Durational Characteristics of Allophonic Variations in Malayalam Vowel Phonemes

Vivek P[1], Sandesh E PA[2], Lajish V L[3]
Department of computer science
University of Calicut
Kerala, India
([1]vivek_dcs@uoc.ac.in, [2]sandeshepa@gmail.com, [3]lajish@uoc.ac.in)

Abstract—Phonemes are considered to be the basic unit for speech processing. But in actual utterances, each phoneme is manifested differently in different contexts. The effect due to neighboring phonemes and position are generally known as co-articulation effects. Such effects are observed to be encoded in the allophone characterization of Malayalam, a major Dravidian language spoken by around 38 million people. A well-defined allophone formation rule set exists for Malayalam. In this work long and short vowel allophones in Malayalam are identified and listed. Detailed analysis on the durational characteristics are performed on the allophonic inventory. The work establishes the allophonic variations in durational properties of vowel phonemes in Malayalam. This can be considered as a first attempt towards an allophone based paradigm shift in Malayalam audio visual speech processing.

Keywords—*Phoneme; Allophone; Vowel; Duration; Speech Recognition; Speech Synthesis.*

I. INTRODUCTION

Malayalam, the mother tongue of the south Indian state Kerala is spoken by more than 40 million people in Kerala, Lakshadweep, Mahe etc. [1]. It is the youngest in the Dravidian language family and conferred the classical language status by the Government of India in 2013.It is included in the list of top 30 languages spoken in the world and top 10 languages in India. The dual rooted evolution of Malayalam from Sanskrit and Tamil make it unique in Indian languages [2].

Speech recognition and speech synthesis in Malayalam is an active research area [3-5]. But the attempts have not yet reached the realm of successful development of Malayalam speech technology solutions that are useful for the masses. More study is needed on the basic structure and peculiarities of the language in the perspective of Malayalam language computing. Malayalam is an alpha syllabary language with the aksharam (character) as its core. Almost one to one correspondence exists between orthographic symbols in the alphabet and phoneme in Malayalam [6].

Phones are considered to be the basic unit in speech and a phone is a physical sound produced when a particular phoneme is articulated. There will be an infinite number of phones corresponding to a phoneme due to the underlying variability of vocal tract configurations. Allophone is a class of phones corresponding to a specific variant of a phoneme. The utterance of a phoneme is affected by the context it appears. This co-articulation effect is mainly responsible for allophonic variations. Forward or anticipatory co-articulation refers to the changes due to anticipatory phonemes. Backward or preservatory co-articulation is caused by the inertia of the uttered phonemes [7]. It is evident that each phoneme existing in speech will be in any of its allophonic forms. Co-articulation happens either due to the inertia of lip, jaw and tongue or due to the low level neural processes within the brain [8].

Co-articulation effects are modelled using allophonic models, target based models or hierarchical models [9]. A well-defined rule-based structure exists for almost each allophone in Malayalam [2, 6]. So an allophone based frame work is the natural choice to model the co articulation effect in Malayalam. These contextual rules explain the allophonic variability in terms of the place of occurrence of the phone (including initial, middle and final) and the neighboring phonemes. These rules can be effectively used for speech synthesis process where concatenating allophones are employed. The sequence of graphical signs named as graphemes, can be converted to the corresponding sequence of allophones. To the best of our knowledge, no detailed study has been reported in Malayalam which places allophones as the basic unit in speech synthesis and recognition.

Allophone based speech processing studies are reported in many languages. Piotr Kozierski et. al used allophones instead of phonemes for polish language speech recognition [10]. Here, instead of using the entire set of allophones, a proper subset of allophones is used based on the frequency of occurrence. Some rarely occurring allophones are omitted in their study. A variation of the same is also used for the Japanese language by Long Nguyen et. al [11]. Allophonic properties are explored in Korean speech recognition also in a work reported by Xu, Ji et. al [12]. Much concatenative text to speech synthesis systems uses allophones as the basic unit. Imedjdouben et. al introduced a phone to allophone conversion algorithm for speech synthesis dedicated to the Arabic language [13]. A report by Pavel A. Skrelin describes the principles of the allophone extraction from Russian natural speech flow, ways of forming synthesized speech, modification of the acoustic parameters [14]. Barkhoda et. al implemented three synthesis systems for the Kurdish language based on syllable, allophone, and diphone and showed all systems' Intelligibility is acceptable using various tests [15]. Cawley, G. C., and P. D. Noakes attempted allophone based concatinative synthesis using neural networks [8]. Mazin et al. uses allophone/ diphone concatenation method for speech synthesis [16]. Alexey Karpov et al. used allophones and multi allophones for text to speech synthesis system [17]. An allophone rule set based strategy is employed by Ka-Ho Wongto et al. for developing a language learning framework in 2011 [18]. Gregor A. Kalberer et al. has used an allophone-viseme transcription system for visual speech synthesis [19].

