A SCREENING TEST OF ARTICULATION IN KANNADA USING DYNAMIC TIME WARPING (DTW) – A PROTOTYPE

Akshay M. Mendhakar

Register No.: 15SLP001

A Dissertation Submitted in Part Fulfilment of Degree of Master of Science

(Speech-Language Pathology)

University Of Mysore

Mysore

ALL INDIA INSTITUTE OF SPEECH AND HEARING

MANASAGANGOTHRI, MYSORE-570 006

May, 2017

CERTIFICATE

This is to certify that this dissertation entitled **"A *screening test of articulation in Kannada using Dynamic Time Warping (DTW) – A prototype*"** is a bonafide work submitted in part fulfilment for degree of Master of Science (Speech-Language Pathology) of the student Registration Number: 15SLP001. This has been carried out under the guidance of a faculty of this institute and has not been submitted earlier to any other University for the award of any other Diploma or Degree.

Mysore
May, 2017

Prof. S.R. Savithri
Director
All India Institute Of Speech and Hearing
Manasagangothri, Mysore-570006

CERTIFICATE

This is to certify that this dissertation entitled **"A *screening test of articulation in Kannada using Dynamic Time Warping (DTW) – A prototype"*** has been prepared under my supervision and guidance. It is also being certified that this dissertation has not been submitted earlier to any other University for the award of any other Diploma or Degree.

Mysore
May, 2017

Guide
Prof. S.R. Savithri
Professor in Speech Sciences
Dept of Speech-Language Sciences
All India Institute of Speech and Hearing
Manasagangothri, Mysore-570006

DECLARATION

This is to certify that this dissertation entitled **"A *screening test of articulation in Kannada using Dynamic Time Warping (DTW) – A prototype*"** is the result of my own study under the guidance of Prof S.R. Savithri, Director, All India Institute of Speech and Hearing, Mysore, and has not been submitted earlier to any other University for the award of any other Diploma or Degree.

Mysore, ***Registration No. 15SLP001***

May, 2017

Dedicated to My Parents...

ACKNOWLEDGEMENTS

It gives me an immense pleasure in acknowledging the reason for my successful endeavour; I would like to sincerely thank everyone who has contributed in building my dissertation.

It all started with that quote,

> People who are really serious about software should make their own hardware.
>
> \- Alan kay

This saying has been a game changer in the field on technology; Apple took it as design your own software and hardware to be the number one.
Well, I just understood the following...... in order for me to excel, I need software skills along with SLP skills too.

Special people to thank my parents, "Mohan & Asha; well guys you are just awesome; none of this WOULD have been possible without your immense hard work and sacrifices".

My cool brother Vivek, "Thanks for all your support man!!".

Thank you world's special grannies; pushpa & saroja

Firstly a huge THANK YOU!! To Prof. S. R. Savithri, "Madam you are a true inspiration and the source of immense knowledge and patience, every encounter with you has been a learning experience."

Secondly, Dr. N. Sreedevi, "Thank you for believing in me and encouraging me to learn and execute my software skills, I myself had no idea that I would be capable of something like this!!

Special thanks to Dr. Vasanthalaxmi for her valuable statistical inputs and guidance. "Thank you madam"!!

A huge gratitude to all the academic and clinical staff who have inculcated the knowledge in me. “ Thank you” !!

Thanks to all those participants that helped me build my database on time. “Thank you everybody”

My gratitude and thanks to all my teachers in St. Matthias school, JSS laxmipuram high school, Marrimalappa pre university college & AIISH for making me who I am!

All those crazy friends, all those crazy experiences in my life; AIISH has truly been a learning experience that has helped me in shaping myself as a person.

A very special thank you- Eliza,Charan, Heena, Pritismita & Shafi for always being there for me!!

-------X-------

TABLE OF CONTENTS

Chapter No.	Contents	Page No.
	Table of contents	viii
	List of tables	ix
	List of figures	x - xi
I	Introduction	1 - 4
II	Review of literature	5 - 40
III	Method	41- 47
IV	Results	48 - 56
V	Discussion	57 - 58
VI	Summary and conclusions	59 - 62
	References	63 - 71
	Appendix	72 - 76

LIST OF TABLES

Sl. No.	Title	Page No.
1.	Overview of articulatory milestones in typically developing children.	09
2.	Selected ten words from KDPAT targeting consonantal productions	41
3.	Distribution of Target words under correct, substitution, omission and other errors.	48

LIST OF FIGURES

Sl. No.	Title	Page No.
1.	Mel Frequency spaced Filter banks	30
2.	Log spectrum of a vowel segment in Hz scale (top) and in Mel Scale (bottom)	31
3.	(a) Magnitude squared spectrum in mel scale is multiplied by a triangular bandpass filter response. (b) A number of triangular filters constitute a filter bank. The filters are equally spaced in the mel frequency scale.	32
4.	Mel Frequency Cepstral Coefficient Block diagram	33
5.	Shows the time alignment between the test and the training pattern using DTW.	34
6.	Dynamic Time Warping of Input sample and stored template.	36
7.	Depicting an orthographic representation of the target word.	42
8	Pictographic representation of Module 01 of Computerized Screening Articulation Test	46
9	Pictographic representation of Module 02 of Computerized Screening Articulation Test	47
10	Shows the opening page of CAT's	49
11	Shows the Calibrate page of CAT's	50
12	Shows the Start Articulo page of CAT's	50

13	Shows the Module 01 page of CAT’s	51
14	Shows the Selection of the speaker page of CAT’s	52
15	Shows a sample word from Module 02 page of CAT’s	53
16	Shows a test completion page of CAT’s	54
17	Shows the results of perceptual rating and Developed tool results.	55

Chapter I

Introduction

"If the tongue had not been framed for articulation, man would still be a beast in the forest."

\- Ralph Waldo Emerson

\-

Articulation, as the word applies to speech, is the production of sounds with identifiable acoustic characteristics. Sound or air is transmitted from the larynx to the resonators, which select certain frequencies of the complex tones they receive for amplification and other frequencies for dampening (Flanagan, 2013). The articulators (tongue, lips, teeth, velum and others) are specialized structures that alter the sizes, shapes and resonation (Gordon-Brannan & Weiss, 2007). For over a century, acquisition of speech similar to that of adults in children has ensnared speech language pathologists i.e. the development of articulation has drawn great deal of attention in the field of Speech Pathology (Geers, 2002; Sander, 1972). It is generally accepted that the ability to produce each of the sounds of a language is acquired at a particular age (Oxford, 1991).

Articulation entails all the processes involved in planning and execution of overlapping sequential gestures to result in connected speech (Fey, 1992). This implies firstly that learning of articulatory skills is a developmental process in which children move from acquisition of gross to fine articulatory movements in a precise and rapid manner. Secondly that errors observed during articulation are a resultant of relative peripheral disturbances of articulatory process (Shriberg & Kwiatkowski, 1982). Therefore, articulation can be considered as a gradually developing skill involving the peripheral motor processes (And & Dodd, 1996).

Errors in articulation are typically classified based on the child's age, highlighting the stages of developmental processes (Dodd & Iacano, 1989). Younger children exhibit errors as they are in early stages of this development but whereas older children are at later stages of this development or have crossed these processes. Based on the child's age certain articulatory errors may be considered to be usual (age appropriate errors) or unusual (non age- appropriate errors). Articulation errors are also viewed as failure on the part of a speaker to perceive the significant contrast between the standard sound and the sound which s/he produces (Gierut, 1998). A child can exhibit a delay in articulation when the acquisition of speech sounds follow a sequence but the developmental errors persist beyond the age as expected (e.g. when a 4 year old continues to say /ta:r/ for car). A child is said to have an articulation disorder when his/her speech sound error and/or sound acquisition sequence digress from those seen in typically developing children. Misarticulations occur normally during the early stages of speech development (Spriestersbach & Curtis, 1951). Thus, when some articulation errors occur at certain age levels, the child is not considered to have an articulation disorder. Rather use of such articulation patterns is characteristic of normal phonologic acquisition. The deviances in articulation could be due to organic factors, emotional conflicts, acoustic and perceptual deficiencies, difficulties in phonetic discrimination, poor motor coordination, poor model or it may be functional (Anthony, Aghara, Dunkelberger, Anthony, Wouldiams, Zhang, 2011; Nober, 1967; Sherman & Geith, 1967).

Articulatory disorders are the most frequently occurring among all types of speech disorders (Law, Boyle, Harris, Harkness, & Nye, 2000). They are a group of disorders that are highly variable in terms of its nature and onset. Different individuals exhibit a variety of articulatory errors, making the assessment challenging task for a Speech-Language Pathologist. The nature of articulatory errors requires a descriptive and perceptual approach which involves description of an individual's speech sound production and relating it to normal or standard in that specific language and community (Gierut, 1998). A thorough assessment is required to arrive at an accurate diagnosis, identification of etiology and a foundation for intervention.

In order to evaluate the articulation of these patients, tests of articulation are essential. The test of articulation is a basic tool of the speech pathologist. Articulation tests are typically designed to (a) evaluate whether an individual's speech sound system is sufficiently different from normal development to require intervention, (b) establish a treatment direction and targets that can be used by the client, (c) comment on prognosis in relation to the phonological skills with / without therapy, and (d) monitor phonological changes across time (Baker, 2012).

Traditionally, an articulation test is a technique employed to measure the general phonemic capabilities of an individual (Van Riper & Irwin; 1958). Articulation tests can be used in screening, diagnostic, predicting articulation disorders or deep testing. Over the years, many investigators have developed and established norms using various articulation tests.

The purposes of articulation test vary and hence the nature and scope of the articulation test inventory varies (Kent, Miolo, & Bloedel, 1994). If the purpose is to assess the general adequacy of articulation in order to determine whether a child would need speech correction, a screening test can be used. If a detailed description and analysis of articulation is desired in order to determine the direction the speech correction should take, a diagnostic test is needed. The value of the information obtained from articulation testing is related both to the precision of the instrument and to the sophistication of the examiner in its administration and interpretation. A descriptive diagnostic approach is time consuming and when valued over the case load, the traditional diagnostic approaches used for assessment of articulatory errors sound obsolete (Deepa, 2010). Hence, a screening objective measure that could capture the multiplicity and error patterns of individuals with articulatory disorders is essential in the field of Speech-Language Pathology. Various articulation tests have been developed based on specific languages to represent their own phonological systems and standardized accordingly such as Wellman , Case , Mengert, and Bradbury 1931; Poole, 1934; Templin, 1957; Sander, 1972; Prather, Hedrick, & Kern, 1975; Arlt & Goodban, 1976; Smit, Hand, Freilinger, Bernthal & Bird, 1990 in English; Babu, Ratna, and Bettagiri 1972 in Kannada; Usha, 1986 in Tamil; Padmaja 1988 in Telugu; Banik, 1988 in Bengali and Maya, 1990 in Malayalam.

Even though the above tests have been considered as gold standard tests in day to day clinical assessment, various recent studies by Roberts, Burchinal, & Foot, (1990); Bharathy, (2001); Rahul, (2006); Sreedevi and Shilpashree (2008); Deepa, (2010) have pointed its drawbacks and associated the reason as children of present generation are more exposed to different environments at a very early age due to advancement in the technology, education, nurture, awareness, and increased speech stimulation. Therefore, it is very important to re-standardize and validate the test according to the present scenario.

Further, the effects of computers and other technology in the field of communication sciences are gaining more popularity day by day (Chen & Huang, 2007; Maier, 2009). Efforts to come up with computerized tools for evaluation of phonological processes have been tried out by various researchers; however these lack ease of use and flexibility (Sreedevi, John, & Chandran, 2013). Hence, there is need for automatic and interactive tool for assessment of articulation. In this context, the present study developed a Computerized Screening Test of Articulation in Kannada[1]. Specifically the test used a database to compare the test word with the help of Dynamic Time Warping.

The specific objectives of the study were to (a) build the Computerized Screening Articulation Test using Matlab®, and (b) investigate the effectiveness of the developed tool by preliminary analyses.

[[1]*Kannada is a Dravidian language spoken predominantly by Kannada people in South India, mainly in the state of Karnataka, and by linguistic minorities in the states of Andhra Pradesh, Telangana, Tamil Nadu, Maharashtra, Kerala, and Goa. The language has roughly 40 million native speakers who are called Kannadigas (Kannaḍigaru), and a total of 50.8 million speakers according to a 2001 census. It is one of the scheduled languages of India and the official and administrative language of the state of Karnataka. Retrived from* https://en.wikipedia.org/wiki/Kannada].

Chapter II

Review of literature

An extensive review has been provided under the following headings:

1. Articulation and normal aspects of articulation.
2. Development of articulation

   Western studies on development of articulation.

   Indian studies in development of articulation
3. Assessment of articulation
4. Computerized tools
5. Speech recognition and Mel frequency Cepstrum Coefficient
6. Dynamic time warping.

1. Articulation and normal aspects of articulation

Articulatory skills of an individual acts a medium to convey a person's thoughts, his/ her language content, what s/he conceptualizes and his/ her views through sounds, words, phrases and sentences. Articulation may be defined as the process by which sounds are produced in the mouth through movements, adjustments and contacts of the lips, tongue, mandible and soft palate associated with respiration and phonation (Wardill and Whillis, 1936). Articulation is regarded as a phenomenon where constant flow and flux of neuromotor adjustments takes in sync with acoustic signals which are a resultant of modification of the outgoing breath stream and voice through alterations in sizes and shapes of resonators.

Fey (1992) views articulation as an array of motoric movements and the way articulators are placed during the production of a speech sound. He postulates that speech is a resultant of a motor process solely, in which sequential motor planning and execution of simultaneous gestures takes place.

Based on this definition it can be inferred that acquisition of articulatory skills is a developmental process in which sequential gestures of the articulators become more rapid and precise, suggesting articulation is a resultant of a motor learning processes. Thus, it indicates that articulatory errors are caused when the peripheral motor processes responsible for articulation is impaired with no errors in the central language capabilities of an individual.

Other definitions of articulation emphasize on the importance of higher order cognitive-linguistic abilities of an individual and how the peripheral and central processes are intertwined for articulatory productions. Considering the above, Gordon-Brannan & Weiss (2007) summarized the final act of articulation to be contributed by many events - first, the vocal tract, the articulators and the intact nervous system enable an individual to execute the sensory (auditory, tactile, kinesthetic, proprioceptive) and the motor functions essential for the controlled movement. Second, the cognitive-linguistic component conceptualizes the need to "say something". This thought is processed linguistically wherein the phonological rules of one's language are applied and based on this. The phonemic elements are selected and ordered depending on the semantic lexicon, the syntax and the pragmatic context. Third, the sensorimotor component that involves motor programming and learning of the actual sequences of physical movement of the articulators comes into play, following which, the vocal and oral cavity get activated transferring acoustic vibrations to meaningful utterances achieved by the manipulation of the oral articulatory structures -.

An individual is said to have articulation disorder if s/he has difficulties in motor productions necessary for speech sound productions (Elbert & Gicrut 1986). Articulation disorder is a sub-category of speech difficulties characterized by erroneous productions categorized as substitutions, omissions, additions and distortions. These erroneous productions are considered to impair the overall intelligibility of speech. Another category of articulatory errors could happen when the comprehension impairment of sound system that governs speech sound productions. Such a class of errors is defined as phonological disorder and is thought to be a sub-category of language disorders as opposed to articulatory errors, a subcategory of speech disorders. Cognitively or linguistically based articulatory errors (production of sound is present but the individual is unable to use the sound in appropriate contexts) are referred to as phonological disorders.

Articulation disorder can be secondary either to organic etiologies such as hearing impairment, Cleft palate, developmental dysarthria and other syndromic conditions or Functional articulation disorder which can exists when no apparent cause is noted yet deficiencies in peripheral motor processes is seen (Bauman Waengler, 2000).

The effect of an articulatory disorder may not be readily evident to the listener but they could have extensive effects on the person's social, emotional, interpersonal, and academic and other such aspects. Because articulation is deliberate, so apparent visibly and auditorily, a slight disturbance provokes judgments and penalties by the listeners, sometimes so severe to an everlasting socio-psychological impact on the speaker. Van Riper and Erickson (1996) have clearly pointed out that right from childhood; a person with articulatory disturbances could be victimized with comments, derision, loneliness, labeling, harassment and ultimately frustration. Over the years, this frustration of not being able to communicate as precisely and as efficiently as the peer group, and as the insight towards the negative attitudes of the society grows, the person goes into a stage where s/he sees himself inferior to others, sometimes avoiding speaking situations or avoiding to speak that particular word which s/he misarticulates. Poverty in grades, depletion in personality and emergence of disruptive behaviors may become an everyday issue. Hummel and Prizant (1993) report a 50-70 % co-occurrence of speech & language problems, behavioural problems, and emotional problems in children/adolescents in various set-ups like schools, community speech and language clinics, and in-patient as well as day treatment psychiatric settings. The exact relation between socio-emotional impairment and communication deviancies might not be clear but, it is evident that many emotional and behavioural disturbances are manifestations of the communication difficulties or disorders. If the emotional-behavioural issues are because of communication disorders, it is the role of a SLP to alleviate such disorders in early childhood itself, provided the child has access to early intervention programs. If such issues are kept in mind, a child may grow up to a confident and efficient adult rather than growing up as an efficient yet less confident adult living life full of contempt and embarrassment.

A clinician evaluating the articulatory skills of an individual with articulatory deficits must be aware of the developmental trend of speech sounds, in terms of its acquisition and mastery. The earliest age at which a child is correctly articulating a speech sound holds to be an important milestone in speech sound acquisition. However, it is practically impossible to ascertain when a particular child uttered a particular speech sound for the first time. So the concept of speech acquisition came into picture. A child is said to have acquired a particular speech sound if s/he is uttering the speech sound correctly most of the times (Sander, 1972). Therefore a thorough knowledge regarding the concept of articulatory acquisition and articulatory mastery is a must for a SLP.

2. Development of articulation

As emphasized by various authors the learning of articulatory skills is a developmental process involving the gradual acquisition of the ability to move the articulators in a precise and rapid manner (Fey, 1992). The most significant development of articulation takes occurs between 1-3 years of age. The following table summarizes various milestones of articulatory development in a typically developing child.

Stage 01	Stage 02	Stage 03	Stage 04
I (0-12 months)	II (12- 18 months)	III (18- 36/42 months)	IV (42-84 months)
Marked by production of first word utterance.	Acquisition of first 50 words These toddlers have developing cognitive system with variant mental representations of speech noted. Highly variant speech sound errors. They may also produce adult like words with similar structural complexity differently. In some children, gradual deterioration of accurate productions occurs such that the erroneous speech sound productions are produced as consistently as the other spoken words indicating that the child's linguistic rules are not fully established.	Children in this age range produce many articulatory errors which follow consistent patterns or processes. Error processes are due to simplification of adult forms of words. This is the stage when children acquire mastery over phonemes in English. Sequence of mastery in English: Vowels→Nasals →glides→stops→affricates→ fricatives. (Poole, 1934; Prather, Hedrick & Kern, 1975: Sander, 1972: Templin. 1957: Wellman, Case, Mengert & Bradbury. 1931). Children develop the concept that a change in the phonemic contrast can alter the meaning of an utterance during this period.	Remaining phonemes are mastered during this age group. Correct production of polysyllabic words and few rules of English morphosyntax are also mastered during this age. During this last stage, children continue to combine phonemes in different ways to alter the meaning in different context.

Table 1: Overview of articulatory milestones in typically developing children.

Western studies on development of articulation

Poole (1934) and Templin (1957) concluded that in early years, diphthongs, vowels, consonant elements, double consonant blends and triple consonant blends are learnt and produced with accuracy in the said order. Nasals are acquired first followed by stops, fricatives, combinations and semivowels with voiceless cognates produced more accurately than the voiced sounds. Overall, by the age of 8 years all speech sound productions are mastered.

