tomas's study whereas 25.1% in Zipf and Roger's study. Similar studies by Quilis and Esgueva (as cited in Guirao & Garcia Jurado, 1990) and Guirao and Borzone (as cited in Guirao & Garcia Jurado, 1990) calculated the vowel and consonant percentage of occurrence of phonemes in Spanish and Roman languages respectively. /e/, /a/ and /o/ were most occurring vowels in both the studies. The most occurring consonants in Quilis and Esgueva's study were /s/, /t/, /d/, /l/ whereas /s/, /n/, /r/, /t/ occurred most frequently in Guirao and Borzone's study. But the vowel and consonant percentage in these two studies were similar. All the vowels constituted 47% of the total data whereas consonants were 52% of the total data.

Table 2.4

Frequency of phonemes data from several studies in American Spanish (in %)

	Zipf and Rogers	Navarro Tomas	Guirao and Borzone	Quilis and Esgueva
Vowels	43.5%	43.5%	47.2%	47.6%
Consonants	56.5%	56.5%	52.8%	52.4%

Zurinskas's (2002) study on phoneme frequency in English indicated that 40.2% of the total data constitute vowels. The vowels /i/, /u/ and /æ/ constituted 18.1% of the vowel data. Other 59.6% of the total data constituted consonants among which /n/, /t/, /s/, and /d/ were the most frequently occurring consonants constituting 23.8%.

e) Frequency of occurrence of phonemes in Indian Languages

Very limited studies are available for the frequency of occurrences of phonemes, but majority of these researches are limited to non-Indian languages. Initial study on Indian language available is by Bhagwat (1961), where he calculated the phonemic and morphemic frequencies in Marathi. He suggested this could be a source or database for devising a speed script.

Among the studies during 1960s, relative contribution by Ghatage was very important as he studied the frequency of phonemes in different languages. Ghatage (1964) calculated the phonemic and morphemic frequencies in Hindi. He considered written source of

materials for the study. Results showed that vowels occurred more frequently than consonants.

Followed by Hindi, Ghatage (1994) studied phonemic and morphemic frequencies in Malayalam. He considered 1,00,000 words from various written materials like light literature (novels, poems, plays and poems), periodicals, books, scientific articles etc and also spoken data from radio plays. The frequencies of phonemes and morphemes were calculated forming a database in Malayalam. The results indicated /a/ and /i/ were the most occurring vowels. Among consonants bilabial nasal /m/, palatal nasal /n/ and velar /k/ were the most occurring consonants.

Kelkar (1994) studied phonemic and morphemic frequencies in Oriya. The sources were similar to that of Ghatage's study. The results indicated that the vowel /ə/ was most occurring followed by /a/ and /i/. Among consonants /r/, /k/ and /t/ occurred most frequently.

The research on frequency of occurrence of phonemes in Kannada was first studied by Ramakrishna (1962). He analyzed the speech sounds in Kannada. The results report that long vowels and aspirated phonemes are used relatively less frequently. Also vowel /a/ is the highest occurring vowel and consonants like /r/, /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), 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 was /a/, /i/, /u/, /r/, /d/, /a:/, /e/, /n/, /t/, /k/, /g/ and /v/.

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 a series of vowels and consonants by rank order of their frequency of occurrence and the order was /a, I, n, r, u, d, ɔ, 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. The most frequently occurring vowel was /a/ followed by /i/, /u/ and /e/. Vowels constituted 43.5% of the total data.

Manjula, Geetha, Sharath and Antony (ARF project, 2011-12) calculated the frequency of occurrence of phonemes in Kannada. 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. The results of their study indicated that the most frequently occurring vowel was /a/ followed by /i/. Among consonants /n/, /r/, /t/, /l/ were the most frequently occurring phonemes.

A pilot study was carried out by Sreedevi, Smitha and Vikas (in Press) on frequency of occurrence of phonemes in Kannada as a part of the present study. Recorded conversation speech samples of 21 speakers in five different recording sessions were considered for calculating phoneme frequency. Each recording involved conversations between three to five people. A total of 13,215 words accounted the data in which 69,624 phoneme occurrences were observed. The most frequently occurring phonemes in the descending order were /a/, /n/, /i/, /l/, /e/, /d/, /a:/, /r/, /t/, /u/, /g/ and /k/ which constituted 70.2% of the total data. Figure 1 shows the mean percentage of most frequently occurring phonemes.

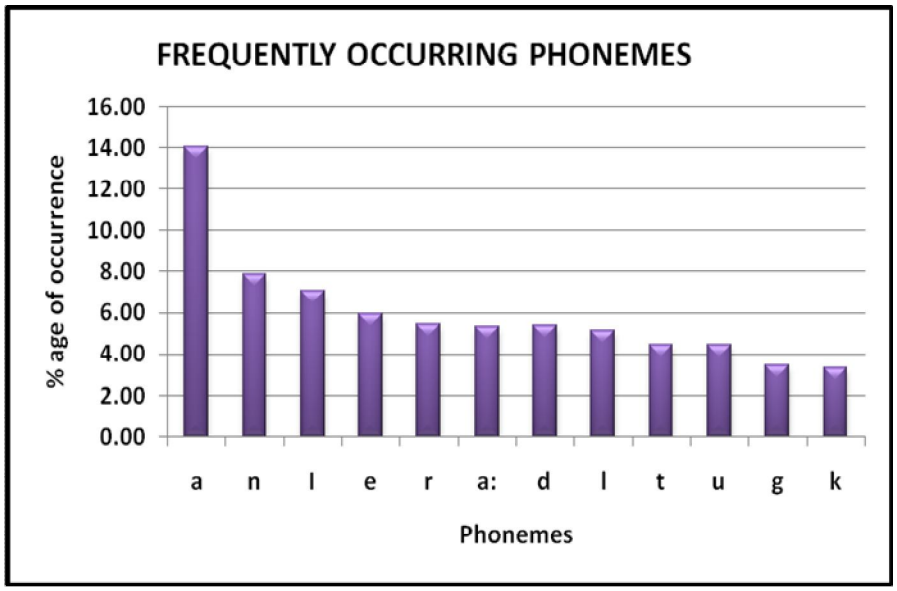

Figure 1: Mean percentage of frequency of occurrence of frequently occurring phonemes in Kannada

The consonants /m/ and /v/ occurred 2.7% of the total data. Vowel /o/ occurred 2% whereas phonemes /h/, /s/, /j/ occurred less than 2% and /p/, /ʧ/, /ʃ/, /iː/, /oː/ occurred in less than 1% each of the total corpus.

The review of research on frequency of occurrence of phonemes indicates that it has been extensively studied in various languages across the world. Data from sources like written as well as spoken are available in several languages. The phoneme frequency varies across languages and also when the data is compared across written and spoken sources of the same language, and there exists small differences in the phoneme frequency and order. A common finding from most of the studies across the Indian and non-Indian languages is that consonants occurred more frequently compared to vowels.