As indicated by Louis C.W. Pols et al. the recognition performance can be further improved by incorporating “specific knowledge” (such as duration and pitch) into the recognizer [20]. They conducted a detailed study on phone duration modeling and its potential benefit on ASR. Durational models are developed in many languages to address the dynamic nature of phoneme duration. Rule-based approaches built on experimental data and model-based approaches are reported in various languages such as English, Korean, Turkish and Czech [21,22,23,24]. A Classification and Regression Tree CART) approach is used by Alexandros et al. [25] to model the duration of greek phonemes. A Long Short Term Memory Recurrent Neural Network framework, which tries to model the countable duration values are used by Bo Chenn et al. [26]. A neural network based Arabic phoneme prediction system is proposed by Yasser Hifny et al. [27]. Giedrius performs a decision tree based phoneme duration analysis for Lithuanian language [28]. Janne et al. employed an expanded state HMM for modelling Finnish language duration modelling [29]. To improve the intelligibility of synthesised speech an HMM and Multilayer Perceptron hybrid model duration prediction is introduced by kalu et al. [30]. Tree based machine learning approaches is also used for Serbian language duration modelling [31].

Durational modeling and characteristic analysis were performed in various Indian languages. Samudravijaya, K. studied the durational characteristics of Hindi phonemes as well as stop consonants in detail [32, 33]. K. Sreenivasa Rao and B. Yegnanarayana proposed a syllable duration prediction system for Indian languages. The analysis is performed in Hindi, Telugu and Tamil languages, in order to predict the duration of syllables in these languages using SVM regression model [34]. N. Sridhar Krishna et. al reported a preliminary attempt on data-driven modeling of segmental (phoneme) duration for two Indian languages Hindi and Telugu [35]. S. R. Savithri identified some of the variables influencing the durations of Kannada vowels in the initial position [36]. Deepa P. Gopinath et al. proposed a preliminary Malayalam phoneme duration model for speech synthesis system [37]. Bindhu K Rajan et al. proposed Malayalam phoneme duration modeling based on Classification and Regression Tree (CART) for synthesis application [38].

Malayalam is one of the few languages in which allophone formations due to contextual and positional variability can be explained in a rigorous rule based approach. So the co articulation effects can be modelled in Malayalam speech using an allophone centric approach. An allophone based durational model can improve the accuracy of speech recognition systems and intelligibility of synthesised speech in Malayalam. This work progress towards developing an allophone based durational model in Malayalam which accommodates the contextual and positional variability of phonemes in actual utterances. In this work, long and short vowel allophones in Malayalam are identified and listed with its allophonic variations. Section II describes the properties of Malayalam phones and allophones. Section III discuss the process of finding the extensive rule set for the formation of Malayalam vowel allophones. Section IV introduces the TEMU Malayalam phonetic archive. Section V discusses durational properties of Malayalam vowel allophones and section VI concludes the work.

II. MALAYALAM PHONEME SET AND ALLOPHONES

A phone refer to the instances of phonemes in the actual utterances. The smallest meaningful distinctive sound unit in a language is called a phoneme. The phoneme set varies considerably from one language to another. International Phonetic Alphabet (IPA) defines around 150 phones among all languages. American English has around 40 phonemes. Malayalam has 11 vowel phonemes, 2 diphthongs, and 37 consonant phonemes together constitute a 50 member phonemeset [13]. The following section describes the characteristics of phonetic inventory of Malayalam language and its allophonic variations in detail followed by the structure of the Malayalam phonetic archive used in this study.