Templin (1957) in his study on 480 children in the age range of 3-8 years concluded that at least 75% of the children mastered the speech sounds in initial, medial and final word positions tested at 176 elements in 3 positions and reported results in consensus with the findings of Wellman, Case, Mengert, and Bradbury (1931) and Poole (1934).

Sander (1972) criticized Templin's concept of age levels based on three position criteria being unrealistic and used a different approach to analyze the same data. He reassigned the sounds to age groups where they were able to correctly produce the said sound in two out of three word positions, his results were reported in two levels: I) the age at which 90% of the children produced the sounds correctly in two out of three word positions, and 2) the age at which 50% of the children produced the sounds correctly in two out of three word positions.

Stoel- Gammon (1985) carried out a descriptive study on the articulatory inventory of 34 children (19 boys and 15 girls) aged between 15 to 24 months and concluded the following: /b/, /d/ and /h/ were in the inventory of 50% of the children in initial word positions by 15 months of age. Nasals /m/ and /n/ were present in the inventory of 50% of the children by 18 months of age. The only phoneme at this age in final word position was /t/ in 50% of the children. /b/, /d/, /h/, /m/, /n/ and /t/ were present in initial position of the inventories of 50% of the children and in final position, only /t/ and /n/ were present in the inventories of 50% of the children at 21 months of age. Overall at the age of 24 months /b/, /d/, g/, /t/, /k/, /h/, /m/, /n/, /w/, /f/ and /s/ were in the initial position inventories and /p/, /t/, /k/, /n/, /r/ and /s/ were in the medial position inventories of 50% of children.

Smit, Hand, Freilinger, Bernthal and Bird (1990) reported the following results of 1049 children in an age range 2-9 years from Iowa and Nebraska as follows: The phonemes /m/, /n/, /h/, /p/, /f/, /w/ and /h/ were acquired by 3 years of age. Girls acquired sounds earlier than boys, although this reached statistical significance only by 6 years of age. The phonemes /l/, / tʃ/, /k/, /g/, /j/ /d/ and /t/ were acquired by 7 years of age. The phonemes / ŋg/ and /s/ were acquired by 7- 9 years.

Watson and Scukanec (1997) profiled the phonological abilities 2-3 year old children (3 months interval) and concluded that at 2.0 years, /p/, /t/, /k/, /b/, /d/, /m/, /n/, /s/, /h/, /w/ and /j/ were acquired in initial word position, and /p/, /t/, /k/, /m/, /n/, /s/, /z/ in final word position. By 2.3 years, phonemes /f/ and /l/ in initial position and /d/ in final position were also acquired. By 2.6 years, in initial position, /g/, / tʃ/, clusters /pw/ and /bw/, and in final position, /l/, /r/ clusters / ŋd/ and / tʃ/ were acquired. By 2.9 years, clusters /pï/ was acquired in initial position and /nt/ and /nz/ at final positions. By 3.0 years, / ð/ was added up to the existing phonetic repertoire along with clusters /st/ and /sp/.

McLeod, Doom and Reed (2001) in a meta analysis of seven decades on acquisition of consonant clusters in English concluded the following: The acquisition of consonant clusters is a gradual process and follows a typical developmental sequence. It is not an 'all-or-nothing' process. Two- year old children can produce consonant clusters one element consonant clusters and two element consonant clusters can be produced by two and three year old children respectively but these might not be similar to the ambient language. Consonant clusters containing plosives (e.g., /pl/, /kw/) are acquired earlier than clusters containing fricatives (e.g., /st/). Cluster reduction, epenthesis & coalescence are common in speech of younger children, while metathesis is rarely seen. Despite developmental trends noted individualistic variations do exist.

Dodd, Holm, Hua and Crossbie (2003) using the 90% criteria of speech sound acquisition and mastery, reported the following: By 3.0-3.5 years, plosives (/pt/, /t/, /k/, /b/, /d/, /g/), nasals (/m/, /n/, / ŋ/) and fricatives (/f/, /v/, /s/, /z/, /h/) are acquired. By 3.6-3.11 years, approximants, (/w/, /l/, /j/) are acquired. By 4.0-4.5 years, fricative /z/, affricate / tʃ/ and / dʒ/ are acquired. By 5.0-5.5 years, fricative / ʃ/ is acquired. By 7 years and above, fricative / θ/ and / ð/are acquired.

These above mentioned western studies cannot be generalized to Indian children while evaluating their articulatory skills during assessment and management. India being a multilingual and multicultural country, a child in Indian scenario is exposed to more than one language right from a very young age. Keeping these issues in mind, several studies have been conducted in the Indian context to investigate the age of acquisition and the age of mastery of speech sounds. Such studies have been embarked on in Kannada, Telugu, Malayalam, Tamil, Hindi, Bengali, Oriya, Marathi, Gujarati and various other languages.

Indian studies in development of articulation

The earliest effort to develop a test of articulation in Indian languages was made by Babu, Rathna and Bettagiri (1972) when they developed the test of articulation in Kannada which was not standardized. Sreedevi (1976) conducted an observational study on 4 children (2+years of age) where they were evaluated four times during the course of the study at equal intervals in terms of their age, between two subsequent recordings. She reported that acquisition of voiced to unvoiced distinction was acquired before aspirated to unaspirated distinction. By 2.6 years of age, distinction between long and short vowels was acquired. Stops acquired more fully than sibilants, trills and laterals. Among nasals, bilabial and alveolar nasals were acquired earlier than other nasals. Among the sibilants, the alveolar and palatal sibilants were acquired earlier than retroflex. Identical clusters were acquired earlier than non-identical clusters.

Following this, Banu (1977) made used the Kannada Articulation Test developed by Babu, Rathna and Bettagiri (1972) and standardized the same on 180 Kannada speaking children in the age range of 3-6.6 years and concluded that articulatory patterns varied significantly across different age groups. A gradual but definite change from age to age was observed. Fricative /h/ not acquired even by 6.6 years of age and no significant gender difference was noticed.

Various other studies to contemplate upon the phonological processes in Kannada speaking children have also been carried upon in the recent past. Sreedevi (2008) studied the articulatory skills of eight children aged between 1.6 to 2 years and reported retroflex fronting (substitution of retroflex sounds by easier denials) to be the most dominant phonological process during this age range along with several other processes like initial consonant deletion, vowel lowering, trill deletion, cluster reductions, etc. In another study, Sreedevi & Shilpashree (2008) studied the phonological processes occurring in children in the age range of 2.6 - 3 years and concluded that by this age, retroflex fronting is reduced and most of the phonemes are mastered by this age including fricatives and trills by the age of 3 years and draw future researchers the need to revise the existing articulatory norms in Kannada.

Almost after 3 decades that the KAT was standardized, Prathima (2009) studied the articulatory acquisition patterns of urban Kannada speaking children in the age range of 3.4 years. She concluded that vowels and most of the consonants except /r/, /h/, /ʃ/, retroflex /d/ and retroflex /l/ were acquired by 3- 3.6 years of age. Nasals, semivowels /v/ and /j/, affricates /tʃ/ and /dʒ/ were acquired by 90% of the children by the age of 3- 3.6 years. Glottal fricative /h/ was not mastered even by 75% of the children by 4 years of age. Dental /s/ (I and M positions) and palatal /ʃ/, (word initial position) were reported to be acquired by 90% of the children by 4 years of age. Retroflex /l/ was acquired by the age 3.6- 4 years in 90% of the girls and 84% of boys. Medial clusters (/ski/ and /kra) were acquired earlier than initial clusters in accordance with a previous study by Rupela and Manjula (2007) where in, medial clusters were the first to appear by the age of 18- 24 months.

Deepa (2010) revised and re-standardized the earlier Kannada Articulation Test (Babu, Rathna & Bettagiri, 1972) and developed a picture form test which is now called as Kannada Diagnostic- Picture Articulation Test (KD- PAT). She tested 240 urban Kannada speaking children in the age range of 2- 6 years and concluded that all bilabials. Labiodentals, dentals and velars were acquired by 2.6- 3 years and palatals by 1.6- 4 years. Glottal /h/ was not acquired even by 6 years of age. Retroflex /ḍ/ was acquired much earlier than /ṭ/. All the vowels and diphthongs were mastered by 90 % of the children by 2- 2.6 years and were achieved by 100% of the children by 3 years of age and all the stops were acquired by the age of 3- 3.6 years. Fricatives /s/ and /ʃ/ were 4- 4.6 years and 3.6- 4 years, respectively. Most of the clusters were acquired by 4- 4.6 years.

In Tamil language, Thirumalai (1972) studied the articulatory acquisition patient and reported that all the vowels and consonants (except /ʃ/, /l/ and /r/ were acquired by 3 years of age. Fricative /ʃ/ was not acquired even by 6 years of age. All the stops were acquired by 3 years of age and retroflex /t/ was acquired earlier than the established western norms; all the nasals were acquired by 3 years of age. Among laterals, /l/ was acquired by 3 years of age, but not consistently produced until 6 years of age and /r/ occurred in medial and final position by 5 years but not in initial position until 6 years which is late as compared to other languages.

Usha (1986) developed the Tamil Articulation Test (TAT) and standardized the same on 180 Tamil speaking children in the age range of 3- 6 years and concluded that there was a significant gender difference. All vowels and most of the consonants except /ʃ/, /l/ and /r/ were acquired by 3 years of age. Voiceless retroflex /ṭ/ acquired much earlier than the Western norms. All nasals were acquired by 3 years of age. Flap /r/ was not acquired even by 6 years of age which is significantly late than that indicated in other studies.

Padmaja (1988) developed the Test of Articulation and Discrimination in Telugu and studied the articulatory acquisition pattern in 160 Telugu speaking children in the age range of 2.5 to 4.5 years. She reported no significant gender differences. All vowels and most of the consonants except /r/, /s/, /ʃ/, /t/, /d/ and aspirated stops were acquired by 2.5 years of age. All nasals were acquired by 2.6 years of age. /s/, /r/, /l/, retroflex /ḷ/ and aspirated consonants were acquired by 3.3 years of age. Phonemes /t/, / ʃ / and clusters were acquired by 3.5 years of age

Neethipriya (2007) studied phonotactics and acquisition of clusters in typically developing Telugu speaking children in the age range of 3.6 years. The results indicated that medial clusters occurred with a frequency of 60- 70% in this age range and among the medial clusters, 30- 40 % were geminated clusters. Medial non- geminate clusters occurred with a frequency of 45- 55%. Nasal plus homorganic stops (/nt/, /nd/, /nk/, /mt/, etc) were predominantly observed in samples following fricatives plus plosives combinations. Three consonant clusters were rarest to occur in the age range of 3- 6 years.

Usha Rani (2010) tested four consonant clusters in Telugu in 2-3 year old children and concluded that none of the clusters were acquired with 75% accuracy even by 3 years of age in both the genders. However, all clusters crossed 60% by 3 years of age.

Sneha (2012) studied acquisition of clusters in 120 typically developing Telugu speaking children in the age range of 4-6 years and reported that medial clusters are acquired earlier than initial clusters; clusters were produced more correctly in word repetition task when compared to picture description task; the acquisition of clusters was a gradual process and followed a typical developmental sequence. In a repetition task, by 6 years of age, 3 word - initial and 6 - medial clusters were mastered whereas in picture description task, 4 word - initial and 8 - medial clusters reached the target criteria and in story telling task, 2 word - initial and 5 - medial clusters reached the target criterion.

Some studies have been conducted to embark on the articulatory acquisition in children in Malayalam language. One of such preliminary attempts was undertaken by Maya (1990) who developed and standardized the Malayalam Articulation Test. The results indicated that articulatory scores increased as age increased, but even at 7 years of age, 100% scores were not achieved. Females had higher articulatory scores than males. All the vowels were acquired by 3 years of age. Most of the consonants except fricative /s/, tap /r/, trill */r/,* lateral /l/ and aspirated phonemes, were acquired by 3 years of age. The acquisition was in the order of un-aspirated stops followed by fricatives, affricates and aspirated stops. Compared to other studies, the articulatory acquisition was earlier in Malayalam speaking children than non-speakers of Malayalam.

Divya (2010) studied the acquisition patterns in Malayalam speaking children in the age range of 2- 3 years and found that all vowels except /u/ and /u: / were acquired by 2.3 years. By 2.6 years, these two vowels reached 90% criteria. Unaspirated sounds were acquired earlier than aspirated sounds. Bilabials, labiodentals, dentals and velars were acquired earlier than alveolars, palatals, retroflex sounds and glottal. None of the clusters reached 90% criteria by 3 years of age.

Neenu (2011), Vipina (2011) and Vrinda (2011) revised and revalidated the Malayalam Diagnostic Articulation Test (Maya, 1990) in the age range of 3-4 years, 4-5 years and 5-6 years, respectively. They also added initial and medial clusters to the existing test. Neenu (2011) reported that all vowels arc acquired by 3-3.3 years of age; most consonants acquired 100 % criteria by 4 years of age; none of the clusters reached 90% criteria by 4 years of age but a few medial clusters reached 90% criteria by 3.3 years of age itself. Children acquired bilabials, labiodentals, denials and velars earlier when compared to alveolars, palatals, glottal and retroflex sounds. Unaspirated sounds were acquired earlier than aspirated sounds. Fricatives, trill /r/, lateral /l/, /ḷ/ and aspirated /k^h/, /b^h/, /d^h/ did not reach 100% criteria even by 4 years of age. Vipina (2011) reported that all aspirated phonemes and a few unaspirated phonemes like /ng/, / ḍ / and /h/ were not achieved by the age of 4-5 years. When compared to subjects of lower age groups (Neenu, 2011), more initial and medial clusters were achieved by children according to the 90% criteria.

Vrinda (2011) reported that the consonants not achieved even by 5.3 years included aspirated phonemes and dental / ḍ / and glottal /h/ in medial position. By 6 years of age, most of the initial as well as medial clusters met the 100% criteria.

Among the Indo- Aryan languages, articulation tests have been developed in Hindi, Marathi, Gujarati, Bengali, Oriya, etc., a few of which are summarized below;

Banik (1988) developed the diagnostic test of articulation in Bengali and reported existence of a definite pattern of acquisition of sounds; Earlier acquisition than English speaking children; Females superior than males in terms of articulatory abilities; All the vowels acquired by 2.5 years; Most of the consonants (except / dʒ/, /h/, /r/ and few consonant clusters) were acquired by 3 years of age; Most misarticulations were in terms of substitution or omission; addition and distortion were rarely seen.

Animesh (1991) constructed the Deep Test of Articulation in Bengali- Picture Form and concluded that articulation of speech sounds improve with age. On task analysis, it was observed that /r/ and / ḍ / were the most difficult sounds to produce. Ease of production was better in an environment of voiced consonants when compared to voiceless consonants. Trills and fricatives were the most difficult to be produced.

Deepa (1998) developed the Deep Test of Articulation in Hindi-Picture Form and administered the same on 120 Hindi speaking typically developing children in the age range of 3-7 years. The results indicated that articulation skills increase with age till 5 years. The environment with aspiration was most difficult to produce. By 5 years of age, children articulated all phonemes correctly. On item analysis, it was found that / ḍ/, /$ḍ^h$ / and / $ṭ^h$ / were the most difficult to articulate.

Panda (1991) developed the Test of Articulation in **Oriya** and administered it on 120 Oriya speaking children in the age range of 3-6 years. He found significant gender differences. All vowels and consonants except /s/, /r/ and aspirated stops were acquired by 3 years of age. Aspirated stops were acquired later (5-6 years) than other languages (except Malayalam). Order of acquisition follows the trend unaspirated stops followed by fricatives, aspirated stops and last to develop were affricates.

3 Assessment of articulation

Assessment of articulatory skills of an individual is a complex phenomenon which requires careful monitoring of various factors. Diagnosis of articulatory disorders involves, not only labeling the child's problems, but also understanding the child's problems, his/her strengths and weaknesses meticulously. Initial evaluation is carried out, mainly to determine if there is a problem? If so what is its causative factor? To streamline effective and potential treatment approaches (Haynes & Pindzola, 2004). Several factors must be kept in mind before initiating the process of assessment of articulation and phonological disorders. Factors such as awareness of the anatomy and physiology of the speech mechanism, phonetics, phonological development, dialectal variations etc are critical for performing a diagnostic evaluation.

The main purpose of an assessment of articulation or phonological disorders includes (Bernthal,Bankson & Flipson 2009) (a) description of the articulatory- phonological development and the current status of the individual, (b) determination whether the individual's speech is sufficiently deviant from that of normal and requires intervention, (c) identification of factors that relate to the presence or maintenance of the speech disorder, (d) determination of the course or treatment, (e) decision regarding prognostic judgments about an improvement in articulatory-phonological skills with and without intervention, and (f) monitoring changes in the articulatory- phonological performance over time. Assessment of phonological or articulation disorders usually begins with a detailed case history and ends with the administration of a standardized articulation test, its scoring and interpretation.

Bernthal, Bankson and Flipson (2009) postulated that a clinician employs several assessment instruments and sampling procedures when carrying out a phonological analysis *as* a single assessment procedure might not provide all that is required to make a diagnosis and, to ascertain the need and direction of effective intervention. A phonological evaluation typically includes collecting speech sample of varying length (e.g., syllables, woods, phrases), phonetic contexts, and as a response to various elicitation procedures (e.g., picture naming. imitation. conversation). This assortment of samples is frequently referred to as an assessment battery. The authors suggest a comprehensive phonological evaluation battery which include (a) collection of articulatory samples via methods like conversational speech sampling, single- word or citation form sampling, (b) stimulability Testing where the ability of an individual to repeat the target utterance correctly when provided with a model. Basically, it tests how well a client imitates in one or more phonetic contexts (isolation, syllables, words, phrases, etc.), sounds that were produced erroneously during testing, (c) contextual testing via deep test of articulation is pivotal in solving issues of inconsistency of errors noted, in children, and (d) selection of Phonological Assessment Instruments from screening, diagnostic or deep tests of articulation for formal articulation testing. The choice of assessment tool used will influence the amount of sample obtained for specific speech sounds (consonants, consonant clusters vowels, diphthongs etc) in various positions, the types of analysis chosen like phonetic and/or phoneme analysis of consonant and vowel sounds of the language, sound productions in initial, medial and final words positions, place, manner, voicing analysis, phonological process analysis, age appropriateness of phonological productions, transcription and Scoring Responses procedure varies from very simple correct/ incorrect in screening tests or detailed reporting like each type of articulatory or phonologic error (substitution, omission, distortion and addition), and analysis and interpretation of responses. The information about sound system errors can be analyzed using two ways (Stoel- Gammon & Dunn, 1985) - In **Independent Analysis**, the child's utterances are described without reference to an adult model. Rather, the child's productions are not described in terms of errors. The clinician makes a note of what the child can produce. In **Relational Analysis**, the child's speech sound productions are compared to the adult forms. In this method, a clinician can perform a traditional sound-by-sound analysis by classifying an error into (SODA) substitution/omission/distortion or an addition, a manner-place-voicing analysis, a distinctive feature analysis based on the presence or absence of certain features called as distinctive features, and a phonological process analysis which provides an overall description of the child's phonological system by targeting a phonological pattern rather a single speech sound.

Skahan, Watson and Lof (2007) inspected the assessment procedures employed by randomly selected SLPs to assess children with suspected Speech Sound Disorders (SSDs) using survey method. They concluded that most frequently SLPs used commercially available articulation tests, estimated their intelligibility, assessed their stimulability and conducted an auditory screening to ascertain the intactness of auditory perceptual skills.

Among the various types of articulation assessment procedures, the traditional procedure is the most common. It holds that each phoneme of a language must be tested in the initial, medial and final positions of words. Typically, these words are elicited from the client by the help of pictures, word lists, sentences, or conventional sampling.