Method

Participants: Adult fluent native speakers of Mysore dialect of Kannada in the age range of 20 to 50 years with a minimum of 10 to 12 years of education in Kannada medium were selected for the study. All the participants were native speakers of Mysore dialect of Kannada and resided in Mysore city. Participants did not have any clinical background of speech, language, hearing or any neurological problems. A group of 3 to 5 participants were involved in each recording session. Each recording session had different participants engaged in conversation. Data consisted a total of 20 recordings which included 74 participants out of which 28 were females and 46 were males. Table 3.1 shows the details of the participants included in each recording session.

Table 3.1

The number of participants in each recording session

Recording Sessions	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
No.of Participants	4	5	4	4	4	4	3	4	3	4
Males	2	4	3	1	2	2	2	3	2	2
Females	2	1	1	3	2	2	1	1	1	2
Recording Sessions	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20
No.of Participants	4	3	4	3	3	3	4	5	3	3
Males	2	2	3	3	3	2	2	2	2	2
Females	2	1	1	0	0	1	2	3	1	1

Instrumentation: A digital recorder (Olympus WS 100) was used for recording the conversation samples. Dell (Inspiron 1545) laptop with headphones (i-ball 1600 MV) was used for the transcription and analysis.

Procedure: The demographic data, education, general medical history of the participants were noted prior to the each recording. The data was collected through conversation in controlled natural environments which lasted for about 25 to 30 minutes of duration. All the participants were made to sit comfortably facing each other and familiarize with other participants of the session. The digital recorder was kept at equidistant from all the speakers. Participants were not given any prior topic about the conversation to avoid the high occurrence of certain phonemes. Any current topic of interest was initiated by the participants themselves and conversations were carried out for half an hour. The participants were instructed to avoid words from other languages as much as possible and to speak in Kannada only as naturally as possible. They were not restricted from using commonly used loan English words (E.g.: Bus, ticket, phone, car etc). Appendix 1 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

Topics of conversation in each recording session

Recording Sessions	Topic of Conversation
Recording 1	Current Education system and political issues.
Recording 2	Family issues, corruption, present government
Recording 3	Casual talk on garments, provisions, postal dept and government officials.
Recording 4	Various cultural and therapeutic programs in Mysore city
Recording 5	Legal issues, family values and studies abroad
Recording 6	Casual talk, food items.
Recording 7	Educating children, general conversation.
Recording 8	Casual talk on current issues in the society.
Recording 9	General conversation, home maintenance issues.
Recording 10	Water problems in the city, political issues.
Recording 11	Casual talk on jewellery items and price hike of such items.
Recording 12	Corruption issues, issues on drainage system.
Recording `13	Protest and developments in the state.
Recording 14	Casual talk on terrorism, general public issues.
Recording 15	Political parties and its duties, issues regarding problems in the nation.
Recording 16	Corruption, lokayukta force and political issues.
Recording 17	Casual talk on laws and its enforcement, women's rights.
Recording 18	Casual talk on devotional aspects and role of youth.
Recording 19	General conversation, festivals and celebrations.
Recording 20	Current political issues, general conversation.

Data Analysis: The conversation samples were transcribed using IPA transcription. The repetitive words (eg: I may go to go to bank) and exclamatory remarks were excluded from analysis. English words that are very 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 in the software SALT RV version 9 for the frequency count. 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, pragmatic, rate, fluency, and error categories. This software analyzes 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.

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 Kannada phonemes was prepared and saved in the editable standard wordlists of SALT software. The database consisted of all the phonemes available in Kannada adapted from Sridhar (2007) which was then modified (e.g: /t^h/ is modified as /th/, /ʃ/ is modified as /sh/, /ʧ/ is used as /c/) according to SALT conventions. The whole conversation data is edited with spacing after every phoneme and this file is loaded into SALT. The Salt software compares the database and the loaded phoneme file and provides the phoneme count.

Inter-judge Reliability: Sample of each conversation recording was subjected to inter judge reliability measures. Three judges including two post graduate speech language pathologists and a clinical linguist 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% sample of the 30 minutes recording of each sample was transcribed by each of the three judges. The recorded samples

were played to the judges independently. They were not allowed to discuss about the transcription of the sample before or after the task. The frequency of occurrence of phonemes from the three judges was noted and was subjected for statistical analysis. Cronbach alpha test was used and a reliability index (alpha) of 0.95 was obtained for inter judge reliability.

Intra Judge Reliability: For intra judge reliability 10% of each of the 20 recordings were separately transcribed and analyzed by one of the authors (a speech language pathologist) after transcribing all the samples completely once. Both the transcriptions were analyzed for the frequency of occurrence of phonemes. The phoneme counts were subjected for statistical analysis. Cronbach alpha score of 0.9 was obtained for intra judge reliability.

Results

The present study aimed to obtain the frequency of occurrence of phonemes in conversational speech in Mysore dialect of Kannada. The results of phoneme count of 20 recordings were tabulated. The vowel and consonant counts from each recording were noted. The results are discussed under the following headings

a) Total phoneme counts
b) Vowel and Consonant percentage
c) Most frequently occurring phonemes
d) Frequency of occurrence of all phonemes

a) Phoneme Counts

A total of 55,503 words were considered for calculating the frequency of occurrence of phonemes from the obtained conversation samples of 20 sessions of recording which constituted 2,68,625 phoneme counts. Table 4.01 represents the number of words and Table 4.02 represents the number of phonemes obtained from each of the 20 recording sessions.

The word counts in the 20 samples ranged from 1,971 words to 3,873 words with a mean of 2,775 words. Recording 4 (R4) consisted of maximum number of words and recording 2 (R2) consisted of minimum number of words. Out of the 20 recordings, recording 1 (R1) had the maximum number of phonemes and recording 7 (R7) had the minimum number of phonemes.

Table 4.01

Total number of words obtained from each of the 20 conversation samples

No.of the Recording Session	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
No of words obtained	3,571	1,971	2,293	3,873	3,583	2,539	2,027	2,322	2,249	2,895
No.of the Recording Session	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20
No of words obtained	3,194	2,189	2,368	2,322	2,386	3,675	3,214	3,175	2,984	2,673

Table 4.02

Total number of phonemes obtained from each of the 20 conversation samples

No.of the Recording Session	R1	R2	R3	R4	R5	R6	R7
Total no of Phonemes obtained	18,221	9,557	9,574	15,962	16,310	10,520	9,411
No.of the Recording Session	R8	R9	R10	R11	R12	R13	R14
Total no of Phonemes obtained	11,458	12,566	14,354	16,424	12,402	13,060	11,572
No.of the Recording Session	R15	R16	R17	R18	R19	R20	
Total no of Phonemes obtained	13,068	13,694	16,542	16,510	15,962	11,458	

b) Vowel and Consonant Percentage

Using descriptive statistics, the mean and standard deviations of the frequency of occurrence of vowels and consonants were calculated. Table 4.03 represents the frequency of vowels and consonants calculated from all the 20 recording sessions. The total vowel count

was 1,19,296 which constituted 44.37% of the total data whereas consonant count accounted for 1,49,269 which is 55.3% of the total data.