A. Malayalam vowel phones

The vowels are produced by exciting an essentially fixed vocal tract shape with quasi-periodic pulses of air caused by the vibration of the vocal cords [16]. The sound difference is produced by changing the shape and position of lip and tongue. Diphthongs are considered as the combination of vowels as an in-between smooth transition happens between the vowel configurations. Malayalam has 11 monophthongs and 2 diphthongs [27]. The list of Malayalam vowels and diphthongs with its allophones is shown in fig 1. Based on articulation, among the 10 vowels 4 are classified as front vowels (ഇ [i], ഈ [i:], എ [e], ഏ [e:]) two as central vowels (അ [a], ആ [a:]) and four as back vowels (ഉ [u], ഊ [u:], ഒ [o], ഓ [o:]). Ideally, the frequently used semivowel (˘ /ə/) can either be treated as a separate phoneme or an allophonic variation of ഉ /u/. In this work we have considered ˘ /ə/ as an allophone of /u/. Another sound ഋ /r̥/ is not included in the study as it is rarely used in normal speech.

As part of the study, 107 allophones in Malayalam which include 76 consonant allophones, 28 vowel allophones and 3 allophones corresponding to diphthongs are identified. The vowels have relatively high degree of allophonic variability. It is also reported that, while comparing with the ideal position in the cardinal vowel figure, Malayalam vowels are slightly displaced towards the inside from the vowel area borders [6]. We have selectively used 238 Malayalam words accommodating all vowel allophonic variability for further analysis.

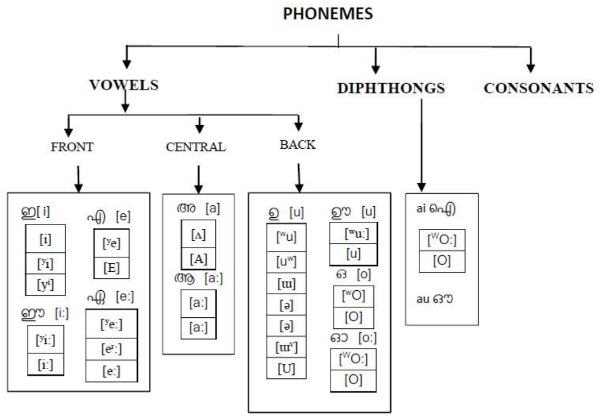


FIGURE I. MALAYALAM VOWELS AND DIPHTHONGS AND ITS ALLOPHONIC VARIATIONS

III. AN EXTENSIVE RULE SET FOR THE MALAYALAM VOWEL ALLOPHONE FORMATION

The variation in the duration of a phoneme can be attributed to many factors. Some factors such as contextual variability can be detected from the text while some others such as dialect cannot be detected from the text. This work addresses the durational variability of vowel allophones in Malayalam. In their studies, Asher and V.R. Prabodhachandran Nair described the rules of Malayalam allophone formation [2,6]. They have proposed linguistic descriptions for defining the allophones of each Malayalam phones. It reveals that most such findings can be converted to position and neighborhood-based rule set. Thus certain rule set for Malayalam vowel allophones based on position and neighboring information’s are created. The following section describes the rule set for the formation of allophones in detail.

A. Rule set for Malayalam vowel allophones based on the position and neighbourhood information

A rule set for the formation of vowel allophones in Malayalam based on the position and neighboring information are formed. In the case of vowel ഇ/i/ the three allophones [I],[^{y}i] and [i^{y}] correspond to medieval, initial and final position of occurrence of the phone respectively. The vowel phone, ഉ /u/ has 6 allophones in Malayalam. The allophone [^{W}u] is characterized by the initial position and the second allophone [u^{w}] is characterized by the presence of off-glide in the final position. The third allophone [ɯ] occurs in the middle syllable. The fourth allophone [ɯ^{y}] occurs in Sanskrit originated words after certain phonemes. The fifth allophone [U] happens if preceded by an initial consonant in a word and finally the last two allophones correspond to the semivowel [ə]. In general, these allophonic categorizations in Malayalam is sufficient to capture the variations due to place and co-articulation effects. Table II represents these newly framed rule set to denote the formation of Malayalam vowel allophones.