Several studies in the past have indicated that children tend to produce more singleton consonants correctly in single-word sampling contexts than in connected speech on traditional testing from the viewpoint of articulation as opposed to phonological process analysis where no significant difference was observed between single- word sampling and connected speech analysis. Mostly, in developmental phonology, single word articulation tests are being used since many decades, where responses are elicited as single words, called as citing, more than conversational sampling even, though many differences exist in both the methods of data elicitation (Morrison & Shriberg, 1992).

The same authors reviewed literature on methods of articulation sampling as a part of their study "Articulation Testing versus Conversational Speech Sampling" and stated that sample sizes varied from 1 to 240 and the children's age varied from under 3 to above 12 years. Most studies which have used citation as the mode of response elicitation have used spontaneous rather than imitative responses. Since, single word articulation tests use standard picture stimuli for each subject, the variability reduces. However, even the size and colour of the picture stimulus is shown to produce more variance in the results (Bernthal, Grossman & Goll, 1989).

Wolk and Meisler (1998) compared two methods of phonological assessment: conversation and picture naming. They reported that although both methods of speech sample elicitation are useful clinical tools for SLPs, picture naming taps a child's phonological system more deeply and extensively, and represent the index of phonological abilities in a better way. In the context of severity of involvement, more articulatory errors were observed in talking compared to citing, but severity ratings were similar or lesser when using citation forms (Andrews & Fey, 1986; Dubois & Bernthal, 1978; Faircloth & Faircloth, 1970, Klein, 1984). Although citation form (single word elicitation) testing might not provide overall information about a child's speech sound production system when compared to connected speech sampling (Andrews & Fey, 1986; Bernhardt & Holdgrafer. 2001a; Morrison & Shriherg, 1992) several disadvantages of connected speech sampling compel the SLPs to rely on single- word articulation tests not only to identify the children having speech sound difficulties (SSDs) but also to analyze productions and to determine goals (Hodson, Scherz & Strattman. 2002; Khan, 2002; Tyler & Tolbert, 2002).

Several standardized tests are available for the purpose of articulatory analysis in various languages. A few of the traditional articulation tests in English are the Arizona Articulation Proficiency Scale, Third Revision (AAPS- 3; Fudala, 2000), the Cioldman-Fristoe Test of Articulation, Second Edition (G-FTA-2; Goldman & Fristoe; 2000), the Photo Articulation Test, Third Edition (PAT-3; Lippke, Dickey, Selmer & Soder, 1997), the Templin- Darley Test of Articulation; Templin & Darley, 1969) and the Weiss Comprehensive Articulation Test (WCAT; Weiss, 1980). These tests sample utterances in single- words, conversations, contextual testing or stimulability testing. Data is analyzed mostly for SODA errors, phonological processes, etc. Phonemes are tested for accuracy in word initial, medial and final positions. In case of continuous speech sampling, only the target words are transcribed and analyzed for ascertaining the error patterns. These tests follow a norm based protocol wherein the child tested is compared against his/her peer group for his articulatory skills. Another similarity of these traditional articulation tests is that all these tests believe in acquiring a phonetic inventory from the child, basically for the purpose of comparison with norms.

Construction of an Articulation Tests: To serve the purpose of diagnosis, an articulation test ought to be brief and precise keeping in mind the patient load and time constraints in routine clinical set-up. So, while constructing a test of articulation, the case of test administration, scoring and analysis, and the time effectiveness of the test material must be kept in mind. It usually follows two stages;

1. Ascertaining the language in which the test of articulation has to be constructed, the inventory of phonemes of the language and the basic phonemic alignment of the language in terms of occurrence in word positions (initial, medial and final)
2. Choosing picturable, unambiguous and familiar words with all the phonemes of the language in all naturally occurring word positions.

   Babu, Rathna and Bettagiri (1972) selected 372 words, Padmaja (1988) opted to choose 283 words, Maya (1990) initially selected 350 words, and Usha (1986) used 218 words with all naturally occurring phonemes in Kannada, Telugu, and Malayalam and Tamil languages, respectively. While re-standardizing the test of articulation in Kannada, Deepa (2010) used 485 words. The authors advocate that there be at least 5 words with each phoneme in each natural word position of the language it tends to test.

3. Familiarization of the selected words with a number of adult native speakers of the language to determine the existence of the word in the language and its usage in day to

day conversation. The task employed is generally to judge if a word is most-familiar, familiar or unfamiliar in the language. Out of all the words, the words rated as most familiar by more than 50% of the judges (usually odd numbers are preferable for case) are listed out and among these, the words which are better in wins of picturability, non-ambiguity and familiarity with children or the population the articulation test intends to test are chosen as stimuli words of the test of articulation.

4. The next step is to pictorially represent the stimuli words. In former days, simple line drawings were employed to elicit spontaneous naming of the stimuli words. Shanks, Sharpe and Jackson (1970) and several others support that coloured pictures or photographs are better as the 3-Dimensional nature of the same help children correlate better with images of real objects when compared to line drawings of the same. Once the pictures are selected, they are stacked in the form of 3¯X4" cards or computer based program modules such as PowerPoint slides, photo gallery, etc depending on whether the test is manual, computerized or semi- computerized.

5. The test is now ready to administer with a pre-decided scoring pattern, way of transcription and analysis.

4. Computerized tools

Various computerized tests of articulation are also available. These are mostly used for phonological analysis purpose and the process is called as computer assisted phonological analysis. Klein (1984) pointed out that such computer based programs save time and provide greater detail of analysis than one could possibly do manually. Computer phonological analysis (CPA) software involves proving an input of phonetic transcriptions from a computer keyboard and/ or selecting from predetermined stimuli, displaying the data on the screen, and ultimately printing results of an analysis. Analyses often include both relational and independent analyses of consonants and vowels, word position analysis, syllable shapes used, patterns among errors, and calculation of percentage of consonants correct. A few of such computer assisted tools of phonetic/ phonemic analysis are as follows:

- Computer Analysis of Phonological processes (CAPP), Hodson (1986)
- Programs to Examine Phonetic and Phonological Evaluation Records Version 4.0 (PEP-PER). Shriberg (1986)
- Computer Profiling (CP), Long & Fey (1988)
- Process Analyses (PAC), Weiner (1986)
- Interactive System of Phonological Analysis (ISPA). Masterson and Pagan (1993)
- Pye Analysis of Language (PAL) , Pye (1987)
- The Computerized Articulation and Phonology Evaluation System (CAPES), Masterson and Bernhardt (2001)
- Computerized Profiling (Version CP941.exe.), Long, Fey and Channell (2002)
- Hodson Computer Analysis of Phonological Patterns (HCAPP). Hodson (2003)
- Phonological profile in Kannada: A study on Hearing Impaired, Ramadevi (2006)
- Computer based Assessment of Phonological Processes in Malayalam (CAPP-M), Merin & Sreedevi (2010)
- Computerized Assessment of Phonological Processes in Kannada (CAPP-K), Sreedevi, John, & Chandran (2013)

In the Indian context, Ramadevi (2006) developed a phonological profile in Kannada for children with hearing impairment. This module assists in computerized presentation of stimuli. Responses are transcribed manually using broad transcription. Sreedevi and Merin (2011) and Sreedevi, John and Chandran (2013) developed the

Computer based Assessment of Phonological Processes in Malayalam (CAPP- M) & Computerized Assessment of Phonological Processes in Kannada which uses Adobe Flash framework and can run on any computer which is preloaded with this software. These soft ware's displays the stimulus in the form of coloured pictures and the correct target utterance printed in IPA below it and automatically assesses the phonological processes in various age groups. It also portrays three possible erroneous productions and an extra option to include any unspecified production under the head of ˉothers". This also gives a scope of analysis and at the end, depicts the phonological processes the child manifests along with its frequency of occurrence.

The chief advantage of using a computer based tool for analysis of articulatory skills of an individual lies in expected time saving for the analysis, and a potential for obtaining and organizing large amount of data in a more systematic fashion. Moreover the accuracy of quantitative data derived through computer, analyses is more certain. However, computer doesn't "do it all"; most of these computer analysis procedures are not yet sophisticated enough to determine the articulatory errors due to the usage of outdated technology of execution.

Though computer based phonological assessment has set its roots in clinical assessment procedure, it still has to be popularized and be used on a large scale. As of now, "paper pencil" tests are being used by many SLPs worldwide. Even though the presentation of stimulus has been picturized using the above mentioned soft ware's most of the analysis is still carried out by the clinicians. Administering a test requires a lot more than just eliciting the speech sample from the child. Articulatory changes occur so fast and are so transient that a clinician must be careful and fast enough to capture the articulatory changes and must have a very good discretion so as to make judgments regarding the correctness of a particular stimulus. Winitz (1969) advocates that SLPs' judgments are fairly reliable in binary decision making tasks where they just have to predict if an utterance is produced correctly *or* incorrectly. But the reliability deteriorates while the challenge is to make finer decisions like the nature of articulatory error, the context in which the articulatory errors occur, the consistency with which they are produced and so on. This can be assisted with an objective tool that could point the type of articulatory effort exhibited by an individual rather than just assisting in presentation of the stimulus.

As seen above, the current trend in articulation test is to use current technology for making a computerized tool along with development and standardization of a traditional pen-paper articulation test. The question arises as to which exact technology is suitable for building a tool in articulatory assessment.

5. Speech recognition and Mel frequency Cepstrum Coefficient (MFCC)

The field of automatic speech recognition (ASR) has its glimpses as old as 1940's. But the advancements in the field of speech recognition in the last couple of decades have been remarkable. In this 21st century which is considered to be an era of electronic gadgets and intelligent machines, ASR is applicable in our day to day activities as being in available in our native languages ASR has successfully established its grounds in all fields. ASR is a process of automatically recognizing words, spoken by a human, using the input speech signal (Anusuya & Katti, 2009). Speech to text is a important product of ASR which has had its applications in almost every fields such as computing (Cooke, Morris, & Green, 1997), voice dialing (Monrose, Reiter, Li, & Wetzel, 2001), call routing (Gorin, Riccardi, & Wright, 1997; Hakkani-Tur, Tur, Rahim, & Riccardi, 2004), domestic appliance control (Damper, 1984; Noyes & Frankish, 2009), preparation of structured documents (Hansen et al., 2005; Misu & Kawahara, 2007), health care (Young & Mihailidis, 2010), military (Beek, Neuberg, & Hodge, 1977), telephony (Zissman, 1996), home automation (AlShu'eili, Gupta, & Mukhopadhyay, 2011; Wang, Lee, Wang, & Lin, 2008). Application of ASR in the field of communication sciences have also been tried out with increased implications in the recent past (Lippman, 1997).

Speech recognition is generally classified based on the type of input utterance used such as isolated word, connected speech, continuous/ spontaneous speech etc. Generally, isolated speech recognition is carried out in most of the ASR systems where single utterances with pauses in between them. The processing of the input elements is done during the pauses in between.

Input Speech signal into the ASR systems are a combination of various relevant information such as fundamental frequency, formant related information etc and other irrelevant information such as background noise, interfering signals etc. Therefore, an ASR system should be capable of filtering relevant from that of irrelevant information. This process of capturing vital information from the input signal is called as feature extraction. Various feature extraction techniques have been demonstrated to be useful in ASR, such as Linear Predictive Co-efficient (LPC), Mel Frequency Coefficients (MFC), Mel Frequency Cepstrum Coefficients (MFCC), Linear Predictive Cepstrum Coefficients (LPCC) etc.

Psychophysical studies of the frequency resolving power of the human ear has motivated modeling the non-linear sensitivity of human ear to different frequencies (Holmes, 2001). The selective frequency response of the basilar membrane (hair spacing) acts as a bank of band pass filters equally spaced in the Bark scale. Figure 1 shows the linear spacing between 100 Hz to I kHz and the logarithmic spacing above 1 kHz further reduces dimensionality of frame/vector of speech. The low-frequency components of the magnitude spectrum are ignored and the useful frequency band lies between 64 Hz and half of the actual sampling frequency. This band is divided into 23 channels equidistant in Mel frequency domain.

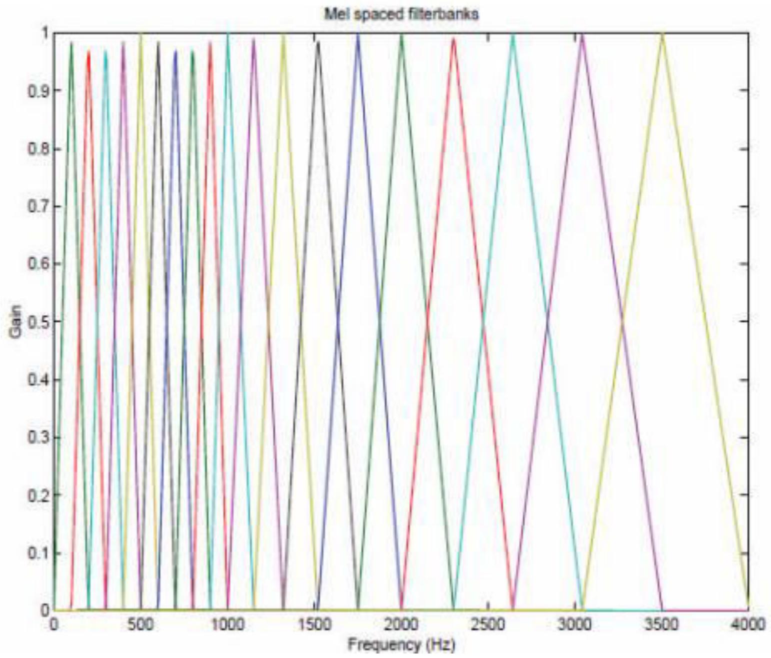


Figure 1: Mel Frequency spaced Filter banks

The model tradition is to represent the log spectrum with axis in Hz scale. It is likely to compute the log spectrum with the frequency axis in Mel- scale (or in bark-scale or in logarithmic scale). Figure 2 shows the schematic representation of log spectrum of a vowel segment in Hz and Mel scale. Note that the low frequencies are packed in. There are standard formulae for converting frequency in Hz to frequency in Mel (or bark) and vise- versa. Discrete Fourier transform (DFT) of log spectrum with the frequency axis in Mel scale gives Mel- scale cepstrum.

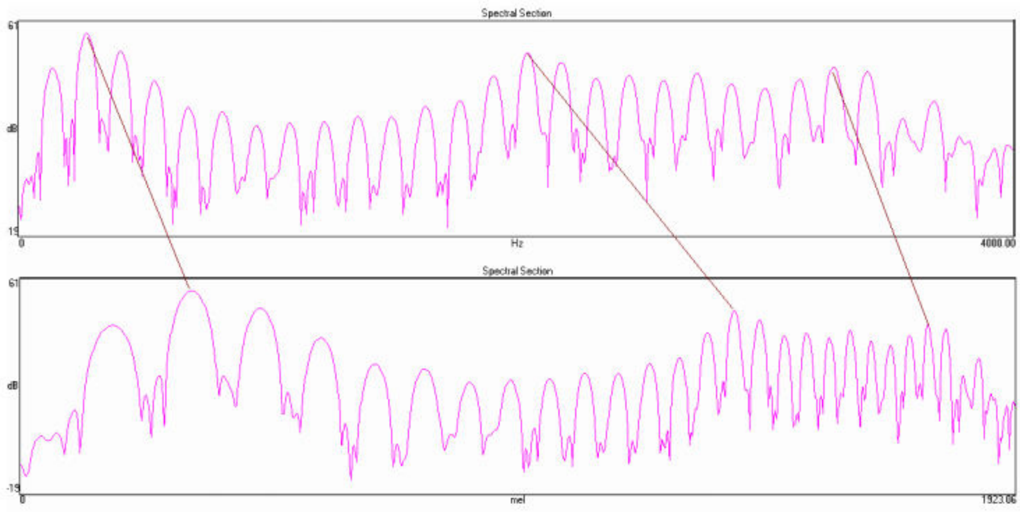


Figure 2: Log spectrum of a vowel segment in Hz scale (top) and in Mel Scale (bottom)

Experimentally, researchers have established that of all the features, MFCCs give the finest performance for automatic speech recognition, speaker recognition, language identification etc. Although, in principle MFCCs are alike to the Mel-scale frequency cepstrum described above but the differences arise in two distinct ways: (a) A filter bank to obtain the log spectral components with frequency axis in Mel scale (b) Instead of a DFT, a transformation called Discrete Cosine Transform (DCT) is used.

Using a filter bank: The magnitude squared spectrum is computed in the standard manner using DFT applied on the windowed pre-emphasized speech segment. The magnitude squared spectrum is warped so that the frequency scale is in Mel. Then the Mel scale magnitude squared spectrum is divided into M bands of equal width by means of overlapping triangular windows, where M is a pre-decided number of coefficients (Figure 2). The magnitude squared spectrum is multiplied by a triangular function which retains frequency components of only one band [Figure 2(a)]. The total energy in the band is the sum of all the non-zero spectral components of that band. The energy is then represented in dB (or log) scale to obtain the filter bank outputs. This procedure gives M filter bank outputs.

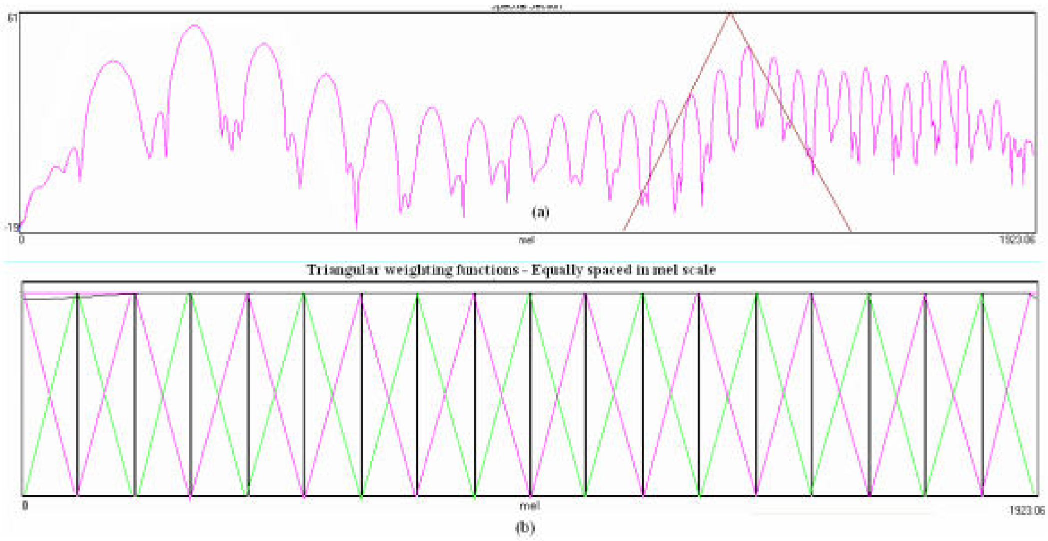


Figure 3: (a) magnitude squared spectrum in Mel scale is multiplied by a triangular band pass filter response. (b) A number of triangular filters constitute a filter bank. The filters are equally spaced in the Mel frequency scale.

Mel frequency Cepstral Coefficient (MFCC) was introduced by Davis and Mermelstein in 1980 and is considered to be one of the most widely accepted feature extraction method. Its computation is based on extraction of cepstral features from Mel scaling frequency domain. Summing up, the process of extracting MFCC from continuous speech is illustrated in Figure 4 given below.