Table 4.03

Frequency of occurrence of vowels and consonants

No.of the Recording Session	R1	R2	R3	R4	R5	R6	R7
No.of vowels	8,173	4,228	4,158	7,001	7,386	4,602	4,227
No.of consonants	10,048	5,269	5,416	8,961	8,924	5,918	5,184
No.of the Recording Session	R8	R9	R10	R11	R12	R13	R14
No.of vowels	5,017	5,525	6,315	7,307	5,506	5,749	5,136
No.of consonants	6,441	7,041	8,039	9,117	6,896	7,311	6,436
No.of the Recording Session	R15	R16	R17	R18	R19	R20	
No.of vowels	5,787	6,411	7, 264	7, 446	7, 001	5,057	
No.of consonants	7,281	7,283	9,278	9,064	8,961	6,401	

The vowel occurrences in the 20 recordings ranged from 4,158 to 8,173 with an average of 5,964 vowel occurrences. Recording 3 (R3) had minimum vowel occurrences whereas recording 1 (R1) had maximum number of vowel occurrences. Similarly consonant occurrences in the 20 recordings ranged from 5,184 to 10,048 with an average of 7,463 consonant occurrences. Recording 7 (R7) had minimum consonant occurrences whereas recording 1 (R1) had maximum number of consonant occurrences. The percentage of vowel frequency in the 20 samples ranged from 43.43% to 46.82% with a mean of 44.3%. Recording 3 (R3) had the minimum vowel frequency whereas recording 16 (R16) had the maximum vowel occurrences. The consonant percentage ranged from 53.18% to 56.57%

with a mean of 55.3%. Recording 16 (R16) had the mimimum consonant frequency whereas recording 3 (R3) had the maximum consonant occurrences.

Table 4.04

Frequency of occurrence of vowels and consonants in percentage

No.of the Recording Session	R1	R2	R3	R4	R5	R6	R7
Vowel %	44.85	44.24	43.43	43.86	45.29	43.75	44.92
Consonant%	55.15	55.13	56.57	56.14	54.71	56.25	55.08
No.of the Recording Session	R8	R9	R10	R11	R12	R13	R14
Vowel %	43.96	43.97	43.99	44.49	44.40	44.02	44.38
Consonant %	56.04	56.03	56.01	55.51	55.60	55.98	55.62
No.of the Recording Session	R15	R16	R17	R18	R19	R20	---
Vowel %	44.28	46.82	43.91	45.10	43.86	43.96	---
Consonant %	55.72	53.18	56.09	54.90	56.14	56.04	---

c) Most frequently occurring phonemes

The results from the SALT software were tabluated for the most frequently occurring phonemes in each of the 20 recorded conversations. Recording 1 (R1) consisted of 3,571 words and 18,221 phonemes. Ten most frequently occurring phonemes of each of the 20 recordings are presented in a series of tables. Table 4.05 represents the frequency of most frequently occurring phonemes of R1. Vowel **/a/** occurred most frequently (occurred for 15.4% of the total data of Recording R1) followed by nasal **/n/** which occurred for 8.5% of the total data. **/n/** was followed by **/i/**, **/r/, /a:/, /e/,** dentals **/d/, /t/, /l/** and vowel **/u/.**

Table 4.05

Frequency of most frequently occurring phonemes in recording 1 (R1)

Phonemes	**/a/**	**/n/**	**/i/**	**/r/**	**/a:/**	**/e/**	**/d/**	**/t/**	**/l/**	**/u/**
Percentage	15.47	8.58	6.84	6	5.84	5.98	5.56	4.88	4.65	4.21
Frequency	2,819	1,563	1,246	1,094	1,065	1,090	1,013	889	847	767

Table 4.06

Frequency of most frequently occurring phonemes in recording 2 (R2)

Phonemes	**/a/**	**/n/**	**/i/**	**/e/**	**/r/**	**/l/**	**/a:/**	**/d/**	**/u/**	**/t/**
Percentage	13.74	7.43	7.20	6.18	5.67	5.46	5.25	5.13	3.99	3.68
Frequency	1,313	710	688	591	542	522	502	490	381	352

Recording 2 (R2) consisted of 9,557 phonemes from 1,971 words. Table 4.06 represents the frequency of most frequently occurring phonemes of R2. Vowel **/a/** occurred most frequently followed by nasal /**n**/. In R2, vowel **/e/** and consonant lateral **/l/** occurred more frequently when compared to recording R1. The order of frequently occurring phonemes are **/a/, /n/, /i/, /e/, /r/, /l/, /a:/, /d/, /u/,** and **/t/.**

Table 4.07

Frequency of most frequently occurring phonemes in recording 3 (R3)

Phonemes	/a/	/n/	/i/	/e/	/r/	/d/	/a:/	/l/	/t/	/u/
Percentage	13.46	7.99	7.07	6.08	5.5	5.5	5.22	4.99	4.5	4.13
Frequency	1,289	765	677	582	527	527	500	478	431	395

2,293 words were present in the recording 3 (R3) which consisted of 9,574 phonemes. Table 4.07 represents the frequency of most frequently occurring phonemes of R3. The order of phonemes of R3 were similar to recording R2.

Table 4.08

Frequency of most frequently occurring phonemes in recording 4 (R4)

Phonemes	/a/	/n/	/i/	/l/	/e/	/d/	/a:/	/r/	/t/	/u/
Percentage	13.76	7.61	7.19	5.82	5.69	5.53	5.16	5.09	5.05	4.49
Frequency	2, 196	1, 214	1, 147	929	909	882	823	813	806	717

3,873 words were present in the recording 4 (R4) which consisted of 15,962 phonemes. Table 4.08 represents the frequency of most frequently occurring phonemes of R4. The most frequent phonemes followed the trend similar to previous recordings that is vowel **/a/** occurred most frequently followed by nasal /**n**/ and short vowel **/i/**. But consonant lateral **/l/** occurred more frequently when compared to recordings R1, R2 and R3. Lateral /**l**/ was followed by **/e/, /d/, /a:/, /r/, /t/** and /**u**/.

Table 4.09

Frequency of most frequently occurring phonemes in recording 5 (R5)

Phonemes	/a/	/n/	/i/	/e/	/u/	/a:/	/d/	/r/	/l/	/t/
Percentage	13.88	7.73	7.1	5.74	5.41	5.29	5.13	4.87	4.77	4.21
Frequency	2,264	1,260	1,158	909	883	863	836	794	778	687

Table 4.09 represents the frequency of most frequently occurring phonemes of recording 5 (R5). 16,310 phoneme occurrences consisted the data of R5. Short vowels **/e/** and

/u/ occurred more frequently when compared to previous recordings. The most frequently occurring phoneme remained same as in previous recordings that is, vowel **/a/** was followed by nasal /**n**/ and short vowel **/i/.**

Tables 4.10 and 4.11 represent the frequency of most frequently occurring phonemes of recording 6 (R6) and recording 7 (R7) respectively. R6 consisted of 10,520 phoneme occurrences from 2,539 words and R7 consisted of 9, 411 phoneme occurrences from 2,027 words. The results of R6 and R7 are similar to that of R4. Consonant lateral **/l/** occurred more frequently when compared to recordings R3 and R5. The order of most frequently occurred phonemes remained almost same across both the recordings.