TABLE I. POSITION AND NEIGHBOURHOOD BASED RULE SET FOR MALAYALAM VOWEL ALLOPHONES

Sl. No.	*Phoneme*	*Allophone*	*Position*	*Rule*
1	ഇ [i]	[i]	Middle	Metadata: Low high front unrounded short vocoid. Neighborhood: Any
		[^{y}i]	Initial	Metadata: High front unrounded long tense vocoid with onglide. Neighborhood: Any
		[y^{i}]	Final	Metadata: High front unrounded short tense vocoid with offglide. Neighborhood: Any
2	ഈ [i:]	[^{y}i:]	Initial	Metadata: High front unrounded long tense vocoid with onglide. Neighborhood: Any
		[i:]	Middle	Metadata: High front unrounded tense long vocoid; medially. Neighborhood: Any
3	എ [e]	[^{y}e]	Initial	Metadata: Higher mid front unrounded short tense vocoid with onglide. Neighborhood: Any
		[^{e}y]	Initial	Metadata: Higher mid front unrounded short tense vocoid with offglide[j]. Neighborhood: Any
		[E]	Middle	Metadata: Mean mid front unrounded short vocoid. Neighborhood: Any
4	ഏ [e:]	[^{y}e:]	Initial	Metadata: Higher mid front unrounded long tense vocoid with onglide. Neighborhood: Any
		[e^{r}:]	Final	Metadata: Higher mid front unrounded long tense vocoid with offglide [j]. Neighborhood: Any
		[e:]	Middle	Metadata: Higher mid front unrounded long tense vocoid. Neighborhood: Any
5	അ [a]	[ʌ]	Inital& Final	Metadata :Low mid back vocoid in the initial syllable and word. Neighborhood: Any
		[A]	Middle	Metadata:Low mid central vocoid in the medial syllable. Neighborhood: Any
6	ആ [a:]	[a:]	-	Metadata: Low back long tense vocoid after velar consonants. Neighborhood: Left : Velar Consonants
		[a:]	-	Metadata: Low central long vocoid after all non-velaric consonants. Neighborhood: Left :Non-velar Consonants
7	ഉ [u]	[^{w}u]	Initial	Metadata: High back rounded tense short vocoid with onglide [w]. Neighborhood: Any
		[u^{w}]	Final	Metadata:Higher back rounded tense short vocoid with offglide [w]. Neighborhood: Any
		[ɯ]	Middle	Metadata: High back unrounded short vociod in the

Sl. No.	*Phoneme*	*Allophone*	*Position*	*Rule*
				medial syllable. Neighborhood: Any
		[ə]	Final	Metadata:Higher mid central unrounded open vocoid, between consonant and vowel in the word boundary with open juncture. Neighborhood: Any Other: Open juncture.
		[ə*]	-	Metadata: In open juncture 'ə' is in free variation after the following some phonemes. Neighborhood: [ɳ] [ɭ] [ṛ] [l]]
		[ɯ^{v}]	-	Metadata: Low high back unrounded vocoid after some phonemes. Neighborhood: Right Labial, dental, palatal, and velar plosives and labial, and dental nasals and non-retroflex fricatives and labio-dental continuants. Other:In Sanskrit words
		[U]	-	Metadata: Low high back rounded tense vocoid, after word initial consonant. Neighborhood: Right :Consonant as the first letter in a word
8	ഊ /u:/	[^{w}u:]	Initial	Metadata; High back rounded long tense vocoid with onglide [w]. Neighborhood: Any
		[u]	Other than Initial	Metadata: High back rounded tense vocoid, elsewhere. Neighborhood: Any
9	ഒ /o/	[^{w}O]	Initial	Metadata:Higher mid back rounded tense short vocoid with onglide [w] in the initial position. Neighborhood: Any
		[O]	Middle	Metadata:Mean mid back tense rounded short vocoid. Neighborhood: Any
10	ഓ /o:/	[^{w}O:]	Initial	Metadata: Higher mid back rounded tense long vocoid with onglide. Neighbourhood:Any
		[O]	Medial and Final	Metadata: Higher mid back tense long vocoid. Neighborhood: Any