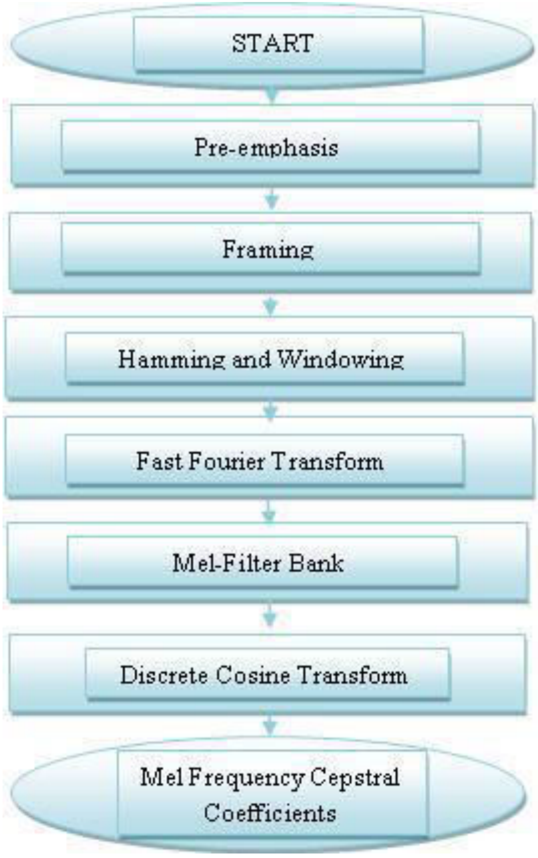


Figure 4: Mel Frequency Cepstral Coefficient Block diagram.

6. Dynamic time Warping (DTW)

One among the earliest methods of speech recognition using isolated words was to store prototypical versions of all words (template) in the database and compare the incoming speech signal with the stored template for the closest match between them. But the problem between template matching was that all the signals being tested had to be exactly of similar duration with that of the template and as we know that no two utterances are exactly the same because speech is a time dependent process.

Therefore, one word can have many variations based on different durations of either the initial, medial or final segments in it. Other factors such as speaking rate may also add to these variations. Hence there is a need to time alignment of all these speech patterns for better and easier speech recognition.

Consider the word "SPEECH" as a template word and input speech as "SsPEEhH". The input speech "SsPEEhH" is a noisy version of the template and when both these are time- time aligned the reference pattern (template) on the vertical side and input pattern along the bottom it can be noted that the 'h' is a closer match to 'H' than anything else in the template.

When the input pattern ``SsPEEhH" is matched against all template available in the database the best matching will be noted where there in minimal distance path of the alignment between the templates and the input pattern. A simple global distance score for a path is simply the sum of local distances that go to make up the path (figure 5).

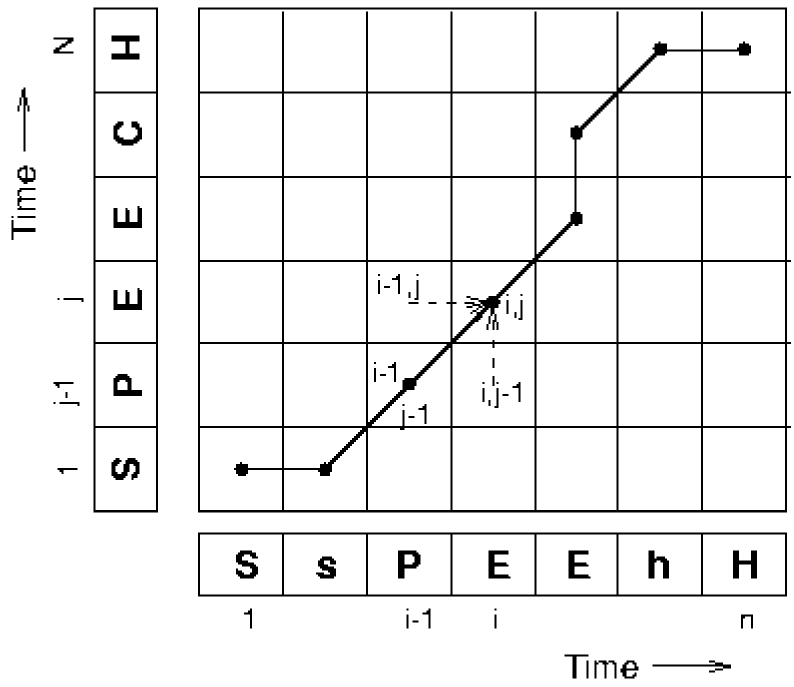


Figure 5: Shows the time alignment between the test and the training pattern using DTW.

Based on the above figure it can be noted that to make the computation faster and easier certain restriction on the direction of propagation of alignments can be made. The constraints include matching paths cannot go backwards in time, every frame in the input must be used in a matching path, and local distance scores are combined by adding to give a global distance. This type of computation is known as **Dynamic Programming** (DP). When applied to template-based speech recognition, it is often referred to as **Dynamic Time Warping** (DTW).

Hence it can be concluded saying that DTW is based on Bellman's principle of Dynamic Programming Problem. It is used to evaluate the similarity/ difference between two time series which vary in speech/ time. This technique is also used to find the optimal alignment between two times series by warping one signal over the other non linearly by stretching or shrinking it along its time axis. The warped signal is used to measure the similarities/ differences between the two time series. Figure 6 shows an example of how one time series is warped to another. In this figure, the horizontal axis represents the time sequence of stored template sample and the vertical axis represents the time sequence of the Input sample. The paths shown as arrows are the resultant minimum distance between the input and template signal.

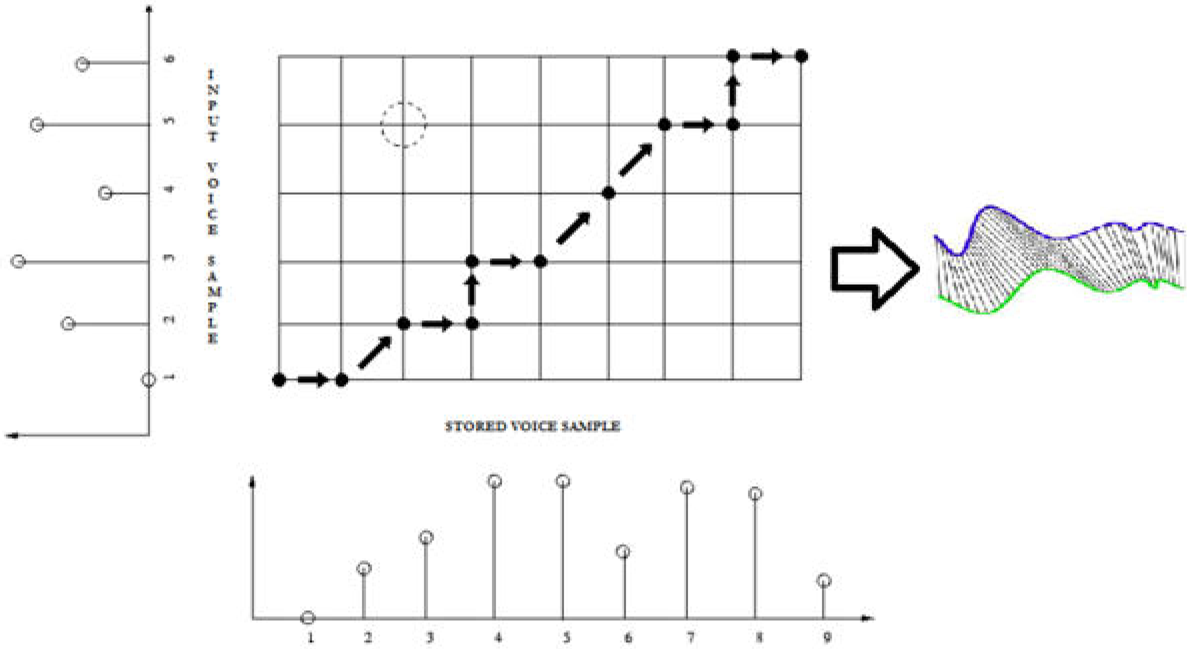


Figure 6: Dynamic Time Warping of Input sample and stored template.

The results of a study conducted by Kinnunen (2003) indicated that the MFCCs are the most apparent example of a feature set that is comprehensively used in speaker recognition. By using MFCCs feature extractor, one makes a postulation that the human hearing mechanism is the forest speaker recognizer. Authors aimed to find the critical parameters that affect the performance and also tried to give some general guidelines about the analysis parameters. They conducted experiments on two speech corpora using vector quantization (VQ) speaker modeling. The corpora were a 100 speaker subset of the American English TIMIT corpus, and a Finnish corpus consisting of 110 speakers. Although noise robustness is an important issue in real applications, however, it was outside the scope of the thesis. The author's main attempt was to gain at least some understanding of what is individual in the speech spectrum. The results indicated that in addition to the smooth spectral shape, a significant amount of speaker information is included in the spectral details, as opposed to speech recognition where the smooth spectral shape plays more important role.

A preliminary study by Chen and Huang in 2006 on Mandarin language proposed an articulation training system with automatic pronunciation detection mechanism which compared base on the modeled articulation errors based on MFCC and Hidden Markov Model (HMM). Results revealed that the average mean opinion scores or matching scores of around 4.13/5 using the proposed method with detection rate of around 82.7%.

Further a follow up studied by Chen, Huang, Yang, Lin and Wu (2007) modified the proposed automatic pronunciation detection system and reported that the new improved system can be implemented in daily clinical usage.

Maier, Noth, Batlier Nkenke, and Schuster (2006) demonstrated the use of ASR as a means to quantify the intelligibility of individuals with Cleft lip and palate. Study concluded saying Automatic speech evaluation by a speech recognizer is a valuable means for research and clinical purpose in order to determine the global speech outcome of children with CLP. It enables to quantify the quality of speech. Adaptation of the technique presented here will lead to further applications to differentiate and quantify articulation disorders. Modern technical solutions might easily provide specialized centers and therapists with this new evaluation method.

Various studies on ASR system using MFCC as feature extractor and DTW as data classifiers in different languages such as Malaysian (Muda, Begam, & Elamvazuthi, 2010), Marathi (Gangonda & Mukherji, 2012; Ghule & Deshmukh, 2015), Hindi (Dhingra, Nijhawan & Pandit, 2013), English (Mohan & Babu, 2014) have demonstrated the reliability of ASR systems using MFCC and DTW.

Similar findings were reported by Pandit and Bhatt (2014) reported a phoneme based speech recognition system with MFCC as feature extractors and DTW as classifier for Guajarati digits. Results revealed 8/10 digits (84.44 %) were matched correctly with the templates in database. Accuracy was reported to be improved to 95.56 % when background noise while recording was eliminated by data segmentation.

Vashisht, Sharma and Dogra (2015) using MATLAB software 7.5 evaluated the factors affecting the recognition accuracy of an ASR system based on MFCC and DTW. Results revealed that factors like Noise, Continuous Speech, Regional and Social Dialects, Movement of Body while recording had significant effect on the recognition scores.

Saastomonen, Fiedler, Kinnunen and Franti (2005) carried out an exploratory study to determine the factors affecting MFCC- based speaker recognition system's accuracy. They explored the affects of (a) linguistic related factors such as text dependency, sample length and native v/s non native language, (b) speaker dependent factors such as reading vs. spontaneous speech & disguise (deliberate cheating), and (c) technical factors such as microphone types and its quality distance to microphone, sampling rate & additive noise. They concluded saying that noise is one of the most significant factor that degrades the accuracy of recognition for about 75% error rates were reported, the second factor was reported that change in microphone (45%) followed by disguise (30%) with least effects for language dependency.

Similar reports from Singh and Rajan in 2011, who evaluated the factors affecting the accuracy of an ASR system based on MFCC and VQ. They evaluated technical factors such as noise, sampling rate, microphone quality and distance to microphone, speaker dependent factors like reading v/s spontaneous speech & deliberate confusing, and voice sample related factors such as text dependent/independent, sample length and speaking languages. They reported that noise & microphone quality had the strongest effect on performance of the speaker recognition with minimal influence from the other factors. They also reported that the error rate came down as low as 10% when noise and microphone quality were controlled.

Various techniques and modifications of ASR systems based on DTW for improving speed and accuracy of the systems are available in literature, Ratanamahatana and Keogh in 2004 evaluated the usage of DTW in various aspects such as handwriting analysis, face recognition, speech recognition etc. They carried out sets of experiments to evaluate the myths of DTW and concluded that DTW based ASR systems with weighted DTW increase speed of processing but hinder the accuracy.

Abdulla, Chow and Sin (2003) developed a new technique for ASR systems which produces templates called a crosswords reference template (CWRTs) which matches a set of examples from the database rather than just one to one reference template matching. They concluded that by combining CWRTs and DTW, resulted improvement of ASR accuracy 85.3% to 99%.

Swamy and Ramakrishnan (2013) reported an efficient speech recognition method of CWRT based matching system of MFCC, vector quantization (VQ) and HMM and reported an overall efficiency of around 98%.

The review provided the usefulness of Mel Frequency Cepstral Coefficients and the effectiveness of Dynamic time warping in building an ASR system. There are no empirical studies which utilize ASR systems to automatically evaluate articulation abilities of an individual. Even though various efforts to come up with computerized tool for articulatory assessment have been made there is no flexibility and ease of usage of these systems. In this context, the present study was to build a fully automatic screening tool for articulation in

Kannada language using Mel Frequency Cepstral Coefficients & Dynamic Time Warping on MATLAB platform using CWRT method. The objectives of the present study were to (a) build the Computerized Screening Articulation Test using Matlab®, and (b) investigate its effectiveness using by preliminary analyses.

Chapter III

Method

The primary objective of the present study was to develop a fully automatic computerized articulation assessment tool with a database to compare articulatory skills in native Kannada speaking individuals. The study was conducted in 2 phases. Phase I was to obtain phonation and target words from the participants and use the built database for creation of the computerized screening tool. Phase II included testing of the developed tool using preliminary analysis.

Phase I: Creation of database and the Development of screening test using Matlab®.

Material: Ten words targeting consonantal productions depicted in table 1 were borrowed from part I of Kannada Diagnostic Photo Articulation Test (Deepa & Savithri, 2010). Words targeting vowels and cluster productions were omitted as the present study was a preliminary screening test. The available photo templates depicting the selected 10 words were edited using Adobe Photoshop (Adobe System Inc., 2011) to add an orthographic representation of the respective word as shown in table 2. These edited photo templates were saved as pictures in .jpeg format which formed the material (Figure 7).

Sl No	Phoneme Tested	Target words	Orthographic representation
01	t̪ in word -Initial	/t̪aʈʈe/	ತಟ್ಟೆ
02	t̪ in word -medial	/ko:t̪i/	ಕೋತಿ
03	d̪ in word -initial	/d̪a:ra/	ದಾರ
04	d̪ in word -medial	/kud̪ure/	ಕುದುರೆ
05	l in word -initial	/lo:ʈa/	ಲೋಟ
06	l in word -medial	/hallu/	ಹಲ್ಲು
07	ʈ in word -Initial	/ʈo:pɪ/	ಟೋಪಿ
08	ʈ in word -medial	/ʧiʈʈɛ/	ಚಿಟ್ಟೆ
09	h in word - Initial	/ha:vu/	ಹಾವು
10	h in word -medial	/simha/	ಸಿಂಹ

Table 2: Selected ten words from KDPAT targeting consonantal productions

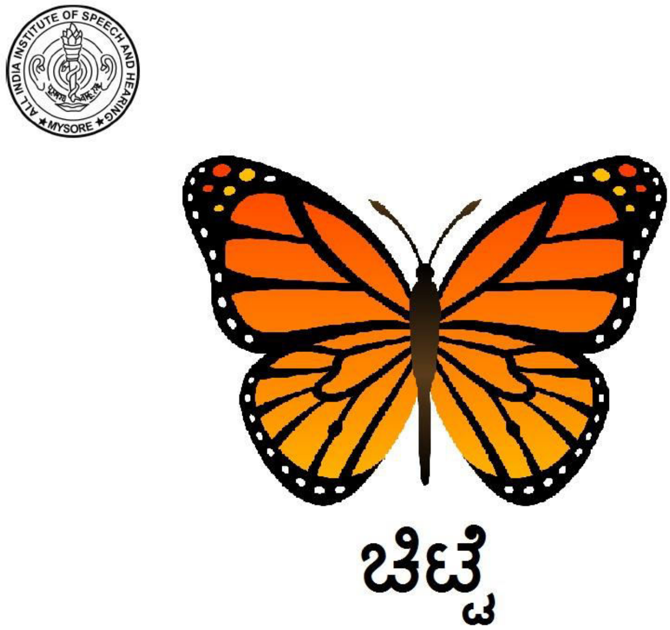


Figure 7: Depicting an orthographic representation of the target word.

Participants: Thirty four individuals participated in the study, Group I consisted of 13 typically developing Kannada speaking children whose age ranged from 4 to 15 years (mean age of 9 years), Group II consisted of 11 normal Kannada speaking males in the age range of 18 to 26 years (mean age of 22 years), and Group III consisted of 10 normal Kannada speaking females in the age range of 18 to 26 years (mean age of 23 years).

Based on the study criterion, only those participants who were native speakers of Kannada and belonging to middle socio economic status were included in the study. All the participants were screened for speech, hearing, cognitive and communication disorders, upper respiratory tract infections, asthma or any allergies at the time of recoding. Brief history and Ling test was also carried out.

Instrumentation: OLUMPUS- LS 100 portable digital audio recorder with a sampling frequency of 8 kHz and a mouth-microphone distance of 10 cm for audio recording, and PRAAT software (Boersma & Weenink, 2015) were used for data segmentation and preparation. MATLAB (The MathWorks Inc., R2015b) was used to design the software.

Procedure: Participants'/ parents (for children) were briefed about the study and before collecting the samples, written consent were obtained from all the participants. Participants in all groups were assessed for general speech-language, oro-facial, voice related, breathing and medical problems by collecting demographic data. All the participants were allowed sufficient time to practice before the recording. They were seated comfortably in a quiet environment. Participants were made to maintain 10 cm distance from their mouth to microphone and recordings were done using OLUMPUS- LS100 portable digital audio recorder at a sampling frequency of 8 k Hz. They were instructed to sustain phonations of /a/, /i/ and /u/ in three trials in their comfortable loudness and pitch after taking a deep breath for 10 seconds each. Along with this, the selected 10 words in 5 trials targeting consonantal productions and their error combinations were recorded from all the participants. The above tasks were demonstrated by the experimenter before recording the same. The recordings of all the samples were done in a quiet sound treated room. The following instructions were given to the participants for 2 tasks.

Task 1: "maintain a distance of 10 cm from your mouth to microphone, do not move away or towards the microphone and when you are ready, take a deep breath and say the vowel /a/, /i/ and /u/ at your comfortable loudness and pitch at least for 7 to 8 sec. Repeat the same thrice".

Task 2: "Read/Repeat the target words at your most comfortable loudness and pitch and repeat it five times."

The recorded phonations and words were loaded on to PRAAT software (Boersma & Weenink, 2015) running on Windows Personal Computer and displayed as waveforms. Segmentation of unwanted signals and labeling of the processed sample were carried out. The segmented samples were loaded onto MATLAB software. By using the VOICEBOX speech processing toolbox http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html of MATLAB, the first 13 Mel Frequency Cepstrum Coefficients (MFCC) along with its first and second - order derivatives were extracted for each sample. This process was repeated for all the samples and corresponding MFCC's and its derivatives were stored as a database (.dat) file for an individual speaker. Similar procedure was carried out for all participants' audio recordings. Therefore, (.dat) files for all speakers were created and labeled accordingly.

MATLAB (The MathWorks Inc., R2015b) was used to design the software. It consists of two modules, first module i.e. ***training module*** evaluates and classified the signal as belonging to Group I, II or III based on the frequency characteristics of the signal. This would be essential as the F0 of each group participant differs and comparing a test signal uttered by Group I with that of the trained signal in Group II or III would be hazardous. The second module i.e. ***error analysis module*** screened the speech sound productions of an individual and displayed results.