Table 4.10

Frequency of most frequently occurring phonemes in recording 6 (R6)

Phonemes	/a/	/n/	/i/	/l/	/e/	/r/	/a:/	/d/	/t/	/u/
Percentage	13.75	7.98	7.04	6.46	5.73	5.45	5.34	5.13	4.51	4.2
Frequency	1,447	839	741	680	603	573	562	548	474	442

Table 4.11

Frequency of most frequently occurring phonemes in recording 7 (R7)

Phonemes	/a/	/n/	/i/	/l/	/e/	/d/	/r/	/a:/	/t/	/u/
Percentage	13.55	7.28	7.08	6.34	6.18	5.53	5.49	5.3	4.52	4.4
Frequency	1, 275	685	666	597	582	520	517	499	425	414

Recording 8 (R8) consisted of 11,458 phoneme occurrences from 2,322 words and recording 9 (R9) consisted of 12,566 phoneme occurrences from 2,249 words. Tables 4.12

and 4.13 represents the frequency of most frequently occurring phonemes of R8 and R9 respectively. Long vowel **/a:/** occurred more frequently compared to other recordings. The order of most frequently occurred phonemes of R8 and R9 remained same. The most frequent occurring phonemes in descending order were **/a/, /n/, /i/, /a:/, /d/, /r/, /l/, /e/, /t/, /u/.**

Table 4.12

Frequency of most frequently occurring phonemes in recording 8 (R8).

Phonemes	/a/	/n/	/i/	/a:/	/d/	/r/	/l/	/e/	/t/	/u/
Percentage	13.39	7.7	6.52	5.92	5.38	5.23	5.16	4.68	4.67	4.51
Frequency	1,534	882	747	678	617	599	591	536	535	517

Table 4.13

Frequency of most frequently occurring phonemes in recording 9 (R9)

Phonemes	/a/	/n/	/i/	/a:/	/d/	/r/	/l/	/e/	/t/	/u/
Percentage	13.23	7.63	6.51	5.97	5.44	5.35	5.09	4.74	4.62	4.58
Frequency	1,662	959	818	750	684	672	639	596	580	575

Table 4.14

Frequency of most frequently occurring phonemes in recording 10 (R10)

Phonemes	/a/	/n/	/i/	/r/	/a:/	/d/	/t/	/e/	/l/	/u/
Percentage	16.35	7.59	6.24	5.89	5.68	5.34	4.63	4.56	3.91	3.62
Frequency	2,347	1,090	895	846	816	767	664	654	561	519

Table 4.14 indicates the frequently occurring phonemes of the recording 10 (R10). A total of 14,354 phoneme occurrences were noted from R10. The findings of the recording are similar to recording R1 where short vowel **/i/** was followed by trill **/r/** and long vowel **/a:/**. Most frequent phoneme remained **/a/** followed by nasal **/n/**, short vowel **/i/** similar to other recordings and the less frequent phoneme was vowel **/u/**.

Table 4.15

Frequency of most frequently occurring phonemes in recording 11 (R11)

Phonemes	/a/	/n/	/r/	/i/	/a:/	/d/	/t/	/e/	/u/	/l/
Percentage	16.04	6.88	6.8	6.26	6.22	5.66	5.07	5.02	4.32	4.23
Frequency	2, 634	1, 130	1, 117	1, 028	1, 022	929	832	825	709	695

Table 4.15 indicates the frequently occurring phonemes of the recording 11 (R11). A total of 16,424 phoneme occurrences were noted from R11. R11 had some salient findings. Consonant **/r/** occurred more frequently than any other recording sessions. Vowel **/a/** was found to be the most frequently occurring phoneme followed by **/n/** and **/r/**. And lateral **/l/** was found to be less frequently occurring.

Table 4.16

Frequency of most frequently occurring phonemes in recording 12 (R12)

Phonemes	/a/	/n/	/a:/	/i/	/r/	/d/	/e/	/l/	/u/	/t/
Percentage	16.82	7.06	6.28	5.85	5.74	4.88	4.18	4.02	3.93	3.89
Frequency	2,086	875	779	726	712	605	518	499	487	482

Table 4.16 represents the frequency of most frequently occurring phonemes of R12. Vowel **/a/** and nasal **/n/** were the most frequently occurring phonemes followed by long vowel **/a:/**. Long vowel **/a:/** occurred more frequently than in any other previous recordings. Dental stop **/t/** occurred less frequently similar to that of recordings R2 and R5.

Table 4.17

Frequency of most frequently occurring phonemes in recording 13 (R13)

Phonemes	/a/	/n/	/i/	/r/	/a:/	/d/	/t/	/e/	/l/	/u/
Percentage	16.51	7.62	6.23	6.03	5.73	5.51	4.59	4.37	3.82	3.68
Frequency	2,156	995	813	787	748	720	600	571	499	480

Table 4.17 indicates the frequently occurring phonemes of recording 13 (R13). A total of 13,060 phoneme occurrences were noted from 2,368 words. The findings of the recording 13 are similar to recording R1 and R10 where short vowel **/i/** was followed by trill **/r/** and long vowel **/a:/**. Most frequently occurring phoneme was **/a/** followed by nasal **/n/,** short vowel **/i/** similar to other recordings and the less frequent phoneme was vowel **/u/.**

Table 4.18 and 4.19 represents the most frequently occurring phonemes of recording 14 (R14) and recording 15 (R15) respectively. R14 consisted of 11,572 phoneme occurrences from 2,322 words and R15 consisted of 13,068 phoneme occurrences from 2,386 words. Long vowel **/a:/** occurred more frequently similar to that of recordings R8 and R9. The most frequently occurred phonemes remained the same, vowel **/a/,** nasal **/n/** and short vowel **/i/.**

Table 4.18

Frequency of most frequently occurring phonemes in recording 14 (R14)

Phonemes	/a/	/n/	/i/	/a:/	/e/	/d/	/r/	/l/	/t/	/u/
Percentage	14.03	7.62	6.28	5.85	5.37	5.33	5.18	5.11	4.62	4.47
Frequency	1,624	882	727	677	621	617	599	591	535	517

Table 4.19

Frequency of most frequently occurring phonemes in recording 15 (R15)

Phonemes	/a/	/n/	/i/	/a:/	/r/	/d/	/e/	/t/	/l/	/u/
Percentage	16.26	7.47	6.71	6.57	6.15	5.45	4.7	4.55	4.37	3.71
Frequency	2,125	976	877	859	804	712	614	594	571	485

Table 4.20

Frequency of most frequently occurring phonemes in recording 16 (R16)

Phonemes	/a/	/n/	/a:/	/i/	/r/	/d/	/e/	/l/	/u/	/t/
Percentage	16.53	6.97	6.16	5.92	5.83	4.91	4.37	3.99	3.96	3.93
Frequency	2,264	955	843	811	798	673	598	547	542	538

A total of 13,694 phoneme occurrences were noted from recording 16 (R16). Table 4.20 represents the frequency of most frequently occurring phonemes of R16. Phonemes **/a/** and **/n/** occurred most frequently followed by long vowel **/a:/**. Long vowel **/a:/** occurred more frequently similar to recording R12. Dental stop **/t/** occurred less frequently similar to that of recordings R2, R5 and R12.