IV. TEMU MALAYALAM PHONETIC ARCHIVE

In this work, an inclusive Malayalam phonetic data set which is being designed and developed as part of the Malayalam phonetic archive project owned by Thunchath Ezhuthachan Malayalam University (TEMU), Kerala, India is used for experimental purposes [39]. It is a fairly comprehensive database created by taking into consideration of a carefully compiled inventory of phones which are currently employed in the Malayalam language. Malayalam phoneme segments are recorded in its standardized orthography followed by a number of examples of its occurrence in phonologically relevant different positions. Allophones are listed together and pronunciation of each example recorded from the natural speech is demonstrated in both male and female voices. The data comprises of 11 vowels, 2 diphthongs and 38 consonants, and its allophonic variation with 900 spoken words as examples. The following section describes the durational properties of the Malayalam vowel allophones derived on the basis of the detailed analysis conducted on the TEMU dataset. This archive is presently available in public domain under creative commons license. The dataset is archived and published in web portal [39].

V. DURATIONAL PROPERTIES OF MALAYALAM VOWEL ALLOPHONES

This section describes a detailed analysis of durational properties of Malayalam vowel allophones. These findings could be the basis for preliminary work to incorporate duration-specific knowledge to a speech recognition or a speech synthesis system. Phoneme segmentation is the most important pre-processing in the phoneme level speech recognition. In phoneme segmentation algorithms, mostly the phonemes are assumed to be of the same length and segmented using a fixed size window. Durational analysis of phonemes performed in many languages reveals the variability in the duration of individual phonemes [20]. So the phone segmentation algorithms must consider the variability in phoneme duration for a better result. The phoneme level duration variability is language specific. Considering these facts a detailed analysis is conducted to establish durational phoneme models for the Malayalam language.

The duration of each individual vowel phones and its allophonic examples are computed from the utterances of the same phone by different male and female speakers. Figure II shows the spectral representation indicating the duration of the allophone [^{y}e] of the vowel എ /e/ obtained from the word എലി [eli]. It is observed that, this duration computed for the allophone varies from the duration of its other allophones [^{e}y] and [E] of the same vowelഎ /e/.

We have selectively used 238 Malayalam words from TEMU dataset accommodating all vowel allophonic variability for the conduct of the experiments. The average durations of all allophonic variations of ten Malayalam vowels are computed distinctly from the selected set of words comprising that allophone taken from the TEMU dataset. Figure III shows the average durations of seven different allophones of the vowel ഉ /u/ computed from the selected words comprising the allophones. From the figure, it is evident that there exist significant variability in the duration of different allophones of the vowel phonemeഉ /u/. The average durations of all allophonic variations of 10 Malayalam vowels are computed separately from the selected word groups comprising all the vowel allophones with respect to each vowel.

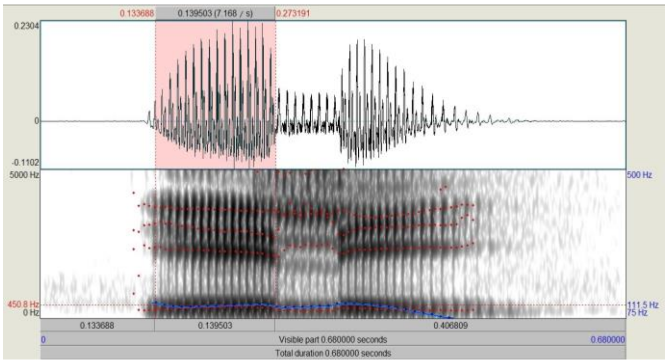


FIGURE II: DURATION OF THE ALLOPHONE [ʸE] FROM THE WORD എലി[ELI].