Phase II: Efficacy testing through preliminary analyses

Participants: Three participants (1 adult male+1 adult female+ 1 child) from Phase I were randomly selected to participate in the second phase of the study.

Procedure: A simulation study was carried out where the participants were instructed to produce the target words with misarticulation (any error combination of initial or medial position of the target words) into the developed Computerized Screening Articulation Test and simultaneous audio recording was carried out using OLYMPUS LS 100 audio recorder. These audio recording of the samples were perceptually analyzed by two SLP's for correctness of articulation. The results obtained from the developed Computerized Screening Articulation Test and SLP's ratings were compared and analyzed statistically using commercially available SPSS software.

Analysis: Analyses was carried out using the developed Computerized Screening Test of Articulation used DTW algorithm. For doing DTW, a reference (train) and a test signal were compared. MFCCs of phonations and words in the database formed the reference/train signal and a new phonation / word from any speaker formed the test signal. These two signals were compared using DTW.

Initially in the ***training module***, phonation produced by an individual was analyzed for MFCCs and based on the comparison of MFCCs in the database s/he was classified as belonging to one of the three Groups (Figure 8).

The ***testing module*** analyzed specific speech sounds by utilizing the image templates (.jpeg files) of the selected words. When using this module an individual speaker produced the target word (test signal) into the computer by clicking on "Record" and "Analyze" buttons accordingly (Figure 9). The produced target word (test signal) was automatically analyzed for their corresponding MFCC's and were compared with the reference templates of the database using DTW and corresponding reference id of the selected word was displayed.

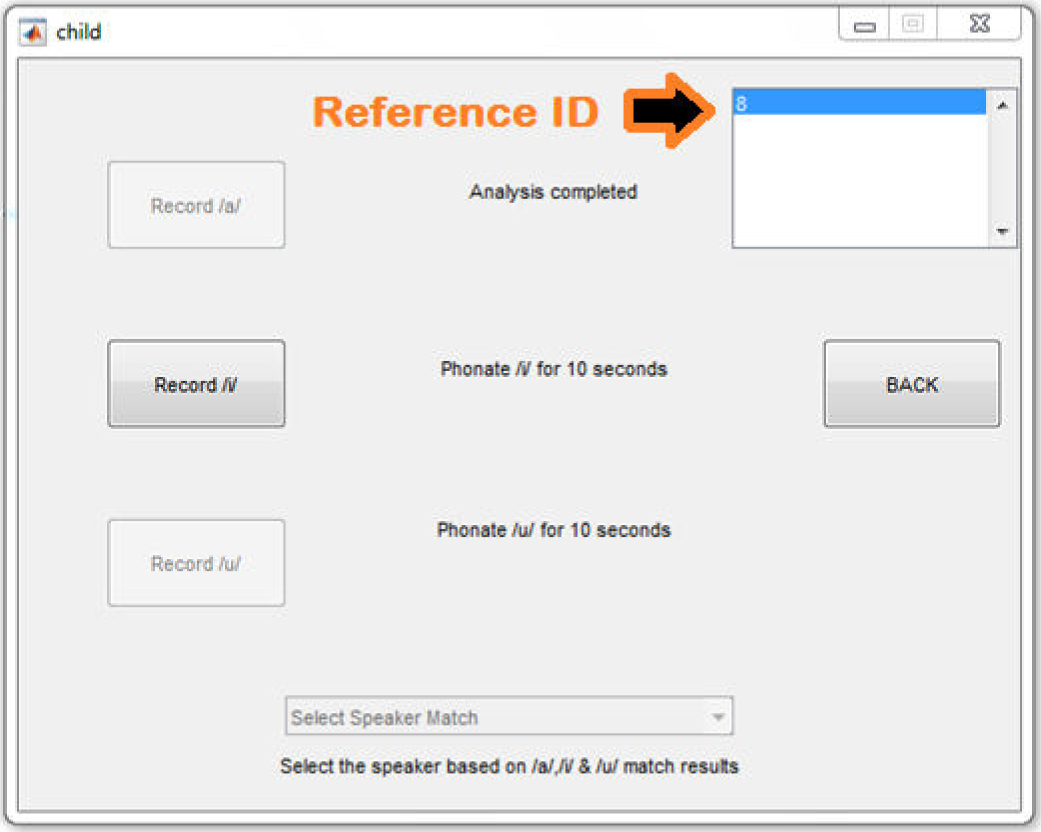


Figure 8: Pictographic representation of Module 01 of Computerized Screening Articulation Test

Figure 9: Pictographic representation of Module 02 of Computerized Screening Articulation Test

The SLPs conducted perceptual analyses and categorized the words into 4 categories, i.e. correct production, substitution of target sound, omission of target sound and other errors.

Statistical Analyses: Statistical Package for Social Sciences (SPSS) version 16.0 was used to analyze the data. Inter rater reliability was checked to compare the results of the SLP's rating and developed software results. Cohen's Kappa coefficient was carried out to investigate the inter rater reliability of SLP's and software of all the participants of Phase II.

Chapter IV

Results

Phase I: Database: Substitution and omission errors were derived along with few other frequently seen erroneous productions for the selected 10 words based on literature review of studies related to phonology (Ramadevi, 2006; Deepa, 2010; Sreedevi, John, & Chandran, 2013). Based on the review, 37 words were finalized under categories of correct production, substitution, omission and other errors which is summarized in Table 3.

Sl. No.	Target/ Correct Word	Substitution	Omission	Other errors	Total
01	ṯaṭṭe	-	aṭṭe / aṯṯe	ṯaṯṯe	4
02	ko:ṯi	ko:ṭi	ko:i	ṯo:ṯi / o:ṯi	5
03	ḏa:ra	-	a:ra/ a:la	ḏa:la	4
04	kuḏure	-	ku:re	ṯuḏure/ kuḏue	4
05	lo:ṭa	-	o:ṭa/o:ṯa lo:ṯa	-	4
06	hallu	-	hau	aḷu	3
07	ṭo:pɪ	ṯo:pɪ	o:pɪ	-	3
08	tʃiṭṭɛ	tʃiṯṯɛ/	ṯiṯṯɛ	ṯɛ/ iṯṯɛ / ṭɛ	6
09	ha:vu	-	a:vu	-	2
10	simha:	-	sima:	-	2

Table 3: Distribution of Target words under correct, substitution, omission and other errors.

Thirty four participants were initially included in the study. Based on the inclusion and exclusionary criteria, three children were excluded from the study due to upper respiratory tract infections (participant 5 & 7) and asthma in the 6th participant. Participant 5 in male group was also excluded due to upper respiratory tract infection.

The samples obtained from the former part of the study which included data recordings and photo templates from KDPAT (Deepa & Savithri, 2010) were used to build the computerized articulation tool.

Framework: The software application was developed using MATLAB (Mathwork solutions, 2015). To run the software, CAT's it is mandatory to install MATLAB runtime Compiler version 8.5 (R2015a). The MATLAB Runtime is a standalone set of sharer libraries that allows the execution of elements compiled using MATLAB, i.e. it helps in running MATLAB components on to computers that does not have MATLAB installed on it. Windows 32/64-bit version of the MATLAB runtime was downloaded from the MathWorks Web site by navigating to http://www.mathworks.com/products/compiler/mcr/index.html.

The application (CAT's) can run on any computer provided the runtime compiler is installed in the system intended for assessment procedure. The application tool CAT's is available for testing in the form of a Compact disk with the compiler.

Working: The complier was installed from the CAT's CD. After the installation, the software program (CAT's) was run by clicking the desktop icon "CAT". The opening page of the tool shows the initial screen about button, institute logo, calibrate options and authors information (Figure 10).

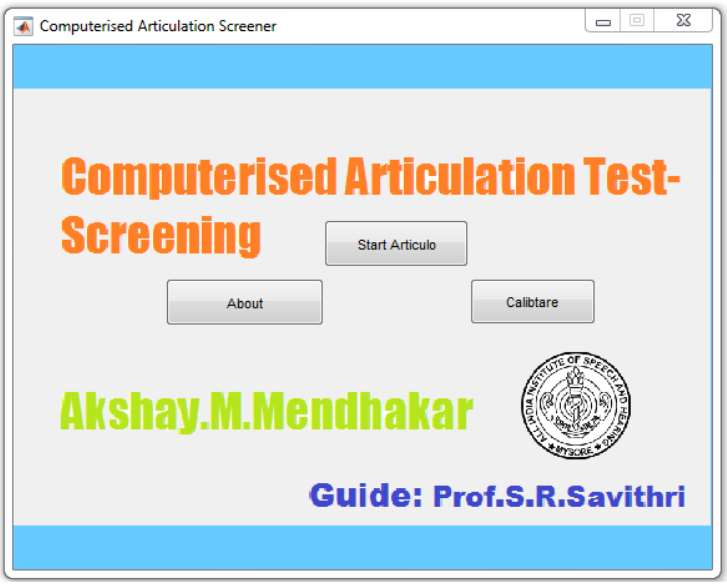


Figure 10: Shows the opening page of CAT's

Calibrate button was used to add or delete sounds from database or make new databases (Figure 11).

Figure 11: Shows the Calibration page of CAT's

"Start Articulo" opens Selector page the with 3 options child, male and female. Based on the individual being tested, the option was selected appropriately (Figure 12)

Figure 12: Shows the Start Articulo page of CAT's

The first module appears on the screen. The buttons and directions was followed to record phonation samples of /a/, /i/ and /u/ (Figure 13). Based on database matching using DTW a reference id of the recognized speaker is displayed on the right hand corner in the result section (Figure 14).

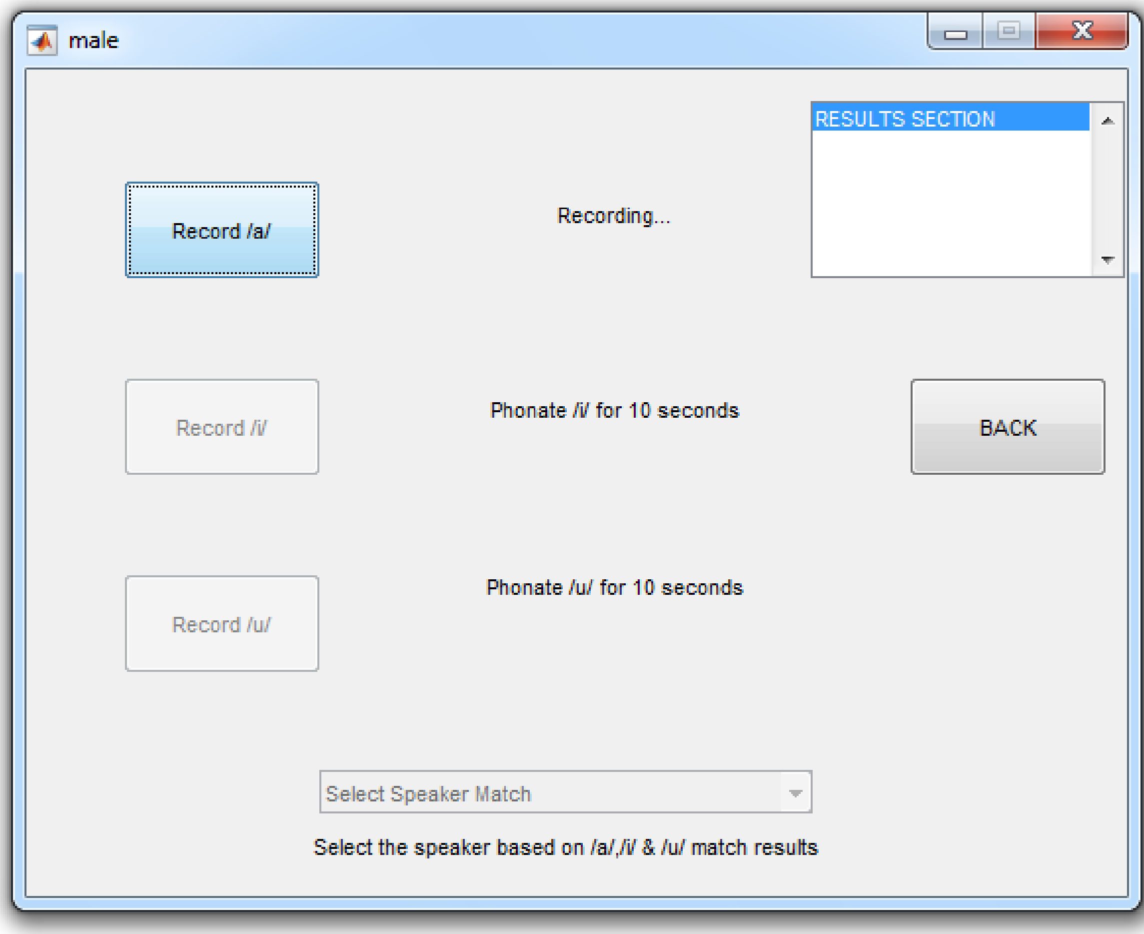


Figure 13: Shows the Module 01 page of CAT's

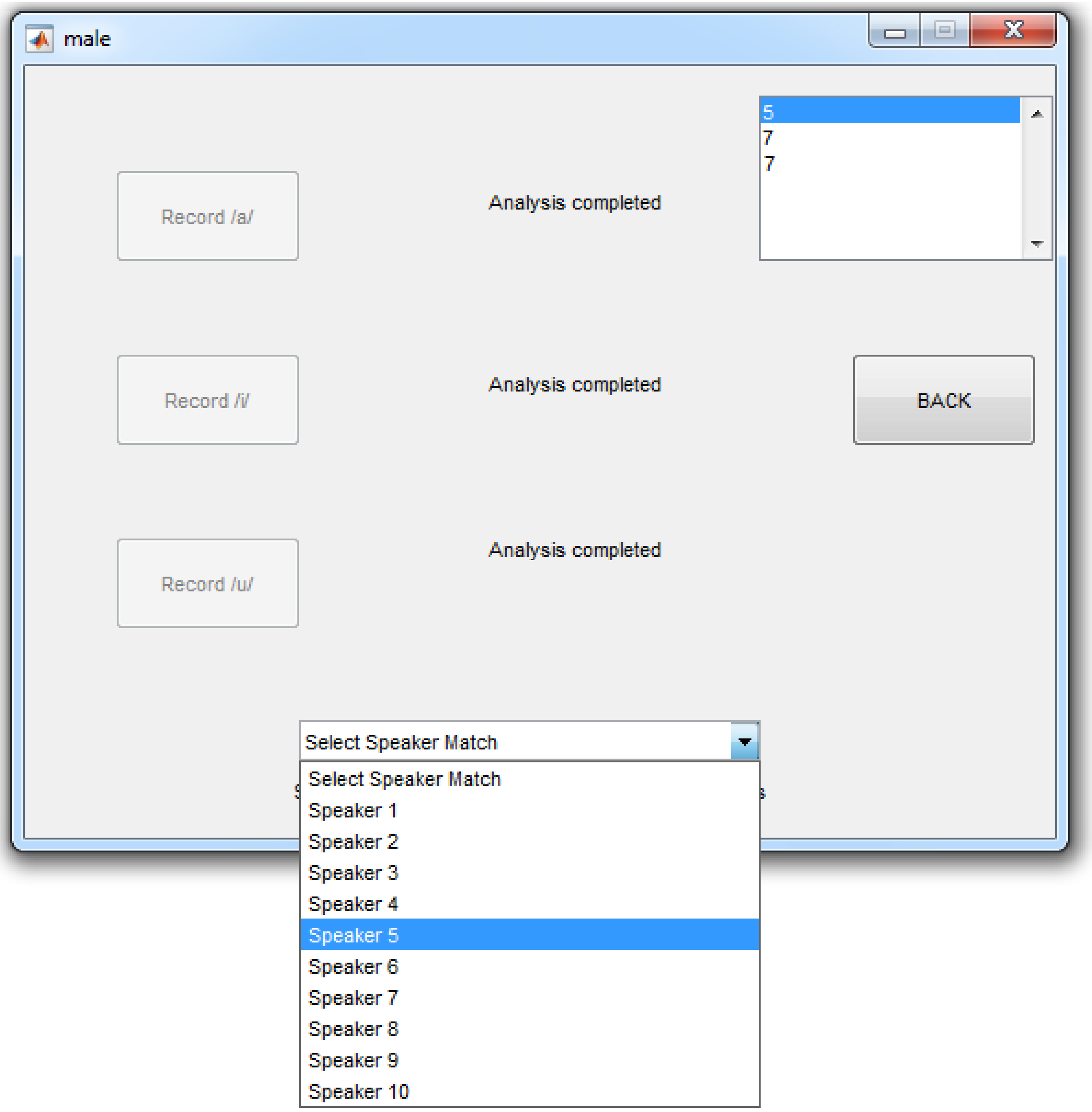


Figure 14: Shows the Selection of the speaker page of CAT's

A sample target word in the tool is displayed in picture form along with its various possible production patterns in target population. The screen contains the picture template on the left side and result section on the right side and 3 buttons (Record, Play & Analyze) at the bottom of the screen (Figure 15).

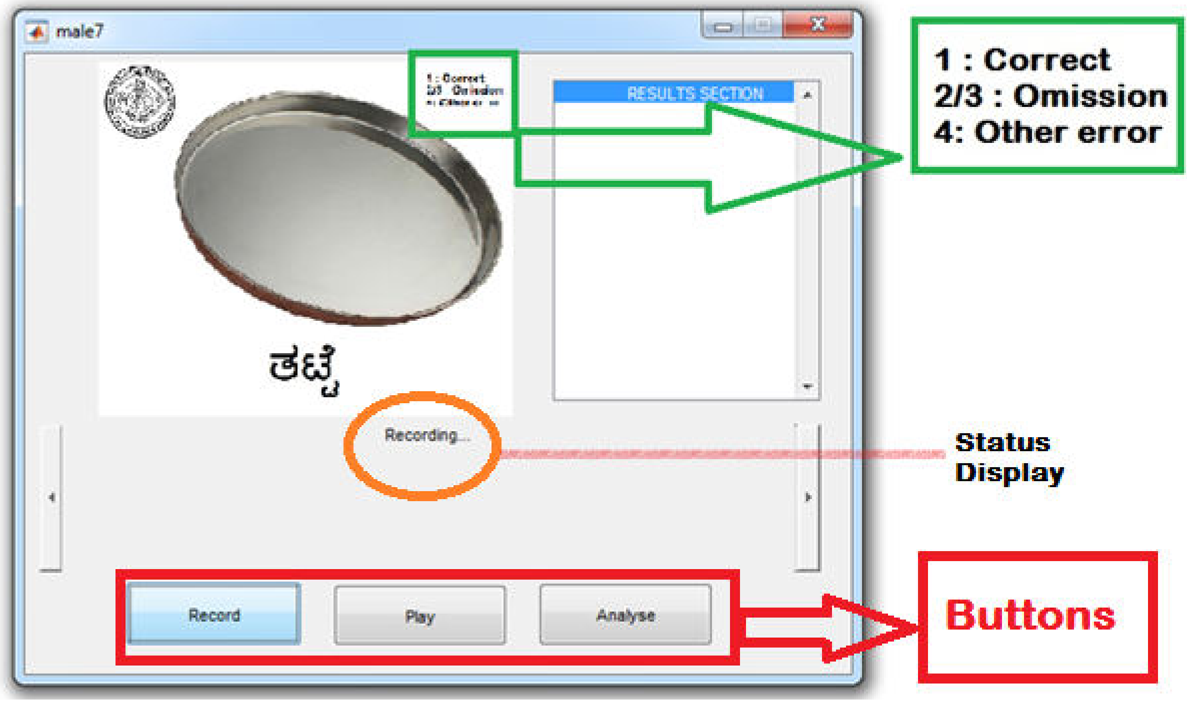


Figure 15: Shows a sample word from Module 2 page of CAT's

The clinician has to navigate with "Next" and "Back" buttons to change the target words and click "Record" and "Analyze" accordingly. Recognized Id will be displayed in the result section on the right side of the screen (Figure 15). Status display provides information regarding ongoing operations of the software like recording, analysis progress etc (Figure 15).