Table 4.21

Frequency of most frequently occurring phonemes in recording 17 (R17)

Phonemes	/a/	/n/	/i/	/l/	/e/	/d/	/r/	/a:/	/t/	/u/
Percentage	13.87	7.67	7.09	5.78	5.68	5.6	5.11	5.08	5.02	4.43
Frequency	2,294	1,269	1,173	956	940	927	846	841	830	732

Table 4.21 represents the frequency of most frequently occurring phonemes of recording 17 (R17). R17 consisted of 16,542 phoneme occurrences from 3,214 words. R17 consists the maximum phoneme occurrences from all the 20 recordings. The results of R17 were similar to recordings R4, R6 and R7. Lateral **/l/** occurred more frequently following short vowel **/i/.** The order of most frequently occurred phonemes remained same as of R7.

Table 4.22

Frequency of most frequently occurring phonemes in recording 18 (R18)

Phonemes	/a/	/n/	/i/	/e/	/u/	/a:/	/d/	/r/	/l/	/t/
Percentage	13.77	7.75	7.14	5.67	5.35	5.29	5.06	4.81	4.71	4.16
Frequency	2,274	1,280	1,178	936	883	873	836	794	778	687

Table 4.22 represents the frequency of most frequently occurring phonemes of recording 18 (R18). A total of 16,510 phoneme occurrences consisted the data of R18. Vowels **/e/** and **/u/** occurred more frequently than compared to previous recordings. The most frequently occurring phoneme remained the same as in previous recordings, that is vowel **/a/** is followed by nasal /**n**/ and short vowel **/i/.**

Table 4.23

Frequency of most frequently occurring phonemes in recording 19 (R19)

Phonemes	/a/	/n/	/i/	/l/	/e/	/d/	/a:/	/r/	/t/	/u/
Percentage	13.7	7.61	7.2	5.82	5.71	5.53	5.17	5.09	5.01	4.51
Frequency	2,186	1,214	1,150	929	911	882	825	813	799	720

Table 4.23 represents the frequency of most frequently occurring phonemes of recording 19 (R19). A total of 15,962 phoneme occurrences were noted from 2,984 words. The results of R19 were similar to recordings R4, R6, R7 and R17. Lateral **/l/** occurred more frequently following short vowel **/i/.**

Table 4.24

Frequency of most frequently occurring phonemes in recording 20 (R20).

Phonemes	/a/	/n/	/i/	/a:/	/d/	/r/	/l/	/e/	/t/	/u/
Percentage	13.39	7.7	6.52	5.92	5.38	5.23	5.16	4.68	4.67	4.51
Frequency	1,534	882	747	678	617	599	591	536	535	517

Table 4.24 represents the frequency of most frequently occurring phonemes of recording 20 (R20). R20 consisted of 11,458 phoneme occurrences from 2,673 words. Vowel **/a:/** occurred more frequently similar to that of R8, R9, R14 and R15. The most frequently occurred phonemes remained same **/a/, /n/** and short vowel **/i/.** Less frequent phonemes were **/e/, /t/** and **/u/.**

The purpose of describing the frequency of occurrence of phonemes in each of the 20 recordings in detail is to show that inspite of the topic variation across recordings, the most frequently occurring 10 phonemes remained relatively stable. Table 4.25 depicts this observation.

Table 4.25

Most frequently occurring phonemes in Mysore dialect of conversational Kannada across the 20 conversational recordings

Recording	10 most frquently occurring phonemes									
R1	/a/	/n/	/i/	/r/	/a:/	/e/	/d/	/t/	/l/	/u/
R2	/a/	/n/	/i/	/e/	/r/	/l/	/a:/	/d/	/u/	/t/
R3	/a/	/n/	/i/	/e/	/r/	/d/	/a:/	/l/	/t/	/u/
R4	/a/	/n/	/i/	/l/	/e/	/d/	/a:/	/r/	/t/	/u/
R5	/a/	/n/	/i/	/e/	/u/	/a:/	/d/	/r/	/l/	/t/
R6	/a/	/n/	/i/	/l/	/e/	/r/	/a:/	/d/	/t/	/u/
R7	/a/	/n/	/i/	/l/	/e/	/d/	/r/	/a:/	/t/	/u/
R8	/a/	/n/	/i/	/a:/	/d/	/r/	/l/	/e/	/t/	/u/
R9	/a/	/n/	/i/	/a:/	/d/	/r/	/l/	/e/	/t/	/u/
R10	/a/	/n/	/i/	/r/	/a:/	/d/	/t/	/e/	/l/	/u/
R11	/a/	/n/	/r/	/i/	/a:/	/d/	/t/	/e/	/u/	/l/
R12	/a/	/n/	/a:/	/i/	/r/	/d/	/e/	/l/	/u/	/t/
R13	/a/	/n/	/i/	/r/	/a:/	/d/	/t/	/e/	/l/	/u/
R14	/a/	/n/	/i/	/a:/	/e/	/d/	/r/	/l/	/t/	/u/
R15	/a/	/n/	/i/	/a:/	/r/	/d/	/e/	/t/	/l/	/u/
R16	/a/	/n/	/a:/	/i/	/r/	/d/	/e/	/l/	/u/	/t/
R17	/a/	/n/	/i/	/l/	/e/	/d/	/r/	/a:/	/t/	/u/
R18	/a/	/n/	/i/	/e/	/u/	/a:/	/d/	/r/	/l/	/t/
R19	/a/	/n/	/i/	/l/	/e/	/d/	/a:/	/r/	/t/	/u/
R20	/a/	/n/	/i/	/a:/	/d/	/r/	/l/	/e/	/t/	/u/

Figure 2 represents the mean and standard deviation of the most frequently occurring phonemes from all the 20 recordings. On overall observation the frequency of occurrence of phonemes in descending order are: **/a/, /n/, /i/, /a:/, /r/, /d/, /e/, /l/, /t/, /u/.**