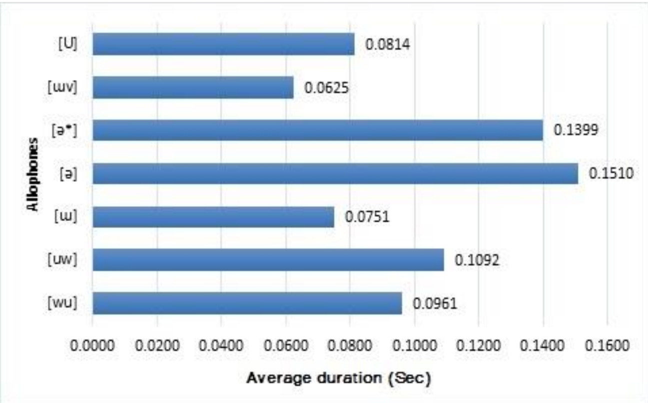


FIGURE III: AVERAGE DURATIONS OF SEVEN ALLOPHONES OF THE VOWEL /U/

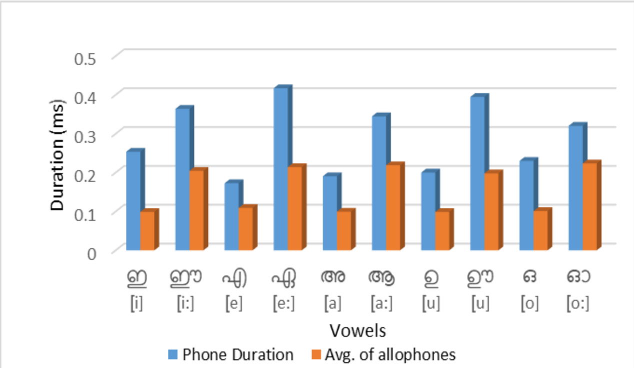
The mean duration obtained for the ten Malayalam vowels including all the allophonic variations is 0.14284 sec for the male speaker and 0.15563 sec for the female speaker. The duration range is from 0.04155 sec to 0.28936 sec for the male, and 0.04316 sec to 0.33001 sec for the female speaker. The detailed statistics including the average duration of allophones corresponding to each Malayalam vowel together with the Mean and Standard Deviation (SD) are computed and listed in table II. The speech samples from the TEMU dataset is used to conduct this analysis.

TABLE II. DURATIONAL STATISTICS OF MALAYALAM VOWEL ALLOPHONES

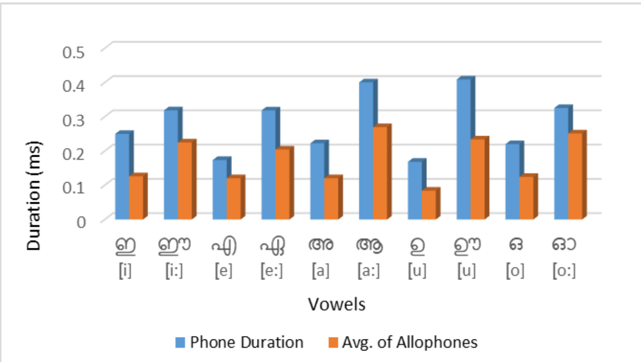
Sl.No	Vowel	Allophone	Average Duration (sec)	
			Male	Female
1	ഇ [i]	[i]	0.08871	0.12866
		[yi]	0.11778	0.14847
		[yi]	0.09356	0.10311
		Mean	0.09871	0.12617
		SD	0.01758	0.02210
2	ഈ [i:]	[yi:]	0.20628	0.22192
		[i:]	0.20229	0.22763
		Mean	0.20429	0.22478
		SD	0.03948	0.04673
3	എ [e]	[ye]	0.12792	0.14015
		[E]	0.08989	0.10110
		Mean	0.10890	0.12062
		SD	0.02454	0.02778
4	ഏ [e:]	[ye:]	0.24340	0.25133
		[er:]	0.20149	0.11787
		[e:]	0.20317	0.24064
		Mean	0.21391	0.20435
		SD	0.03674	0.06900
5	അ [a]	[ʌ]	0.11452	0.15527
		[A]	0.08414	0.08588
		Mean	0.09933	0.12057
		SD	0.01819	0.03656
6	ആ [a:]	[a:]	0.22137	0.26522
		[a]	0.21536	0.27507
		Mean	0.21870	0.26960
		SD	0.02341	0.02772
7	ഉ [u]	[wu]	0.09610	0.10346
		[uw]	0.10917	0.08431
		[ɯ]	0.07510	0.07979
		[ə]	0.15100	0.08847
		[ə*]	0.13990	0.08253
		[ɯv]	0.06245	0.07106
		[U]	0.08139	0.07943
		Mean	0.09856	0.08409
		SD	0.03565	0.01455
8	ഊ /u:/	[wu:]	0.19784	0.24334
		[u]	0.19789	0.21948
		Mean	0.19786	0.23380
		SD	0.02819	0.04677
9	ഒ /o/	[wO]	0.10789	0.11639
		[O]	0.09402	0.13223
		Mean	0.10096	0.12431
		SD	0.01459	0.02296

Sl.No	Vowel	Allophone	Average Duration (sec)	
			Male	Female
10	ഓ /o:/	[wO:]	0.24083	0.26733
		[O]	0.20402	0.23338
		Mean	0.22351	0.25135
		SD	0.03539	0.04161

A comparison is performed on the durations computed for the isolated vowel phones and the average durations obtained for the vowel allophones extracted from the example word set present in the TEMU dataset. Figure IV (a-b) shows the duration of the isolated vowel phones together with the average duration for each vowel phoneme extracted from the allophonic variations obtained from the TEMU dataset for both male and female speakers.