Once the entire list of test words are administered, the page "test completed" will be displayed as seen in figure 16.

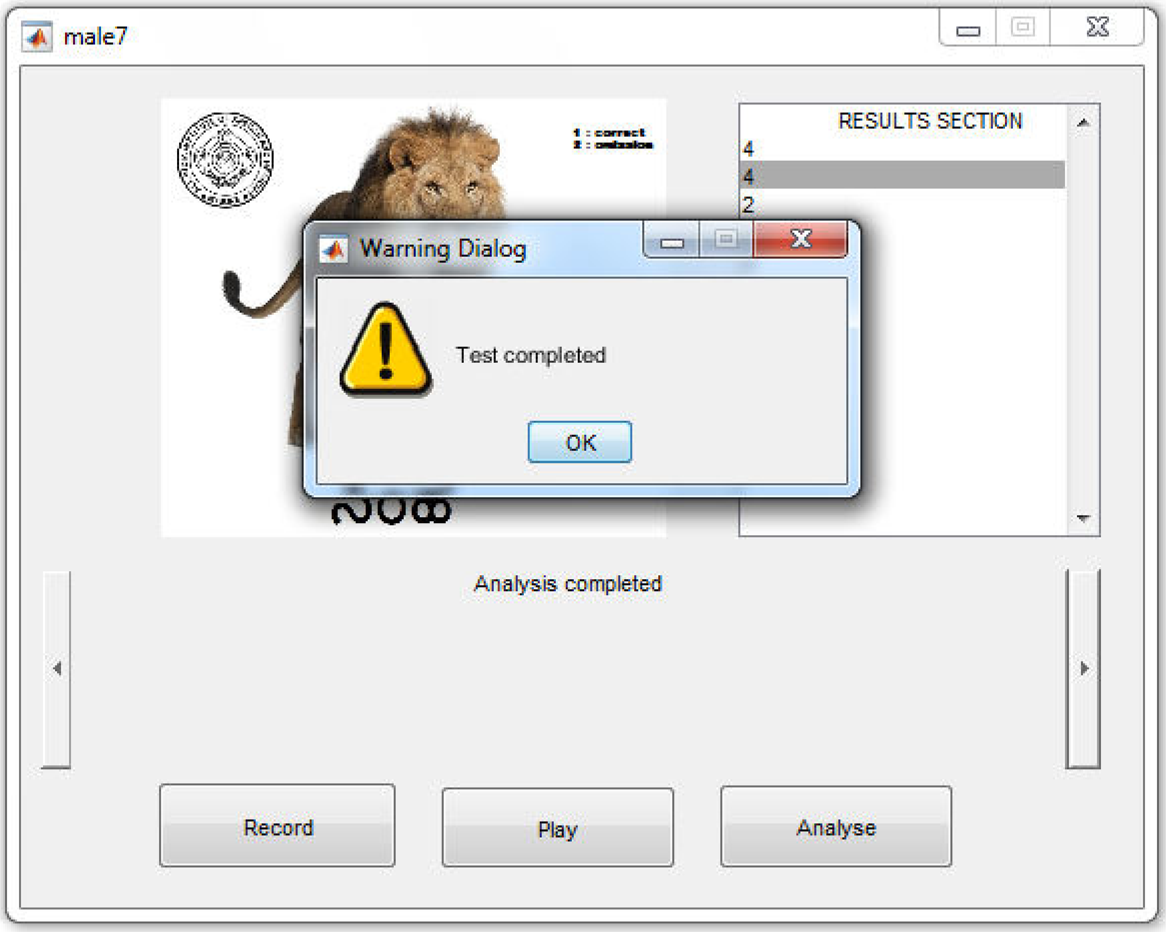


Figure 16: Shows a test completion page of CAT's

Phase II: Based on randomization, a speaker belonging to male, female and children group was selected - i.e. 4th participant in children group, 7th in male group and 1st in female group. The simulation recordings of the selected words were simultaneously analyzed by the developed tool and also the data recording of the same was rated by 2 experienced SLP's and grouped under 4 categories as discussed earlier.

Perceptual analysis of samples by SLP's: Results of perceptual analyses revealed correct production, substitution, omissions and other errors. Results of Cohen's kappa analysis revealed reliability between judges to be 0.927 for child samples, 1.00 for male samples and 0.963 for female samples at p value of 0.05 for all the conditions. Since, the value was > 0.9

in all conditions, it can be concluded as almost perfect reliability and the ratings done by 2 judges are reliable (Figure 17).

Results of the developed computerized tool v/s original words: To verify the reliability of results obtained from the developed computerized tool to that of original word ratings Kappa analysis was done. Reliability between developed tool and original words were found to be 0.916 for child samples, 1.0 for male samples & 0.957 for female samples, with the p value of 0.05 for all the conditions. The results indicated a good agreement between the software's ratings and original words as the value was > 9 (Figure 17).

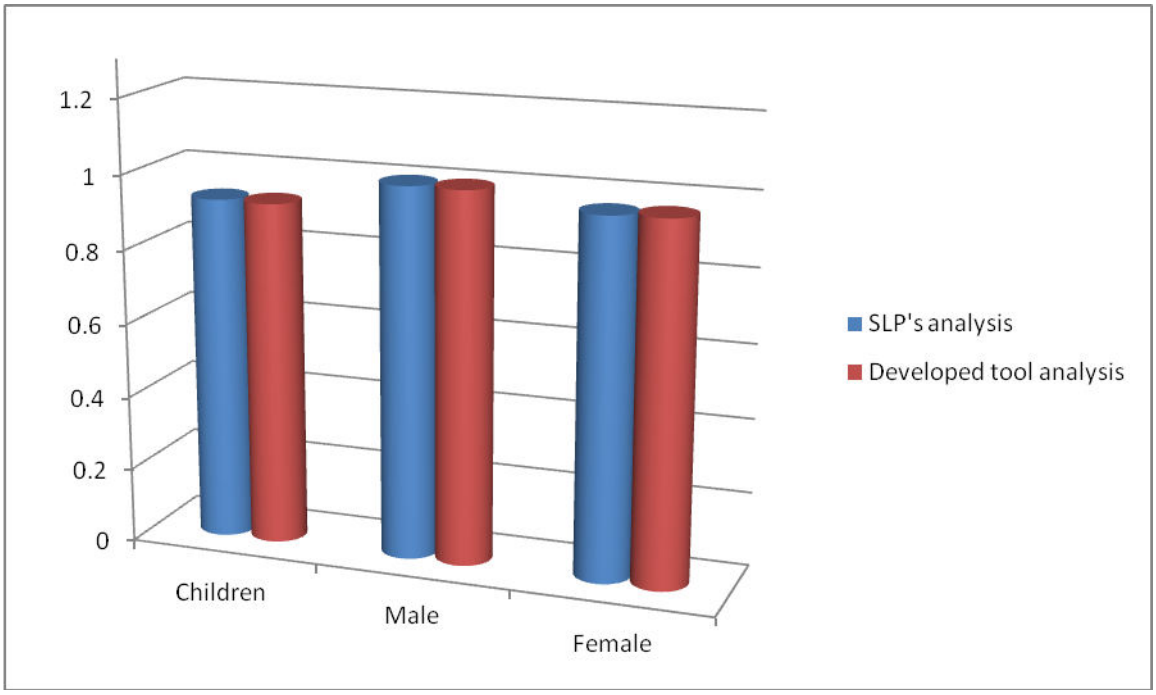


Figure 17: Shows the results of perceptual rating and Developed tool results.

A comparison of the results obtained by the SLP's rating and the developed tool showed a Cohen Kappa coefficient of 0.9 indicating a positive agreement between both the ratings.

Comparison of computerized tool analysis and perceptual analysis by SLP's: Cohen's kappa correlation coefficient was done to see the correlation between the perceptual ratings and computerized tool results. The results revealed perfect positive correlation of 1 for male samples, almost perfect positive correlation of 0.926 & 0.890 for child and female samples respectively at $p<0.05$ for all conditions.

Chapter V

Discussion

The results revealed several interesting points. First of all, it was possible to create a database exclusively for 3 groups of speakers and build the Computerized Screening Articulation Test using Matlab®. Second, the database could identify speaker as male, female or child which eased the analyses. The samples collected were used to extract MFCC's and its first and second order derivatives forming the databases for individual speaker in all groups. The developed tool using MATLAB utilized these built databases to classify the samples based on DTW. This tool was first of its kind.

Third, Cohen's kappa analysis showed a good correlation between the results of the developed tool and SLP's perceptual analysis across all three groups with a p value < 0.05 indicating the effectiveness of the toll. There was a complete positive correlation found between the perceptual analysis and developed tool results for males and almost perfect positive correlation in female and children group. Reliability check showed 100% reliable values by the same judges.

Even though, the obtained results are promising, a curtail part of software's working is missing i.e. the ability to differentiate between silence and utterance and hence fails to segment ambient noises and silence from the signal. This factor has been pointed out to be detrimental in most of earlier studies which evaluated the accuracy of MFCC based ASR's (Saastomonen, Fiedler, Kinnunen & Franti, 2005; Singh & Rajan, 2011; Vashisht, Sharma & Dogra ,2015).

Based on the results obtained from the preliminary analysis, it can be speculated that using the database developed in the present study, it is possible to classify the speakers and their utterances by calculating MFCC's and their derivatives to see whether the sample matches with the database category of a specific individual or not. However, more number of participants can be included in future research work to better differentiate between speakers and to make definite large scale clinical implications.

Chapter VI

Summary and Conclusions

Articulatory development is an ongoing process and its evaluation is a complex task which needs lot of dedication and practice. Due to its complex and time consuming process, evaluation of phonological development is accredited as a task that requires a great deal of expertise. For the analysis of articulatory skills, researchers always rely on an experienced listener or a professional SLP which is always perceptual and subjective. Therefore, there is a need to have an automatic objective tool which can assist professionals in evaluating articulation of an individual. One such objective tool is CAAP-K. Even though this software reduces the efforts of a SLP in transcribing and evaluating phonological processes noted in a child, it lacks flexibility and evaluation is still subjective. Therefore, the present study was undertaken to obtain a full automatic objective tool with speech inputs and extracting MFCC's for the classification of the sample into different categories under correct or erroneous. Also, it made an effort to classify the erroneous productions into substitutions, omission and other errors. Further, the study investigated the efficacy of the developed tool through preliminary analysis.

Ten healthy participants each belonging to male, female and children groups participated in the study based on the predefined inclusionary criteria. The participants were asked to phonate /a/, /i/ and /u/ for 10 seconds and to articulate 10 target words in 5 trails. Adequate practice trials were given for all the participants before the actual recording. The recorded samples were segmented to extract MFCC and its derivatives were extracted to build databases for different individuals. The computerized tool was built using MATLAB software and the databases were loaded onto the software. A simulation study was carried out in the Phase II of the study where 1 individual from each group of male, female and children were selected. Simultaneous analysis by the developed tool and audio recording were done. The audio recordings were perceptually analyzed by 2 experienced SLP’s.

The samples collected were used to extract MFCC’s and its first and second order derivatives formed the databases for individual speaker in all groups. The developed tool using MATLAB utilized these build databases classifies the samples based on DTW. This tool is first of its kind.

Cohen’s kappa analysis showed a good correlation between the results of the developed tool and SLP’s perceptual analysis across all three groups with a p value < 0.05 indicating the effectiveness of the toll. There was a complete positive correlation found between the perceptual analysis and developed tool results for males and almost perfect positive correlation in female and children group. Reliability check showed 100% reliable values by the same judges.

In conclusion it can be stated that even though the evaluation of articulation in individuals being a taxing task, a fully automatic tool that can assist SLP’s during the process of evaluation is finally available. All these years, for all the clinical or research work, traditional pen and paper method of transcribing and analyzing articulatory skills based on perceptual judgment by the experienced SLP were carried out. But the present study developed a computerized tool that can augment and carry out the same activity without depending on a perceptual judgment. The tool can categorize whether the sample was correctly or erroneously produced, based on the comparison of an individual’s utterance and the database files.

Even though, the obtained results are promising, a curtail part of software’s working is missing i.e. the ability to differentiate between silence and utterance and hence fails to segment ambient noises and silence from the signal. This factor has been pointed out to be detrimental in most of earlier studies which evaluated the accuracy of MFCC based ASR’s (Saastomonen, Fiedler, Kinnunen & Franti, 2005; Singh & Rajan, 2011; Vashisht, Sharma & Dogra ,2015).

Based on the results obtained from the preliminary analysis, it can be speculated that using the database developed in the present study, it is possible to classify the speakers and their utterances by calculating MFCC’s and their derivatives to see whether the sample matches with the database category of a specific individual or not. However, more number of participants can be included in future research work to better differentiate between speakers and to make definite large scale clinical implications.

Further, recording conditions and setup used for building the database should be kept constant while using the developed tool on other individuals due to technical limitations such as variations in MFCC’s extracted owing to frequency range of a microphone. Slight change in the recording apparatus can bring in hazardous results of the tool therefore; the above factor seems to be the biggest demerit of the study.

Population considered for building the database is small and the age group considered is also large, hence it is difficult to generalize the results onto a larger population without further research studies on a streamlined group of population.

Notwithstanding the above limitations, the present study was the first attempt to find a fully automatic computerized solution that could assist SLP’s in everyday clinical evaluation of articulatory skills. It can also serve as a screening tool to identify and track articulatory changes in individuals. With further modifications and experimentations the tool can also be used as a therapeutic tool to quantify the improvement due to speech therapeutics. Future research can be done by taking a large sample size with more streamlined age groups in all categories. Similarly studies further can utilize the built computerized tool to build better and larger databases and eliminating the faulty results obtained due to mismatch in database and evaluation recording quality.

References

Abdulla, W. H., Chow, D., & Sin, G. (2003). Cross-words reference template for DTW-based speech recognition systems. In *TENCON 2003, Conference on Convergent Technologies for the Asia-Pacific Region, 4*, 1576-1579.

Adobe Systems Software Ltd (2011). Adobe Photoshop Cs5. *Methods*, Ireland

AlShu'eili, H., Gupta, G. S., & Mukhopadhyay, S. (2011). Voice recognition based wireless home automation system. In *Mechatronics (ICOM), 2011 4th International Conference On Mechatronics* , 1-6.

And, A. B., & Dodd, B. (1996). Do all speech-disordered children have motor deficits? *Clinical Linguistics & Phonetics*, *10*(2), 77–101. http://doi.org/10.3109/02699209608985164

Andrews, N., & Fey, M. E. (1986). Analysis of the speech of phonologically impaired children in two sampling conditions. *Language, Speech, and Hearing Services in Schools*, *17*(3), 187-198.

Animesh, B. (1991). Deep Test of Articulation in Bengali- Picture Form. Unpublished Master's dissertation submitted in part- fulfillment for the Masters – degree in Speech and Hearing. University of Mysore, Mysore.

Anthony, J. L., Aghara, G., Dunkelberger, M. J., Anthony, T. I., Wouldiams, J. M., & Zhang, Z. (2011). What Factors Place Children With Speech Sound Disorders at Risk for Reading Problems? *American Journal of Speech-Language Pathology*, *20*(2), 146–160. http://doi.org/10.1044/1058-0360(2011/10-0053)b

Anusuya, M. A., & Katti, S. K. (2009). Communique: comments on the paper by Perlovsky entitled Integrating language and cognition. *IEEE- Computational Intelligence Magazine*, *4*(2), 48-49.

Arlt, P. B., & Goodban, M. T. (1976). A comparative study of articulation acquisition as based on a study of 240 normals, aged three to six. *Language, Speech, and Hearing Services in Schools*, *7*(3), 173–180. http://lshss.pubs.asha.org/article.aspx?articleid=1749111

Babbie, E., Halley, F. S., Wagner III, W. E., & Zaino, J. (2012). *Adventures in social research: data analysis using IBM SPSS statistics*. Sage Publications.

Babu, R. M., Ratna, N., & Bettagiri, R. (1972). Test of articulation in Kannada. *JAIISH*, *3*, 7-19.

Baker, E. (2012). Optimal intervention intensity in speech-language pathology: Discoveries, challenges, and unchartered territories. *International Journal of Speech-Language Pathology*, *14*(5), 478–485. http://doi.org/10.3109/17549507.2012.717967

Banik, A. (1988). Articulation test in Bengali. *Unpublished Masters Dissertation, University of Mysore*.

Banu, T. (1977). Articulatory acquisition In Kannada: A study of normal children 3-6.6 years. *Research at AIISH*, *1*, 100-101.

Bauman – Waengler, J., A. (2000). *Articulatory and Phonological Impairments: A Clinical Focus. Boston.* Allyn and Bacon.

Beek, B., Neuberg, E., & Hodge, D. (1977). An assessment of the technology of automatic speech recognition for military applications. *IEEE-Transactions on Acoustics, Speech, and Signal Processing*, *25*(4), 310-322.

Bernthal, J.E., Bankson, N.W., & Flipson, P. (2009). *Articulation and phonological disorders: Speech Sound Disorders in Children. (6th ed.).* MA: Allyn & Bacon.

Bernthal, J. E., Grossman, F. M., & Aerts Goll, A. E. (1989). Phonologically delayed children's responses to three types of pictured stimuli. *Journal of Childhood Communication Disorders*, *12*(2), 137-143.

Bernhardt, B. H., & Holdgrafer, G. (2001). Beyond the Basics IThe Need for Strategic Sampling for In-Depth Phonological Analysis. *Language, Speech, and Hearing Services in Schools*, *32*(1), 18-27.

Bharathy, R. (2001). Development of phonological processes of 3-4 year old normal Tamil speaking children. *Unpublished Masters Dissertation, University of Mysore*.

Boersma, P., & Weenink, D. (2015). Institute of phonetic sciences, University of Amsterdam.

Brookes, M. (n.d.). VOICEBOX: Speech Processing Toolbox for MATLAB. Retrieved from http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

Cooke, M., Morris, A., & Green, P. (1997). Missing data techniques for robust speech recognition. In *International conference on acoustics, speech, and signal processing*, *2,* 863-866.

Chen, Y. J., Huang, J. W., Yang, H. M., Lin, Y. H., & Wu, J. L. (2007). Development of articulation assessment and training system with speech recognition and articulation training strategies selection. In *International Conference on Acoustics, Speech and Signal Processing, ICASSP 2007, 4*, 209-212.

Chen, Y. J., & Huang, J. W. (2007). Development of Articulation Training System with Speech Recognition Based Automatic Pronunciation Detection Mechanism. In *3rd Kuala Lumpur International Conference on Biomedical Engineering 2006,* 637-639.

Damper, R. I. (1984). Voice-input aids for the physically disabled. *International Journal of Man-Machine Studies*, *21*(6), 541-553.

Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE- transactions on acoustics, speech, and signal processing*, *28*(4), 357-366.

Deepa, S. (1998). Deep Test of Articulation in Hindi- Picture Form. Unpublished Master's dissertation submitted in part- fulfillment for the Masters – degree in Speech and Hearing. University of Mysore, Mysore.

Deepa, A. (2010). Restandardization of Kannada Articulation Test. *Unpublished Masters Dissertation, University of Mysore*.

Deepa,A., & Savithri, S.R. (2010). Restandardization of Kannada Articulation Test. *Student Research at AIISH*, *8*, 53-65, Part-B, Speech Language Pathology, AIISH, Mysore.

Dhingra, S. D., Nijhawan, G., & Pandit, P. (2013). Isolated speech recognition using MFCC and DTW. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, *2*(8), 4085-4092.