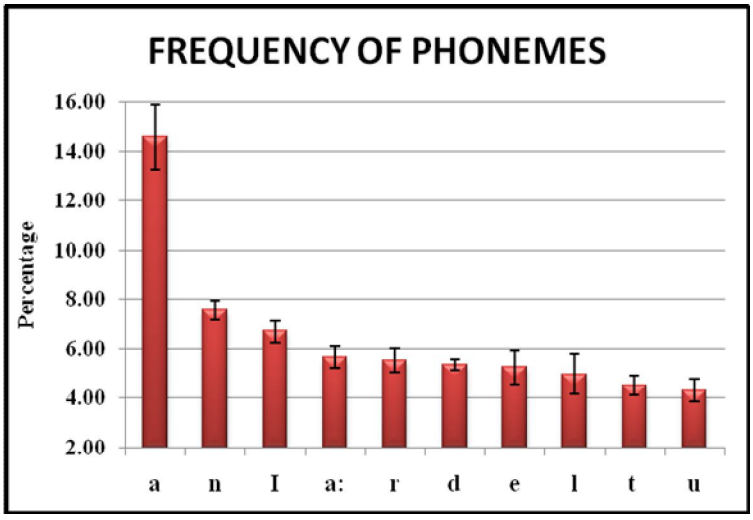

Figure 2: Mean frequency of occurrence of frequently occurring phonemes in percentage in Kannada. Error bars show ±1 standard deviation (SD)

d) Frequency of occurrence of phonemes

The frequency of occurrence of all the phonemes in conversational Kannada from all the 20 conversation recordings were tabulated. Overall the total data consisted of 2,68,625 phoneme occurrences. Mean and standard deviation of the frequency of occurrence of each phoneme from the 20 transcribed conversation samples were obtained. Table 4.26 provides the mean percentage and standard deviation of all phonemes from the 20 recorded conversation samples in Kannada.

Table 4.26

Mean percentage and standard deviation of frequency of occurrence of phonemes in Mysore dialect of conversational Kannada

Vowels		**Consonants**			
	Mean % (SD)		Mean % (SD)		Mean % (SD)
/a/	14.57 (1.3)	/n/	7.59 (0.37)	/ʈ/	1.53 (0.41)
/i/	6.70 (0.44)	/r/	5.53 (0.49)	/ɭ/	1.33 (0.22)
/a:/	5.66 (0.44)	/d/	5.35 (0.22)	/j/	1.24 (0.22)
/e/	5.27 (0.67)	/l/	4.98 (0.79)	/p/	1.22 (0.34)
/u/	4.32 (0.47)	/t/	4.54 (0.39)	/ʤ/	0.53 (0.26)
/e:/	2.23 (0.53)	/k/	3.49 (0.54)	/ʃ/	0.42 (0.25)
/o/	1.85 (0.46)	/g/	3.30 (0.73)	/ʧ/	0.34 (0.14)
/o:/	1.20 (0.28)	/m/	2.76 (0.26)	/ɳ/	0.29 (0.1)
/i:/	1.09 (0.21)	/v/	2.57 (0.16)	/bʰ/	0.09 (0.07)
/u:/	0.55 (0.08)	/ɖ/	2.24 (0.51)	/kʰ/, /tʰ/	0.05 (0.02)
/ ə/	0.35 (0.45)	/b/	2.08 (0.28)	/ ɖʰ/	0.20 (0.8)
Diphthongs		/h/	1.75 (0.38)	/ʈʰ/, /gʰ/	0.01 (0.02)
/au/	0.33(1)	/s/	1.75 (0.21)	/ŋ/, /pʰ/,	0
/al/	0.28(0.23)			/ɖʰ/, / ʤʰ/, / ʧʰ/,	0

The consonants /**m**/ and /**v**/ occurred in 2.76% and 2.57% of the total data. Vowel /**e:**/ occurred 2.23% whereas phonemes /**h**/, /**s**/, /**j**/, /ʈ/, /ɭ/ and /p/ occurred in less than 2% and /ʧ/, /ʃ/, /bʰ/, /ʤ/, /kʰ/, /tʰ/, /o:/ occurred in less than 1% each of the total data. Consonants /ŋ/, /pʰ/, /ɖʰ/, /ʤʰ/, /ʧʰ/ did not occur in the 20 recorded conversational samples albeit their presence in the Kannada phoneme system. The aspirated phonemes were rarely seen. Diphthongs /**ai**/ and /**au**/ occurred in less than 1% of the total data. /**ai**/ occurred in 0.28% and /**au**/ occurred in 0.32% of the total phoneme corpus obtained. Appendix 2 provides the frequency of occurrence of phonemes in spoken Kannada in descending order.

Figure 3 indicates the mean vowel and consonant frequency (in percentage) from the 20 recorded conversation samples. The total vowel count was 1,19,296 which constituted 44.37% of the total data whereas consonants accounted for 1,49,269 which is 55.59% of the total data.

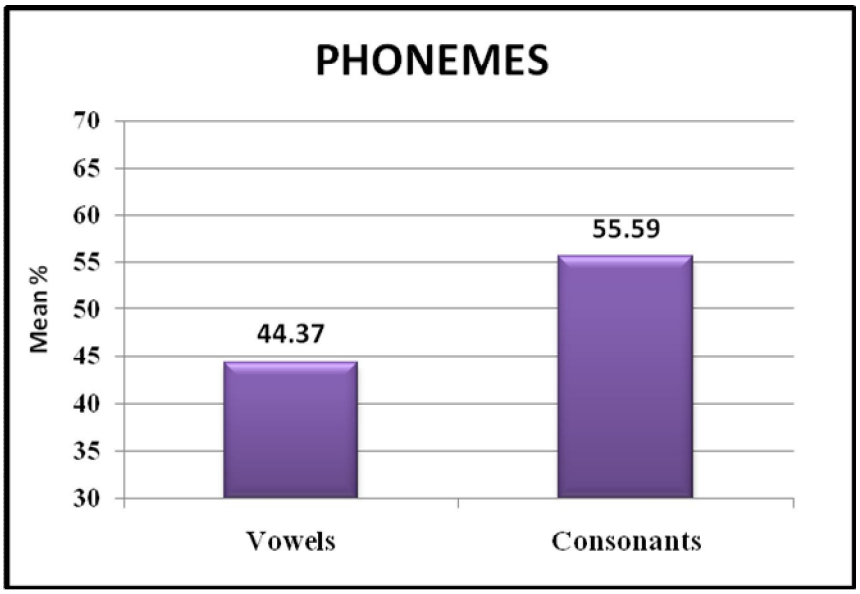

Figure 3: Mean percentage of vowels and consonants

Discussion

The present study aimed to obtain the database for frequency of occurrence of phonemes in conversational Kannada. A total of 74 individuals divided into 20 groups were involved in conversation, each group consisting of 3 to 5 individuals. 20 recorded conversations obtained were transcribed using IPA transcription. The topic of conversation varied across each recording session as to control the repeated occurrence of a particular word. The various studies which involved spoken materials considered lectures, interview with an individual for the data (Ghatage, 1994; Mines, Hanson & Shoup, 1978). But the present study incorporated an environment of natural conversation between 3 to 5 individuals in each session.