A. MALE SPEAKER



B. FEMALE SPEAKER

FIGURE IV (A-B): DURATION OF ISOLATED VOWEL PHONES AND THE AVERAGE DURATION OF ITS ALLOPHONIC VARIATIONS

VI. CONCLUSION

Every phone in any spoken language is pronounced as one of its allophone. For the very reason, the properties of allophonic variations of each phone are very vital in continuous speech recognition and speech synthesis studies. In this work Malayalam vowel allophones are identified, classified and analyzed based on their durational properties. From the experimental results, it is evident that allophonic variations exists in the durational property of vowel phonemes. This aspect can be effectively used to improve the performance of the Malayalam ASR and synthesis. This work can be considered as a first step towards a paradigm shift to allophone based Malayalam speech processing.

ACKNOWLEDGMENT

We would like to thank The Centre for Malayalam Language Technology, Thunjath Ezhuthachan Malayalam University, Kerala, India for providing necessary support to work with the TEMU Malayalam phonetic archive project and for granting permission to use same dataset for experimental purposes.

REFERENCES

[1] "Report of the Commissioner for linguistic minorities: 50th report (July 2012 to June 2013)". Commissioner for Linguistic Minorities, Ministry of Minority Affairs, Government of India. Retrieved 17 September 2016.

[2] Asher, Ronald E., and T. C. Kumari. Malayalam. Psychology Press, 1997.

[3] Soman, Arun, et al. "Corpus Driven Malayalam Text-to-Speech Synthesis for Interactive Voice Response System." International journal of computer application 29.4 (2011).

[4] V.L. Lajish, T.M. Thasleema and N.K Narayanan, 2006. "Hidden Markov Models for isolated wordspeech recognition based on large vocabulary", Proc. of the National Symposium on Acoustics, (NSA2006), NPL, New Delhi.

[5] Kurian, Cini, and KannanBalakriahnan. "Continuous speech recognition system for Malayalam language using PLP cepstral coefficient." Journal of Computing and Business Research 3.1 (2012).

[6] V. R. Prabodhachandran Nair, Svanavijnjaanam (Phonetics), SIL, 1980

[7] Mattheyses, Wesley, and Werner Verhelst. "Audiovisual speech synthesis: An overview of the state-of-the-art." Speech Communication 66 (2015): 182-217.

[8] Cawley, G. C., and P. D. Noakes. "Allophone synthesis using a neural network." Proceedings of the First World Congress on Neural Networks (WCNN-93)(2). 1993.

[9] Turkmani, Aseel. Visual analysis of viseme dynamics. Diss. University of Surrey, 2008.

[10] Kozierski, Piotr, et al. "Allophones in automatic whispery speech recognition." Methods and Models in Automation and Robotics (MMAR), 2016 21st International Conference on. IEEE, 2016.

[11] Nguyen, Long, Xuefeng Guo, and John Makhoul. "Modeling Frequent Allophones in Japanese Speech Recognition." Seventh International Conference on Spoken Language Processing. 2002.

[12] Xu, Ji, et al. "Automatic allophone deriving for Korean speech recognition." Computational Intelligence and Security (CIS), 2013 9th International Conference on. IEEE, 2013.

[13] Imedjdouben, FayÇal, and Amrane Houacine. "Generation of allophones for speech synthesis dedicated to the Arabic language." New Technologies of Information and Communication (NTIC), 2015 First International Conference on. IEEE, 2015.

[14] Skrelin, Pavel A. "Allophone-based concatenative speech synthesis system for Russian." International Workshop on Text, Speech and Dialogue. Springer, Berlin, Heidelberg, 1999.

