IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

Implementation of PSOLA for Comparative Linguistic Analysis of English and Marathi Language

Kavita Waghmare[1], Pravin Yannawar[2], Bharti Gawali[3]

Research student, Department of CS & IT, Dr. B.A.M.U, Aurangabad, India[1]

Assistant Professor, Department of CS & IT, Dr. B.A.M.U, Aurangabad, India[2]

Professor, Department of CS & IT, Dr. B.A.M.U, Aurangabad, India[3]

Abstract: Naturalness is one of the very important parameter to evaluate the speech synthesis system. Besides other prosodic feature parameters like pitch, formant and energy, pitch plays a very significant role in understanding spoken signals. To obtain naturalness in synthesized speech, pitch has to be analysed. Pitch is dependent on language we attempted to present the implementation of PSOLA algorithm on the database of English and Marathi sentences. It is seen that English being irregular language results in different accent whereas Marathi being regular proves to be with good synthesized speech.

Keywords: Speech Synthesis, PSOLA, pitch.

1. INTRODUCTION

Language is the ability to acquire and use complex systems of communication and is any specific example of such a system. The scientific study of language is called linguistics and is the ingredient for speech [1] [2]. Speech processing is the study of speech signals and the processing methods of these signals. It includes the study of speech recognition, speaker recognition, speech coding, speech analysis, speech enhancement, speech compression and speech synthesis. Speech synthesis is the process of converting written text into voice communication [3].

It is also referred as text-to-speech system (TTS). In the process of speech synthesis, mainly two processing components are used; they are NLP (Natural Language Processing) and DSP (Digital Signal Processing) modules [4]. This system first converts the input text into its corresponding linguistic or phonetic representations and then produce the sounds corresponding to those representations. The main aim of this system is to generate natural sounding voice for conveying information to the user in the desired accent, language, and voice [5].

Natural speech can be produced by concatenation technique and it is the simplest and widely practiced techniques for speech synthesis. In this technique, units of phonemes, diphones, syllable, words and phrases are concatenated in order to make the speech output. These words are pre-recorded in the machine memory and retrieved according to need. This proficiency has its own problems like discontinuities at concatenation points, high memory requirements, data collection and labelling takes more time. There are three basic forms of concatenation, they are:

Unit Selection Synthesis
This technique uses large databases of recorded speech. The primary advantage of using large database is to vary prosodic and spectral characteristics and it should be possible to synthesize more natural- sounding speech than that can be produced by a diminished set of controlled units [6]. The main difference between the unit selection and the diphone based method is the length of the used in speech segments.

Diphone based Synthesis
It is one of the method used for creating synthetic voice. It uses minimal speech database. It uses two adjacent phones to make the speech waveform [7]. The quantity of diphones in the database depends on the phonotactics of the desired language. Diphone gives best results in the languages having consistencies in the pronunciation. It suffers with problems of discontinuities.

Domain based synthesis
Domain based synthesis concatenates pre-recorded words to produce entire utterances relevant to specific field. It can be used specifically announcement system in railways, bus, airports, weather condition reports due to restricted vocabulary.

The most complex attribute of speech is the determination of correct/ appropriate prosody information for the sentence. Prosodic features consist of pitch, duration and intensity of synthesized speech [8]. Naturalness and smoothness plays very important role while synthesizing speech samples [9]. It is due to pitch variation. Pitch is one of the most important parameters. So it is very important

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

to extract the pitch specific information [10]. One of the approaches in order to change the voice is to change the pitch of the voice, this is done by shifting the pitch of the voice using certain techniques like the Pitch Synchronous Overlap Add (PSOLA) algorithm [11]. Pitch period is the period of glottal pulse [12]. It is responsible for making some sounds to be sharper than others. The main aim of pitch shifting algorithm is to create a change in pitch without creating a change in the replay rate.

PSOLA works by dividing the speech waveform in small overlapping segments. To change the pitch of the signal, the segments are moved further apart or closer together. To change the duration of the signal, the segments are then repeated multiple times or eliminated and then the segments are combined using add technique. It directly works on the signal waveform without any sort of model and therefore does not lose any detail of the signal. Several types of PSOLA algorithm are present such as Time Domain, Frequency domain and Linear Predictive PSOLA. The TD-PSOLA is most commonly used due to its computational efficiency [13]. The TD-PSOLA algorithm was proposed allowing pitch modification of a given signal without changing the time duration and vice versa. The TD-PSOLA algorithm was suggested to allow pitch modification of a given speech signal without altering the time duration and vice versa. The TD-PSOLA consists mainly of the following three steps:

1. The analysis step, where the original speech signal is first divided into separate but often overlapping short term analysis signals (ST). Short term signals xm (n) are obtained from the digital speech waveform x (n) by multiplying the signal by a sequence of the pitch synchronous analysis window HM (n) as in Eq. 1:

$$Xm(n) = hm(tm - n)x(n) \text{ ---------- } (1)$$

where m is an index for the short-time signal

2. The windows, which are usually Hanning type, are centered on the successive instants tm, called pitch marks. These marks are set at a pitch-synchronous rate in the voiced parts of the signal and at a constant rate on the unvoiced parts.
3. The modification step, where each frame is modified according to the target. The synthesis steps are done such that these segments are recombined by means of overlap adding.