The present study included a large corpus of 2,68,625 phoneme occurrences as compared to previous studies using spoken data which were limited to 1,00,000 to 1,65,000 phoneme occurrences (Guirao & Jurado, 1990; Mines, Hanson & Shoup, 1978). The phoneme occurrences across conversation recordings ranged from 9,411 to 18,221 with an average of 13,431 phoneme occurrences. These variations in phoneme occurrences across recordings is possibly due to several reasons like the pauses between the conversation, and the rate of speech of each participant during the 30 min recording.

The results of the present study had several salient findings. 1,19,296 vowel occurrences and 1,49,269 consonant occurrences were obtained from 20 recordings. The results indicated that Kannada conversational speech consists of 55.59% consonants and 44.37% of vowels. This suggests that the consonants are used more than vowels in our day to

day conversation as they convey the meaning. Similar findings were observed in five out of nine languages studied by Yegerlehner and Voegelin (1957). The results of Delattre's (as cited in Edwards, 2003) study in English and Jayaram (1985) in Kannada also showed that the consonants occurred more than vowels.

In the present study, the most frequently occurring phonemes in descending order are **/a/, /n/, /i/, /a:/, /r/, /d/, /e/, /l/, /t/, /u/.** These frequently occurring phonemes constituted 64.5% of the total data whereas in English it was 47% of the total data (Mines, Hanson & Shoup, 1978). Jayaram (1985) also found that ten most frequently occurring consonant constituted about 64% of the total data in Kannada. The present study indicates that the vowels **/a/, /i/, /a:/, /e/** and **/u/** which occurred most frequently constitute 36.5% of the total vowel data (44.37%) whereas consonants **/n/, /r/, /d/, /l/, /t/** constituted 27.9% of the total consonantal data (55.59%). The order of the frequently occurring phonemes in the present study has some differences from the reports of Ramakrishna (1962) in written Kannada. Though he found that vowel /**a**/ occurred most frequently, the predominantly occurring consonants reported in order were /**r**/**,** /**d**/, /**n**/ and **/t/.**

From the data of conversational Kannada in the present study, it is found that nasal **/n/** is the most frequently occurring consonant which constituted 7.59% of the total data followed by **/r/,** dentals **/d/, /l/** and **/t/**. The findings of Ramakrishna (1962) in written Kannada showed trill **/r/** as the most frequently occuring consonant followed by dental stop **/d/,** nasal **/n/** indicating a difference in the sequence of the frequently occurring phonemes across spoken and written materials. Similar differences have been reported in written English by Sandoval et al (2008) that /**s**/ was the most frequently occurring consonant

whereas Delattre reported it to be /**t**/ as the mostly frequently occurring consonant. The results of these studies clearly indicate that the frequency of use of consonants varies across languages and also across modes, that is written and spoken contexts. Study by Manjula, et al (ARF project, 2011-12) in Kannada, including both written as well as spoken data showed that the most frequently occurring vowel was /a/ followed by /i/. Among consonants /n/, /r/, /t/, /l/ were the most frequently occurring phonemes similar to the present study.

In the present study consonant **/s/** occurred for 1.75% of the total data whereas its frequency was higher in the previous studies by Ramakrishna (1962) and Jayaram (1985) which was 2.2% and 2.6% respectively. The consonants **/m/, /v/, /ɖ/** occurred about 2.2% to 2.8% of the total data whereas consonants **/h/, /ʈ/, /p/** occurred about 1.2% to 1.7% of the total data. These results are in consonance with Ramakrishna (1962) and Jayaram (1985) who reported similar frequency of occurrence of these phonemes.

The aspirated consonants were rarely seen in the present study. Aspirated **/pʰ/, /ʧʰ/, /ʤʰ/** did not occur in any of the recorded conversations whereas /ɖʰ/, **/kʰ/, /ʈʰ/** occurred in 0.05-0.14% of the total data. This is a significant finding of the present study as the other studies by Ramakrishna (1962), Ranganatha (1982), Jayaram (1985) showed greater occurrence (0.5% to 1.35%) of these aspirated consonants in written context. This indicates that although the aspiration is present in the Kannada phoneme system, it is sparingly used in conversation.

Considering the vowels in the present study, vowel **/a/** was the most frequently occurring vowel followed by **/i/, /a:/, /e/** and **/u/.** Vowel /**a**/ occurred 14.5% of the total data. Similar findings were observed by Jayaram (1985) and Ramakrishna (1962) in written

Kannada who showed that vowel **/a/** occurred 19% and 18.7% respectively of the total data. Although **/a/** occurred most frequently in the study by Ranganatha (1982) in written Kannada the order was relatively different. In his study, short vowels **/a/, /i/** and /**u**/ occurred more frequently than any other phonemes.

Diphthongs **/au/** (0.3%) and **/ai/** (0.28%) occurred 0.58% of the total data. But Ramakrishna's (1962) study reported diphthong occurrence of 0.1%. The present study yielded similar order and percentage of phonemes in all the twenty recordings, i.e. vowel /**a/** was the most frequently occurring phoneme followed by nasal **/n/,** and short vowel **/i/**. The relatively less occurring phonemes were dental **/t/** and vowel **/u/**. This indicates that there was a consistency in the occurrence of phonemes albeit the change in the topic of conversations across different recording sessions.

Overall from the present study, it is found that the frequency of occurrence of consonants is more than that of vowels which infers that consonants are being used more in daily conversation. The frequency of occurrence of phonemes also varies across different languages. Interestingly, the order of frequency of occurrence of phonemes varies across spoken and written data within the same language.

Summary and Conclusions

The present study aimed to obtain the frequency of occurrence of various phonemes in spoken Kannada using conversation samples. The frequency of occurrence of phoneme data is crucial to understand the language structure and also has wide applications in audiology and speech language pathology. Earlier studies on frequency of occurrence of phonemes in various Indian languages like Hindi, Malayalam, Oriya, Kannada etc were based on written source of materials.

In the present study, spoken data in Kannada was collected through conversation samples which were recorded in 20 separate sessions, with each session consisting of 3 to 5 participants. 74 fluent native speakers of Mysore dialect of Kannada in the age range of 20 to 50 years were selected for the study and were divided into 20 groups. Conversation sample of each of the groups were recorded separately for 25-30 minutes using a digital recorder (Olympus WS 100). Participants were not given any prior topic for the conversation. Any current topic of interest was initiated by the participants themselves. The participants were instructed to avoid words from other languages, however they were not restricted from using commonly used loan English words (E.g.: Bus, ticket, phone, car etc).