Divya, P. (2010). Articulatory Acquisition in Typically Developing Malayalam Speaking Children: 2-3 years. *Student research at AIISH*, 8, 66-76. Part-B, Speech Language Pathology, AIISH, Mysore.

Dodd, B., & Iacano, T. (1989). Phonological disorders in children: changes in phonological process use during treatment. *International Journal of Language & Communication Disorders*, *24*(3), 333–52. http://doi.org/10.3109/13682828909019894

Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: a normative study of British English-speaking children. *Clinical Linguistics & Phonetics*, *17*(8), 617-643.

DuBois, E. M., & Bernthal, J. E. (1978). A comparison of three methods for obtaining articulatory responses. *Journal of Speech and Hearing Disorders*, *43*(3), 295-305.

Elbert, M., & Gierut, J. (1986). *Handbook of clinical phonology: Approaches to assessment and treatment*. San Diego, Calif: College Hill Press.

Faircloth, M. A., & Faircloth, S. R. (1970). An analysis of the articulatory behavior of a speech-defective child in connected speech and in isolated-word responses. *Journal of Speech and Hearing Disorders*, *35*(1), 51-61.

Fey, M. E. (1992). Articulation and PhonologyInextricable Constructs in Speech Pathology. *Language, Speech, and Hearing Services in Schools*, *23*(3), 225-232.

Flanagan, J. L. (2013). *Speech analysis synthesis and perception* (Vol. 3). Springer Science & Business Media.

Fudala, J. B., & Reynolds, W. M. (1986). *Arizona Articulation Proficiency Scale: Manual*. Western Psychological Services.

Fudala, J., B. (2000). *Arizona Articulation Proficiency Scale, Third Revision.* LA: Western Psychological Services.

Gangonda, S. S., & Mukherji, D. P. (2012). Speech processing for Marathi numeral recognition using MFCC and DTW features. *International Journal of Engineering Research and Applications (IJERA)*, 218-222.

Geers. (2002). Factors affecting the development of speech, language, and literacy in children with early cochlear implantation. *Language, Speech, and Hearing Services in Schools*, *33*(3), 172–183. http://doi.org/10.1044/0161-1461(2002/015)

Ghule, K. R., & Deshmukh, R. R. (2015). Automatic Speech Recognition System Using MFCC And DTW For Marathi Isolated Words. *International Journal of Technology Enhancements and Emerging Engineering Research (IJTEEE), 3(9),* 11-13.

Gierut, J. A. (1998). Treatment EfficacyFunctional Phonological Disorders in Children. *Journal of Speech, Language, and Hearing Research*, *41*(1), S85-S100.

Goldman, R., & Fristoe, M. (2000). *Goldman- Fristoe Test of Articulation – Second Edition.* Circle Pines, Mn: American Guidance Service, Inc.

Gordon-Brannan, M. E., & Weiss, C. E. (2007). *Clinical Management of Articulatory and Phonologic Disorders*. Lippincott Wouldiams & Wilkins.

Gorin, A. L., Riccardi, G., & Wright, J. H. (1997). How may I help you?. *Speech communication*, *23*(1), 113-127.

Hakkani-Tur, D., Tur, G., Rahim, M., & Riccardi, G. (2004). Unsupervised and active learning in automatic speech recognition for call classification. In *International Conference on Acoustics, Speech, and Signal Processing, 2004, Proceedings (ICASSP'04) IEEE,* 1, 429-432.

Hansen, J. H., Huang, R., Zhou, B., Seadle, M., Deller, J. R., Gurijala, A. R., ... & Angkititrakul, P. (2005). Speechfind: Advances in spoken document retrieval for a national gallery of the spoken word. *IEEE Transactions on Speech and Audio Processing*, *13*(5), 712-730.

Hayes, W., & Pindzola, R., H. (2004). Diagnosis and Evaluation in Speech Pathology. New York. Pearson's Education, Inc.

Hodson, B. W. (1986). *Assessment of phonological processes Revised.* Austin, Tex: Pro-Ed.

Hodson, B. W., Scherz, J. A., & Strattman, K. H. (2002). Evaluating communicative abilities of a highly unintelligible preschooler. *American Journal of Speech-Language Pathology*, *11*(3), 236-242.

Hodson, B. W. (2003). Computerized Analysis of Phonological Patterns (HCAPP). *Wichita, KS: Phonomp Software*.

Holmes, W. (2001). *Speech synthesis and recognition*. CRC press.

http://www.mathworks.com/products/compiler/mcr/index.html

Hummel, L. J., & Prizant, B. M. (1993). A socioemotional perspective for understanding social difficulties of school-age children with language disorders. *Language, Speech, and Hearing Services in Schools*, *24*(4), 216-224.

Kannada. (2016, July 31). In *Wikipedia, The Free Encyclopedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Kannada&oldid=732340964

Kent, R. D., Miolo, G., & Bloedel, S. (1994). The Intelligibility of Children's Speech: A Review of Evaluation Procedures. *American Journal of Speech-Lang Pathology*, *3*(2), 81–95. http://doi.org/10.1044/1058-0360.0302.81

Khan, L. M. (2002). The sixth view: Assessing preschoolers' articulation and phonology from the treaches. *American journal of Speech- Language Pathology,* 11(3), 250-254.

Kinnunen.T. (2003). Spectral features for automatic text independent speaker recognition. *Unpublished Thesis, University of Joensuu, Department of Computer Sciences, Finland.*

Klein, H.B. (1984). Proocedures for maximizing phonological information from single- word responses. *Language, Speech and Hearing Services in Schools*, 15(4), 267-274.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). Prevalence and natural history of primary speech and language delay: findings from a re- cent systematic review of the literature. *International Journal of Language and Communication Disorders*, *35*, 165–188.

Lippmann, R. P. (1997). Speech recognition by machines and humans. *Speech communication*, *22*(1), 1-15.

Lippke, B. A., Dickey, S. E., Selman, J. W., & Soder, A. L. (1997). In M. N. Hegde (2000). *Assessment and treatment of articulation and phonological disorders in children.* United States of America. Pro –ED Inc.

Long, S., & Fey, M. (1988*). Computerized profiling, Version 6.0* [Computer program]. Arcata, CA: Computerized Profiling.

Long, S., Fey, M. E., & Channel, R. W. (2002). Computerized profiling (Version CP941. exe)[computer software]. *Cleveland, OH: Case Western Reserve University*.

Maier, A., Nöth, E., Batliner, A., Nkenke, E., & Schuster, M. (2006). Fully automatic assessment of speech of children with cleft lip and palate. *Informatica-Ljubljana*, *30*(4), 477.

Maier, A. (2009). *Speech of children with cleft lip and palate: Automatic assessment*. *Informatica*. Logos-Verlag.

Masterson, J., & Bernhardt, B. (2001). Computerized articulation and phonology evaluation system (CAPES). *San Antonio, TX: The Psychological Corporation*.

Materson, J., & Pagan, F. (1993). *Interactive system for phonological analysis.* San Antonio, TX: Psychological Corp.

MATLAB Compiler. (n.d.). Retrieved April 14, 2017, from https://in.mathworks.com/products/compiler/mcr.html

Maya, S. (1990). Articulation test battery in Malayalam. *Unpublished Master's Dissertation, University of Mysore, Mysore*.

McLeod, S., Doorn, J. V., & Reed, V.A. (2001). Normal Acquisition of Consonant Clusters. *American Journal of Speech- Language Pathology, 10*, 99-110.

Merin, J & Sreedevi. N (2010). *Computer based assessment of phonological processes in Malayalam. Student research at AIISH*, 8, 167-175, Part-B, Speech Language Pathology, AIISH, Mysore.

Misu, T., & Kawahara, T. (2007). Speech-based interactive information guidance system using question-answering technique. In *International Conference on Acoustics, Speech and Signal Processing, 2007, ICASSP 2007, IEEE,* 4,145-149.

Mohan, B. J., & Babu, N. R. (2014). Speech recognition using MFCC and DTW.

In *International Conference on Advances in Electrical Engineering (ICAEE), 2014,*1-4.

Monrose, F., Reiter, M. K., Li, Q., & Wetzel, S. (2001). Cryptographic key generation from voice. In *Proceedings of the 2001 IEEE Symposium on Security and Privacy, May 2001*, 202-213.

Morrison, J. A., & Shriberg, L. D. (1992). Articulation testing versus conversational speech sampling. *Journal of Speech, Language, and Hearing Research*, *35*(2), 259-273.

Muda, L., Begam, M., & Elamvazuthi, I. (2010). Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. *Journal of Computing, 2* (3), 138-143.

Neenu, S., (2011). Revalidation of the Malayalam diagnostic articulation test (3-4 years). *Student research at AIISH*, 9, 124-132, Part-B, Speech Language Pathology, AIISH, Mysore.

Nethipriya, N. (2007). Aspects of Phonotactics in Typically developing Telugu Speaking Children (3-6 years). Unpublished Master's dissertation submitted in part- fulfilment for the Masters – degree in Speech and Hearing. University of Mysore, Mysore.

Nober, H. (1967). Articulation of the deaf. *Exceptional Children*, *33*, 611–321.

Noyes, J., & Frankish, C. (1992). Speech recognition technology for individuals with disabilities. *Augmentative and Alternative Communication*, *8*(4), 297-303.

Oxford, R. (1991). Language acquisition: The age factor. *System*, *19*(4), 475-477. http://doi.org/10.1016/0346-251X(91)90031-J

Panda, K. (1991). Articulatory proficiency in the Mentally Retarded. Unpublished Master's dissertation submitted in part- fulfilment for the Masters – degree in Speech and Hearing. University of Mysore, Mysore.

Pandit, P., Bhatt, S. (2014) Automatic Speech Recognition of Gujarati digits using Dynamic Time Warping, *International Journal of Engineering and Innovative Technology (IJEIT)*, 3(12), 69-73.

Padmaja, B. (1988). Telugu Articulation Test. *Unpublished Masters Dissertation, University of Mysore*.

Poole, I. (1934). The genetic development of articulation of consonant sounds in children's speech. *Elementary English Review*, *11*, 159–161.

Prather, E., Hedrick, D., & Kern, C. (1975). Articulation development in children aged two to four years. *Journal of Speech and Hearing Disorders*, *40*(2), 179-191.

Prathima, S. (2009). Articulatory acquisition in typically developing native Kannada speaking children. Unpublished Master's dissertation submitted in part- fulfilment for the Masters – degree in Speech and Hearing. University of Mysore, Mysore.

Pye, C. (1987). Pye analysis of language (PAL). *Computer Program*, *200*

Rahul, (2006). Study of Phonological processess in 2-3 year old Hindi speaking children. *Unpublished Masters Dissertation, University of Mysore*.

Ralph Waldo Emerson. (n.d.). BrainyQuote.com. Retrieved April 14, 2017, from BrainyQuote.com Web site: https://www.brainyquote.com/quotes/quotes/r/ralphwaldo397441.html

Rama Devi, K. J. S. (2006). Phonological profile in Kannada: A study on hearing impaired. *Unpublished doctoral thesis, University of Mysore*.

Ratanamahatana, C. A., & Keogh, E. (2004). Everything you know about dynamic time warping is wrong. In 3rd *Workshop on Mining Temporal and Sequential Data*. Seattle, WA.

Rupela, V., & Manjula, R. (2007). Phonotactic patterns in the speech of children with Down syndrome. *Clinical linguistics & phonetics*, *21*(8), 605-622.

Roberts, J. E., Burchinal, M., & Footo, M. M. (1990). Phonological process decline from 212 to 8 years. *Journal of Communication Disorders*, *23*(3), 205-217.

Saastamoinen, J., Fiedler, Z., Kinnunen, T., & Fränti, P. (2005). On factors affecting MFCC-based speaker recognition accuracy. In *International Conference on Speech and Computer (SPECOM'2005), Patras, Greece* ,503-506.

Sander, E. K. (1972). When are speech sounds learned? *Journal of Speech and Hearing Disorders*, *37*(1), 55–63.

Shanks, S. J., Sharpe, M. R., & Jackson, B. R. (1970). Spontaneous responses of first grade children to diagnostic picture articulation tests. *Journal of Communication Disorders*, *3*(2), 106-117.

Skahan, S. M., Watson, M., & Lof, G. L. (2007). Speech-language pathologists' assessment practices for children with suspected speech sound disorders: Results of a national survey. *American Journal of Speech-Language Pathology*, *16*(3), 246-259.

Sherman, D., & Geith, A. (1967). Speech sound discrimination and articulation skill. *Journal of Speech, Language, and Hearing Research*.

Shriberg, L. D., & Kwiatkowski, J. (1982). Phonological disorders I: a diagnostic classification system. *Journal of Speech and Hearing Disorder*, *47*(3), 226–41.

Shriberg, L. D. (1986). *PEPPER [Programs to Examine Phonetic and Phonologic Evaluation Records*. Software Development and Distribution Center. University of Wisconsin-Madison.

Singh, S., & Rajan, E. G. (2011). MFCC VQ based speaker recognition and its accuracy affecting factors. *International Journal of Computer Applications*, *21*(6), 1-6.

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa Articulation Norms Project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, 55(4), 779–798. http://doi.org/10.1044/jshd.5504.779

Sneha, B. (2012). Cluster Acquisition in typically Developing Telugu Speaking Children: 4-6 years. Unpublished Master's dissertation submitted in part- fulfillment for the Masters – degree in Speech and Hearing. University of Mysore, Mysore.

Spriestersbach, D. C., & Curtis, J. F. (1951). Misartigulation and discrimination of speech sounds. *Quarterly Journal of Speech*, *37*(4), 483–491. http://doi.org/10.1080/00335635109381704

Sreedevi, S. V. (1976). Aspects of acquisition in Kannada by+ 2 year old children. *Research at AIISH Dissertation Abstracts Vol*, *1*, 77-79.

Sreedevi, N., & Shilpashree, H. N, (2008). Phonological processes in typically developing Kannada speaking children. *JAIISH*, 27, 83-88.

Sreedevi, N. (2008). Study of phonological processes in normal Kannada speaking children: 1.6- 2 years. *Interdisciplinary Journal of Linguistics*, 1, 103-111.

Sreedevi, N., John, M., & Chandran, S. (2013). Computerized assessment of phonological processes in Malayalam (CAPP-M). In *Global Humanitarian Technology Conference: South Asia Satellite, GHTC-SAS 2013 IEEE* (pp. 264–269).

Stoel-Gammon, C. (1985). Phonetic Inventories, 15–24 MonthsA Longitudinal Study. *Journal of Speech, Language, and Hearing Research*, *28*(4), 505-512.

Stoel-Gammon, C., & Dunn, C. (1985). Normal and disordered phonology in children. Baltimore,University Park Press.

Swamy, S., & Ramakrishnan, K. V. (2013). An efficient speech recognition system. *Computer Science & Engineering*, *3*(4), 21.

Templin, M. C. (1957). *Certain Language Skills in Children: Their Development and Interrelationships*. http://psycnet.apa.org/psycinfo/1957-07556-000

Templin, M. C., & Darley, F. L. (1969). *Templin-Darley tests of articulation*. University of Iowa, Division of extension and University services.

Thirumalai (1972). Acquisition of phonology of a child. Unpublished Master's dissertation submitted in part- fulfillment for the Masters – degree in Speech and Hearing. University of Mysore, Mysore.

Tyler, A. A., & Tolbert, L. C. (2002). Speech-language assessment in the clinical setting. *American Journal of Speech-Language Pathology*, *11*(3), 215-220.

The MathWorks Inc. *(2015) MATLAB* (R2015b).

Usha, D. (1986). A test of articulation in Tamil. *Unpublished Masters Dissertation, University of Mysore*.

Usha Rani, K. (2010). Articulatory Acquisition in typically developing Telgu speaking children: 2-3 years. *Student research at AIISH*, 8, 291-299, Part-B, Speech Language Pathology, AIISH, Mysore.

Van Riper, C., & Irwin, J.V. (1958), *Voice and Articulation*, Englewood Cliffs, NJ: Prentice Hall.

Van Riper, C., & Erickson, R. L. (1996). Speech Correction: An Introduction to Speech Pathology and Audiology. Needham Heights: Ally & Bacon.

Van Riper, C., & Erickson, R. (1996). Speech Corrections: An Introduction to Speech-Pathology and Audiology (8^{th} ed). Englewood Clifs, NJ: Prentice-Hall, Inc.

Vani, R. (2011). Revalidation of Malayalam diagnostic articulation test (4-5 years). *Student research at AIISH*, 9, 279-289. Part-B, Speech Language Pathology, AIISH, Mysore.

Vashisht, D., Sharma, S., & Dogra, L. Design of MFCC and DTW for Robust Speaker Recognition. *International Journal of Electrical & Electronics Engineering, 2(3),* 12-17.

Vipina, V. P., (2011). Revalidation of the Malayalam diagnostic articulation test (4-5 years). *Student research at AIISH. 9,* 279-289, part- B, Speech Language Pathology, AIISH, Mysore.

Vrinda, R. (2011). Revalidation of Malayalam diagnostic articulation test (5-6 years). *Student research at AIISH*, 9, 290-301,Part-B, Speech Language Pathology, AIISH, Mysore.

Wang, J. C., Lee, H. P., Wang, J. F., & Lin, C. B. (2008). Robust environmental sound recognition for home automation. *IEEE transactions on automation science and engineering*, *5*(1), 25-31.

Watson, M. M., & Scukanec, G. P. (1997). Profiling the phonological abilities of 2-year-olds: A longitudinal investigation. *Child Language Teaching and Therapy*, *13*(1), 3-14.

Wardill, W. E. M., & Whillis, J. (1936). Movements of the soft palate. *Surg. Gynec. Obstet*, *62*, 836-839.

Weiner, P. S. (1986). The study of childhood language disorders: Nineteenth century perspectives. *Journal of communication disorders, 19(1),* 1-47.

Weiss, C. E (1980). *Weiss Comprehensive Articulation Test.* Autisn, Tx: Pro-Ed.

Wellman, B. L., Case, I. M., Mengert, I. G., & Bradbury, D. E. (1931). Speech sounds of young children. *University of Iowa Studies: Child Welfare, 5*(2).

Wolk, L., & Meisler, A. W. (1998). Phonological assessment: A systematic comparison of conversation and picture naming. *Journal of Communication Disorders*, *31*(4), 291-313.

Winitz, H. (1969). *Articulatory acquisition and behavior*. New York: Appleton-Century drofts.

Young, V., & Mihailidis, A. (2010). Difficulties in automatic speech recognition of dysarthric speakers and implications for speech-based applications used by the elderly: A literature review. *Assistive Technology*, *22*(2), 99-112.

Zissman, M. A. (1996). Comparison of four approaches to automatic language identification of telephone speech. *IEEE Transactions on speech and audio processing*, *4*(1), 31.