[15] Barkhoda, Wafa, et al. "A comparison between allophone, syllable, and diphone based TTS systems for Kurdish language." Signal Processing and Information Technology (ISSPIT), 2009 IEEE International Symposium on. IEEE, 2009.

[16] Hamad, Mazin, and Mustafa Hussain. "Arabic text-to-speech synthesizer." Research and Development (SCOReD), 2011 IEEE Student Conference on. IEEE, 2011.

[17] Karpov, Alexey, et al. "Audio-visual speech asynchrony modeling in a talking head." (2009).

[18] Wong, Ka-Ho, Wai-Kit Lo, and Helen Meng. "Allophonic variations in visual speech synthesis for corrective feedback in capt." Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on. IEEE, 2011.

[19] Kalberer, Gregor A., Pascal Müller, and Luc J. Van Gool. "Speech Animation Using Viseme Space." VMV. 2002.

[20] Pols, Louis CW, Xue Wang, and Louis FM ten Bosch. "Modelling of phone duration (using the TIMIT database) and its potential benefit for ASR." Speech Communication 19.2 (1996): 161-176.

[21] Klatt, Dennis H. "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence." The Journal of the Acoustical Society of America 59.5 (1976): 1208-1221.

[22] Lee, Yong-Ju, and Sook-hyang Lee. "On phonetic characteristics of pause in the Korean read speech." Fourth International Conference on Spoken Language Processing. 1996.

[23] Sayli, Omer. "Duration analysis and modeling for Turkish text-to-speech synthesis." Master's thesis, Department of Electrical and Electronics Engineering, Bogaziei University (2002).

[24] Batusek, Robert. "A duration model for Czech text-to-speech synthesis." Speech Prosody 2002, International Conference. 2002.

[25] Lazaridis, Alexandros, et al. "A CART approach for duration modeling of Greek phonemes." Proc. of SPECOM. 2007.

[26] Chen, Bo, Tianling Bian, and Kai Yu. "Discrete duration model for speech synthesis." Proc. Interspeech 2017 (2017): 789-793.

[27] Hifny, Yasser, and Mohsen Rashwan. "Duration modeling for arabic text to speech synthesis." Seventh International Conference on Spoken Language Processing. 2002.

[28] Norkevičius, Giedrius. Method for creating phone duration models using very large, multi-speaker, automatically annotated speech corpus. Diss. Vytautas Magnus University, 2011.

[29] Pylkkonen, Janne, and Mikko Kurimo. "Duration modeling techniques for continuous speech recognition." Eighth International Conference on Spoken Language Processing. 2004.

[30] Ogbureke, Kalu U., Joao P. Cabral, and Julie Carson-Berndsen. "Explicit duration modelling in HMM-based speech synthesis using a hybrid hidden Markov model-Multilayer Perceptron." SAPA-SCALE Conference. 2012.

[31] Sovilj-Nikic, S., et al. "Tree-based phone duration modelling of the Serbian language." Elektronika ir Elektrotechnika 20.3 (2014): 77-82.

[32] Samudravijaya, K.: Durational characteristics of Hindi phonemes in Continuous Speech. Technical Report, TIFR (April 2003)

[33] Samudravijaya, K. "Durational characteristics of Hindi stop consonants." Eighth European Conference on Speech Communication and Technology. 2003.

[34] Rao, K. Sreenivasa, and B. Yegnanarayana. "Modeling syllable duration in indian languages using support vector machines." Intelligent Sensing and Information Processing, 2005. Proceedings of 2005 International Conference on. IEEE, 2005.

[35] Krishna, N. Sridhar, and Hema A. Murthy. "Duration modeling of Indian languages Hindi and Telugu." Fifth ISCA Workshop on Speech Synthesis. 2004.

[36] Savithri, S. R. "Durational analysis of Kannada vowels." Journal of Acoustical Society of India 14.2 (1986): 34-31.

[37] Gopinath, Deepa P., et al. "Duration analysis for malayalam text-to-speech systems." Information Technology, 2006. ICIT'06. 9th International Conference on. IEEE, 2006.

[38] Rajan, Bindhu K., et al. "Duration modeling for text to speech synthesis system using festival speech engine developed for Malayalam language." Circuit, Power and Computing Technologies (ICCPCT), 2015 International Conference on. IEEE, 2015.

[39] http://www.cmltemu.in/.