The conversation samples obtained were transcribed using International Phonetic Alphabet (IPA) transcription. The raw data obtained were analyzed using the software SALT (Systematic Analysis of Language Transcripts) RV version 9 for the frequency count. The database of Kannada phoneme list was saved in the editable standard wordlist of SALT software. The SALT software compares the loaded transcribed phoneme file and the database

and provides the phoneme count. Inter judge and intra judge reliability of phonetic transcription was evaluated for 10% of the recorded samples.

The results obtained from the SALT software were tabulated and were subjected to statistical analysis. Mean and standard deviation of frequency of phonemes of all the twenty samples were obtained. A total of 2,68,625 phoneme counts were obtained from 20 recorded conversation samples. 44.37% of the total data consisted of vowels whereas consonants accounted for 55.3% of the total phoneme corpus. The results show that vowel **/a/** was the most frequently occurring phoneme among vowels and nasal dental **/n/** among consonants followed by **/i/, /a:/, /r/, /d/, /e/, /l/, /t/** and **/u/** in Kannada. Phonemes **/h/, /s/, /p/, /ʧ/, /ʤ/, /ʃ/, /ʈ/, /ɭ/, /j/** occurred less frequently (less than 2%) in conversational samples. The aspirated consonants **/pʰ/, /ʈʰ/, /ʤʰ/, /ʧʰ/** were not present in any of the 20 conversation samples. Earlier studies in written Kannada showed trill **/r/** as the most frequently occuring consonant followed by dental **/d/.** Comparing these data with the present study, indicates that the frequency of occurrence of consonants varies across modes i.e. written and spoken contexts.

Vowel **/a/** and nasal consonant **/n/** occurred most frequently in all the 20 recordings. But the order of frequency of occurrence of phonemes across 20 recording sessions had some variations. Following nasal **/n/**, vowel **/i/** occurred most frequently in most of the recording sessions. Few of the recording sessions had long vowel **/a:/** as the most frequently occurring phoneme followed by nasal **/n/**. Across the 20 recording sessions, variations in frequency of occurrence of phonemes were mainly observed for phonemes **/r/, /t/, /d/** and **/u/**.

The information on the frequency of phonemes in Kannada obtained in the current study will aid 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 the area of linguistics, speech synthesis and automatic speech recognition tasks.

References

Bhagwat, S. V. (1961). *Phonemic frequencies in Marathi and their relation to devising a speed-script*. Pune: Deccan College

Bloomfield, L. (1914). *Introduction to the study of Language*. New York: Henry Holt.

Campbell, R. A. (1965). Discrimination and word test difficulty. *Journal of Speech and Hearing Research,* 8, 13–22.

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.

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, A. M. (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., & Borzone de Manrique, A. M. (1972). Fonemas, silabas y palabras del espanol de Buenos Aires. Filologia, 15, 135-165.

Guirao, M., & Garcia Jurado, M. A. (1990). Frequency of occurrence of phonemes in American Spanish. *Revue Quebecoise de Linguistique, 19*, 135-149.

Jayaram (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.

Kelkar, A. R. (1994). *Phonemic and Morphophonemic frequency in Oriya*. Mysore: CIIL.

Malecot, A. (1974). Frequency of occurrence of French phonemes and consonant clusters. *Phonetica, 29,* 158-170.

Manjula, P., Geetha, C., Sharath Kumar, K.S. & Antony, J. *Development of PB word list in Kannada.* An ARF Project (2011-12). Report to be submitted

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.

Pei, M. A., & Gaynor, F. (1954). *A dictionary of linguistics*. New York: Philosophical Library.

Ramakrishna, B. S. (1962). *Some aspects of the relative efficiencies of Indian languages: A study from information theory point of view.* Bangalore: Indian Institue of Science.

Ranganatha, M. R. (1982). *Morphophonemic analysis of the Kannada language*. Mysore: CIIL.

Renwick (2011). Phoneme type frequency in Romanian. *In proceedings of the 34th Annual Penn working papers in Linguistics, 17*, 194-204.

Roberts (1965). *A statistical linguistic analysis of American English*. The Hague: Mouton.

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 Castalian 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).

Sreedevi, N., Nair, S., & Vikas (in press). Frequency of phonemes in Kannada: A preliminary study. Submitted to the *Journal of All India Institute of Speech and Hearing.*

Sridhar, S. N. (2007). *Modern Kannada Grammar*. New Delhi: Manohar.

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 methods Instrum comput, 36(3)*, 531-47.

Trager, G. L., & Smith, H. L. (1951). *An outline of English structure*. Studies in Linguistics occasional papers (No. 3). Norman, OK: Battenberg Press.

Upadhyaya (1976). *A Comparative Study of Kannada Dialects: Bellary, Gulbarga, Kumar, and Nanjangud Dialects.* Mysore: Prasaranga, University of Mysore.

Webster, N., & Gove, P. B. (1961). *Webster's third new International dictionary of the English language.* London : G. Bell.

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

Zurinskas, T. E. (2002). Analysis of English in USA Accent Developing a Truespel Leengwuprint. Retrieved from http://lincs.ed.gov/nifl-assessment/2002/0093.html.

==========

Appendix 1

Loan English words considered in analysis

Bank	Gas	Paper (Newspaper)
Bike	Gate	Pass
Book	Helicopter	Pencil
Bucket	Insurance	Permanent
Car	Kilometer	Petrol bunk
Circle	Madam	Phone
Coffee	Marks	Reservation
Company	Medical	Room
Degree	Mobile	Round
Diary	News	Rush
Doctor	Off	School
Engineer	Office	Scooter
Free	Operation	Seat
Fridge	Pant	Shoe
Time	Waste	Sir
Urgent		T.V (television)

Appendix 2

Frequency of occurrence of phonemes in Mysore dialect of spoken Kannada in rank order

Phoneme	Percentage	Phoneme	Percentage
/a/	14.57 %	/j/	1.24 (0.22)
/n/	7.59 %	/p/	1.22 (0.34)
/i/	6.70 %	/o:/	1.20 (0.28)
/a:/	5.66 %	/i:/	1.09 %
/r/	5.53 %	/u:/	0.55 %
/d/	5.35 %	/ʤ/	0.53 %
/e/	5.27 %	/ʃ/	0.42 %
/l/	4.98 %	/ ə/	0.35 %
/t/	4.54 %	/ʧ/	0.34 %
/k/	3.49 %	/au/	0.33 %
/g/	3.30 %	/ɳ/	0.29 %
/m/	2.76 %	/ai/	0.28 %
/v/	2.57 %	/ ɖʰ/	0.20 %
/ɖ/	2.24 %	/bʰ/	0.09 %
/e:/	2.23 %	/kʰ/	0.05 %
/b/	2.08 %	/tʰ/	0.05 %
/o/	1.85 %	/ŋ/	0
/h/	1.75 %	/pʰ/	0
/s/	1.75 %	/ʈʰ/,	0
/ʈ/	1.53 (0.41)	/ ʤʰ/	0

/ʃ/	1.33 (0.22)	/ʧʰ/	0