COMPUTERISED ARTICULATION TEST-SCREENNG (CAT's)

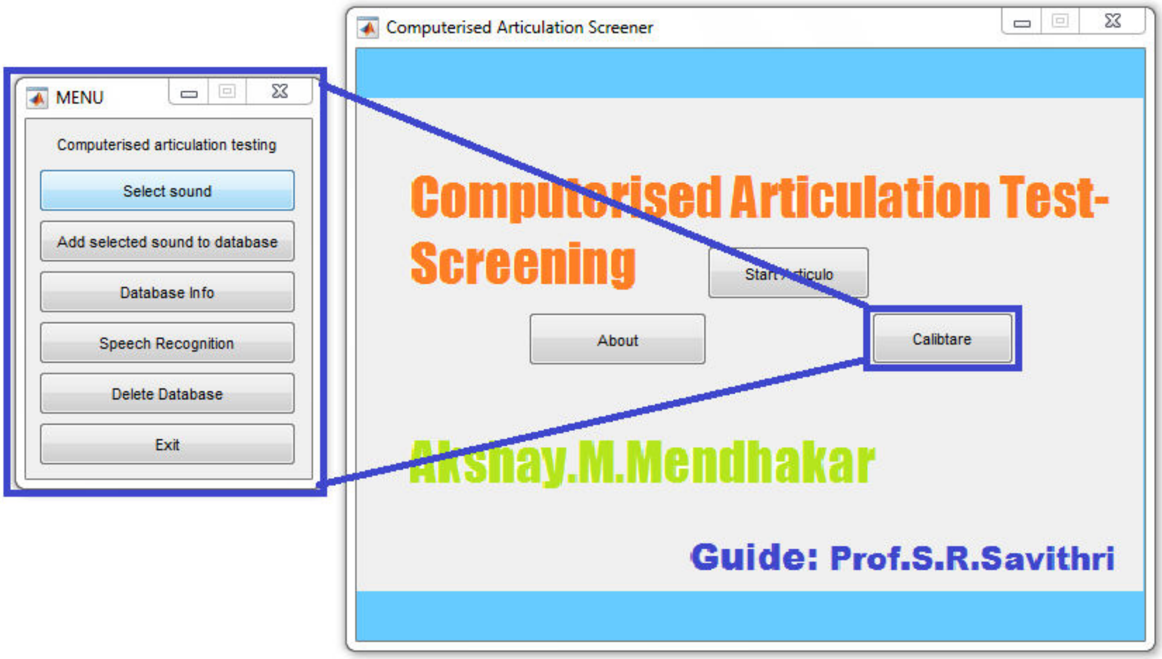
The software is available on the CD and can be downloaded onto any system just by double clicking on the "Cat's_pkg.exe"

A self guiding installation procedure will guide you through the installation of the software.

For any issues/ queries please contact the authors.

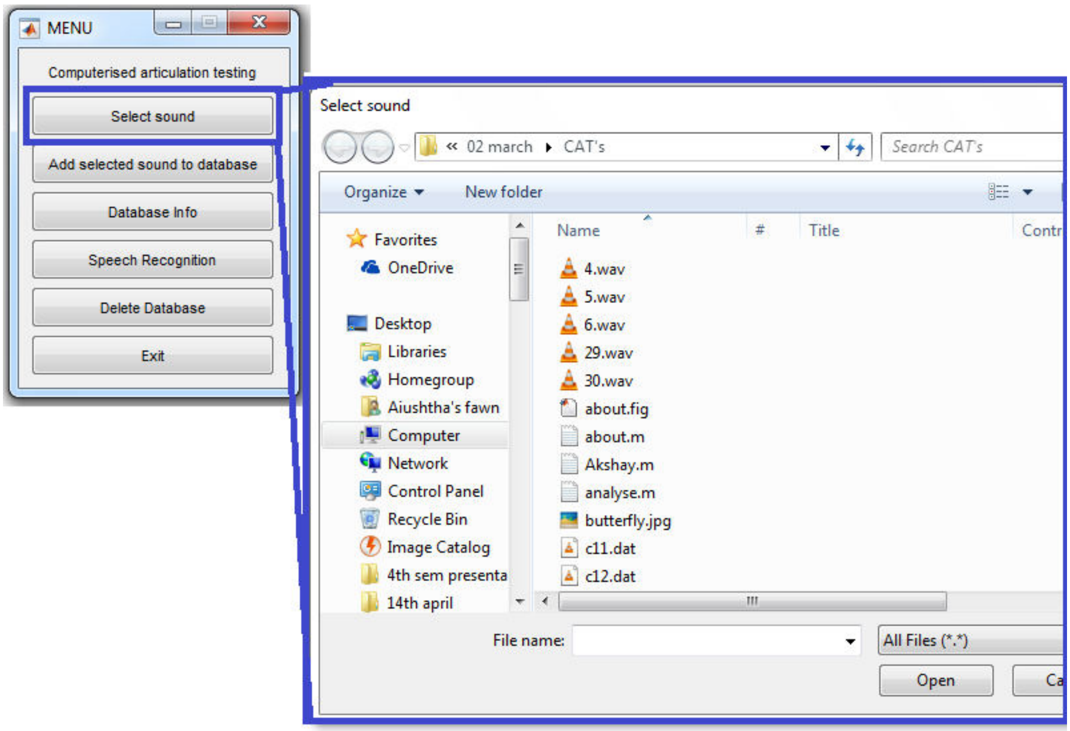
CREATING DATABASE USING CAT's SOFTWARE

Step 1: Select the **"Calibrate"** button from the Main screen of CAT's software. A pop up window as depicted below will open up.

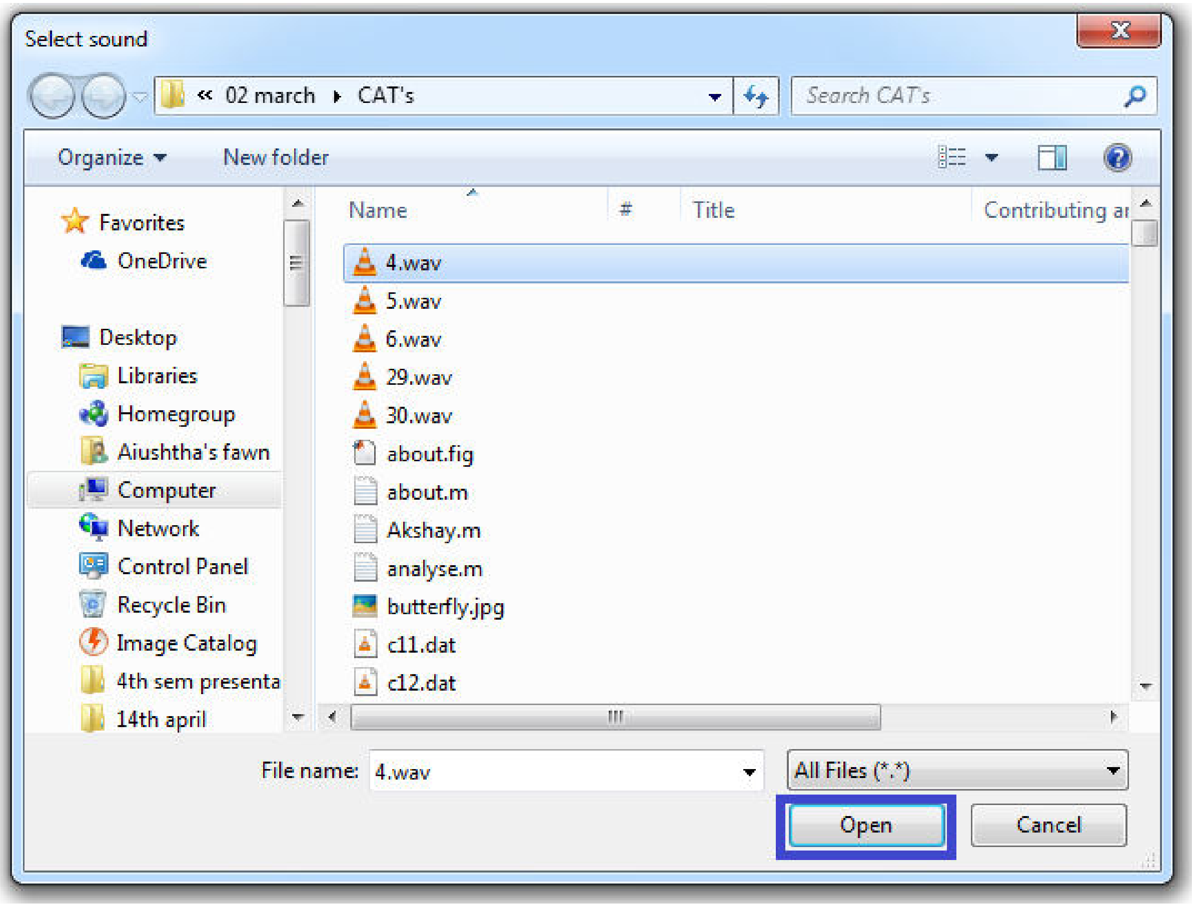


Step 2: Pick the **"Select sound"** button to choose the audio file which you want to add to the database.

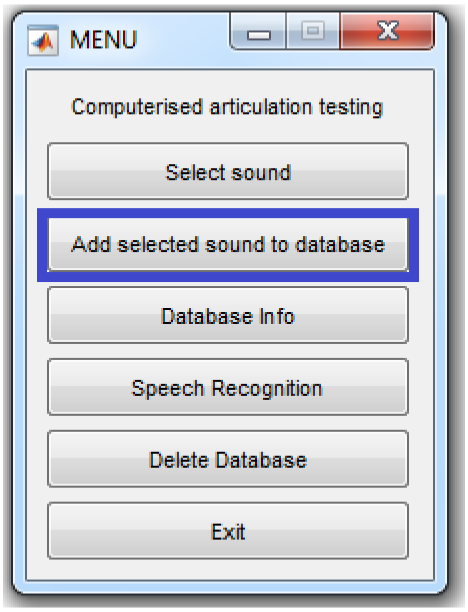
A pop up window opening the current directory as shown in the figure below will appear.



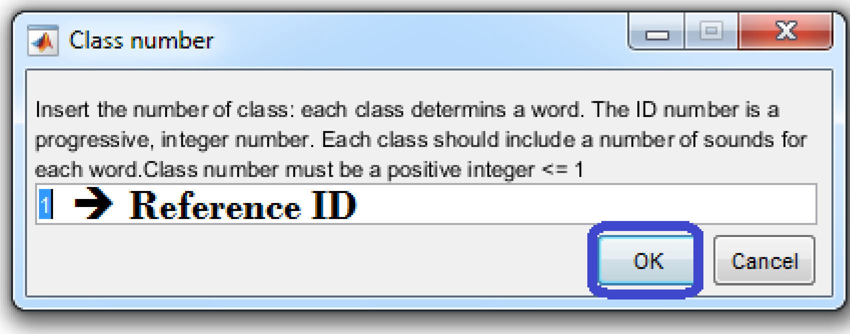
Step 3: Select the file you want to put in your database and click **"Open".**



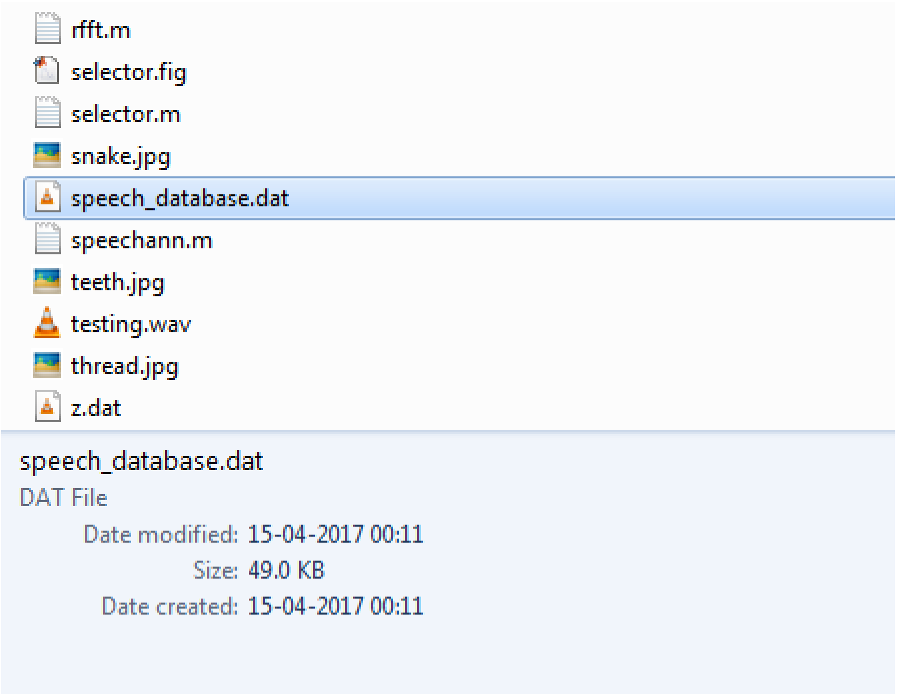
Step 4: Select the **"Add selected sound to database"** button to add the selected sound to the database.



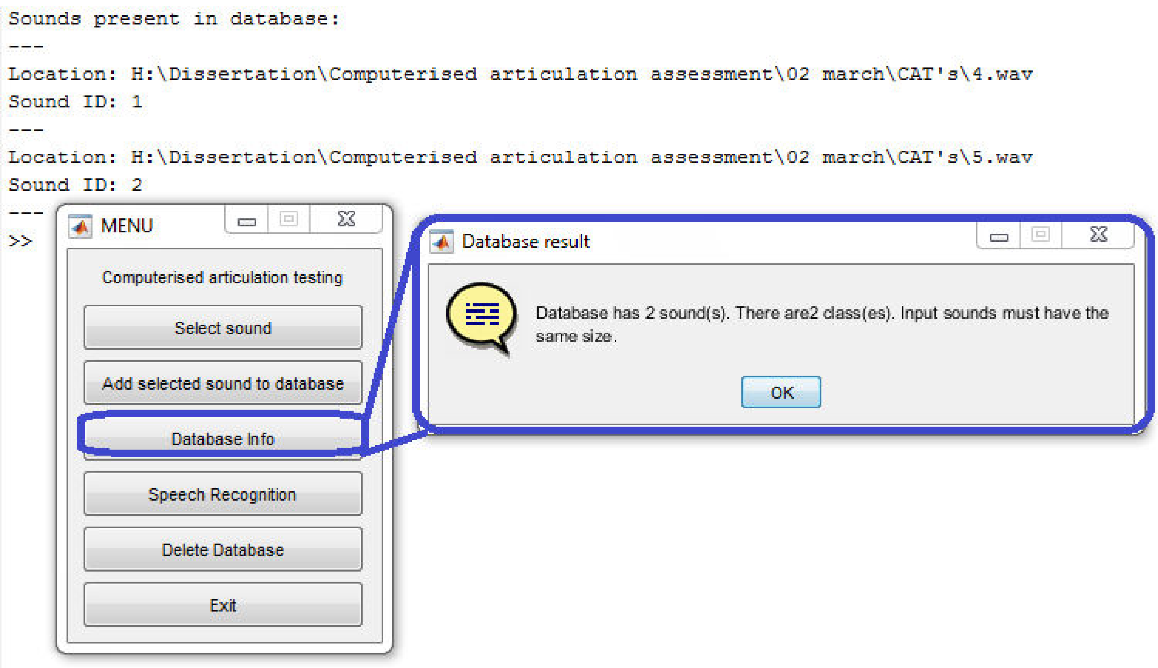
A pop up window **"class number"** is displayed, which contains a **reference ID**, Provide an appropriate ID (Positive integers' only). Followed by click on the **"OK"** button



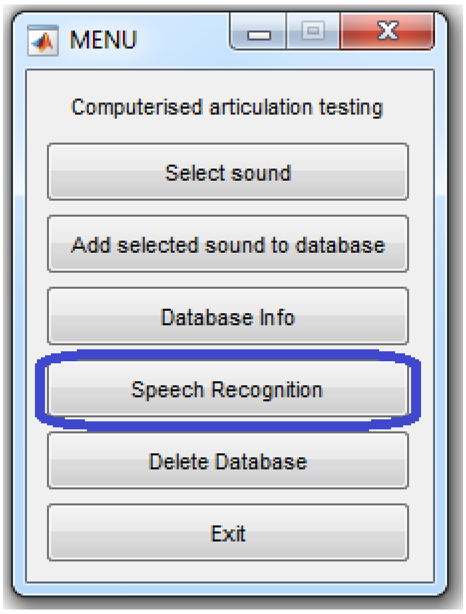
NOTE: A dataset file (.dat) file will be added into the current directory of the software. A default name of "**speech_databade.dat**" will be saved.

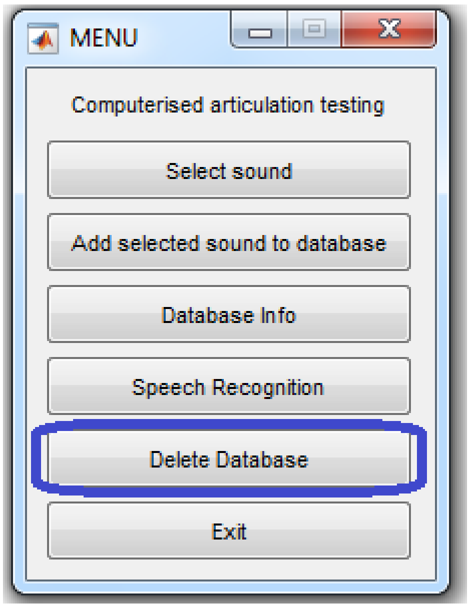


Step 5: Similarly 'n' number of sounds can be added on to the database and the database can be viewed by clicking on **"Database Info"**



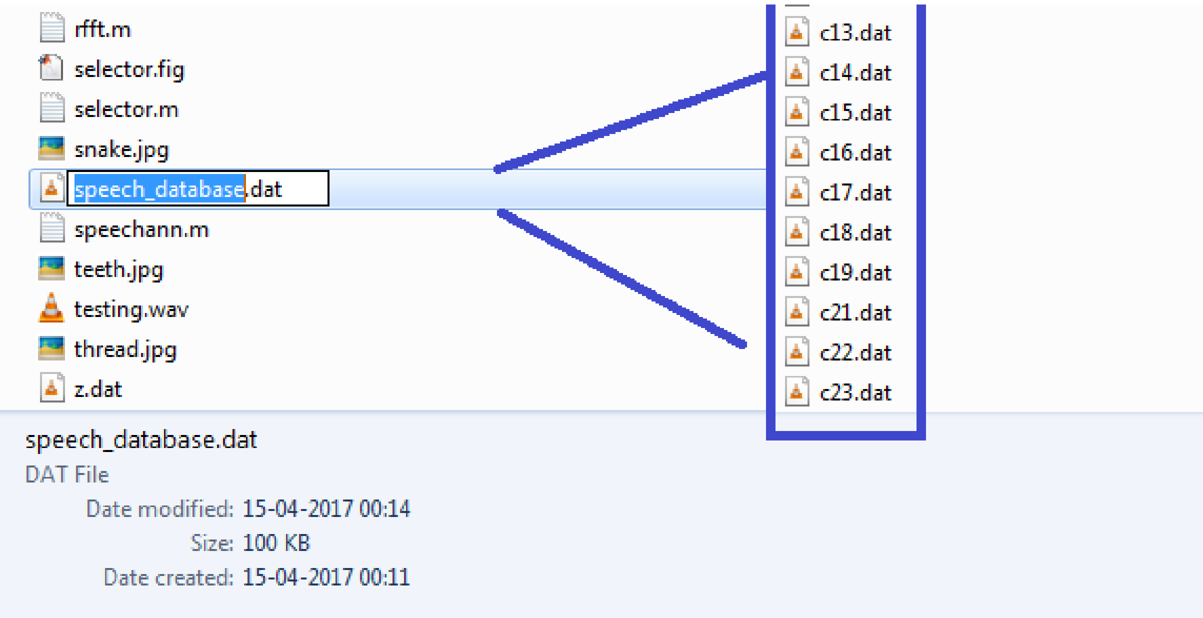
Step 6: one can choose to do direct analysis by comparing the selected file and clicking on **"Speech recognition"** button to instantly compare selected file from the database. Similarly database can be deleted by clicking on **"Delete Database"** button





NOTE: The software by default will search for a database named " **speech_database.dat**" and perform all operations selected on that file itself.

One can also choose to rename databases to create "n" number of databases of your choice by simply renaming the file.



EDITING THE (CAT's) SOFTWARE CODES

The source codes of the software are available online. Please mail the authors'' for a copy on amendhakar@gmail.com, savithri486@gmail.com