Creation of Database

The database is created for the analysis of pitch synchronisation in English and Marathi Language. In English language a male and female speaker has uttered sentences while for Marathi male speaker has uttered all sentences and the database has been downloaded online [14]. These sentences are collected from story books and newspaper articles. The total number of sentences in English language is around 590 and in Marathi it is around 1000. The database collection has been exercised in a recording studio in the afternoon session with a complete noise-free environment. For the Pitch analysis out of these many sentences 30 sentences in each language have been chosen. Each sentence is uttered for one time by the speaker. The total size of database considered for analysis is 120 sentences. After collection of these speech samples, the sentences have undergone though synthesis process. For synthesizing the speech samples Festival Framework has been used, which is a platform for research and development with its highly flexible architecture. The experiment has been done on 30 sentences in each database but for presentation only 10 sentences has been shown. The technical specifications are shown in table 1. While the table 2 depicts The specification of database voices. The table 3,4,5 shows the English and Marathi sentences with their respective labels used for original and synthesized speech samples.

Table 1 the technical specification for database creation

Sr.No	Parameter	Specification
1.	Sampling Frequency	16Khz
2.	Distance from microphone	10cm
3.	Environment	Recording studio
4.	Temperature	25
5.	Channel	Single
6.	Gender	Female:1

Table 2 Specification of Database

Sr.No	Language	Voice	Label	Utterance
1.	English	Male-1 Female-1	R001-R030 B001-B0030	30*1=30 30*1=30 Total=60
2.	Marathi	Male subject 1 Male subject 2	M001-M030 M101-M130	30*1=30 30*1=30 Total=60

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

Table 3 The sentences used in Pitch analysis of Marathi language for subject 1

Sr. No	Sentence	Label used for the original speech file	Label used for the synthesized speech file
1	छत्रपती शिवाजी महाराज यांच्या पत्नी	M001	MS001
2	विकिपीडिया हा एक जनकोश आहे	M002	MS002
3	तुम्ही काही मार्गदर्शन करू शकाल का धन्यवाद	M003	MS003
4	भारत सरकार द्वारे दिला जाणारा एक नागरी सम्मान	M004	MS004
5	मिशिगन राज्यातील सर्वात मोठे औद्योगिक शहर आहे	M005	MS005
6	आता हे शहर पुणे ह्या अधिकृत नावाने ओळखल्या जाते	M006	MS006
7	युनैटेड किंग्डम या साम्राज्याचा एक भाग	M007	MS007
8	चुक निदर्शनास आणून दिल्या बद्दल धन्यवाद	M008	MS008
9	ऑस्ट्रेलिया देशांतर करण्या आधी आय	M009	MS009
10	राजस्थान राज्यातील पाली जिल्हया विषयची लेख	M010	MS010

Table 4 Sentences used in Pitch Analysis of Marathi language of subject 2

Sr. No	Sentence	Label used for the original speech file	Label used for the synthesized speech file
1	मला वाटलं, की पाणी माझ्या हाता खाली श्वास, घेत होते	M001	MS001
2	मी माझे डोळे, घट्ट बंद केले	M002	MS002
3	मी मनात प्रार्थना करत होतो, की माझे आई वडील, त्यात असावे, अचानक, मला एक जोराचा, झटका लागला	M003	MS003
4	साइबेरियन चित्ता, आपल्या समुद्रासारख्या निळ्या डोळ्यांनी, मला, रागीटपणे बघत होता	M004	MS004
5	त्याचे काळे पट्टे, मंत्रमुग्ध करण्यासारखे होते	M005	MS005
6	दुर्भाग्याने, वाइटाचे अनेक रुप असतात, आणि हे बोलत बोलत, चुप होऊन, तो अंतरिक्ष यानाकडे, बघु लागला	M006	MS006
7	मला काचेच्या, छोट्या छोट्या खिडक्यांजवळचे चेहरे, अस्पष्ट दिसू लागले होते	M007	MS007
8	मी श्वास रोखून, प्रतिक्षा करु लागलो	M008	MS008
9	एक मिनिट झाल्यानंतर, दुसरा असं वाटत होते, जसे आत ते लोक, फार वेळ लावत आहेत	M009	MS009
10	मी शिवा आहे. एक मानव, ह्या साहसपुर्ण केलेल्या घोषणेनंतर, जवाबात, पुर्ण शांतता होती	M010	MS010

Table 5 Sentences used in Pitch Analysis of English language in Male voice

Sr. No	Sentence	Label used for the original speech file	Label used for the synthesized speech file
1	Will we ever forget it	R001	RS001
2	If I ever needed a fighter in my life I need one now	R002	RS002

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

3	He was a head shorter than his companion, of almost delicate physique	R003	RS003
4	There was a change	R004	RS004
5	I followed the line of the proposed railroad, looking for chances	R005	RS005
6	I was about to do this when cooler judgment prevailed	R006	RS006
7	It occurred to me that there would have to be an accounting	R007	RS007
8	To my surprise, he began to show actual enthusiasm in my favour	R008	RS008
9	I had faith in them	R009	RS009
10	She turned in at the hotel	R010	RS010

Table 6 Sentences used in Pitch Analysis of English language in Female voice

Sr. No	Sentence	Labels for original speech signal	Label synthesized speech signal
1	I can see that knife now	B001	BS001
2	They robbed me a few years later	B002	BS002
3	I was completely lost in my work	B003	BS003
4	He caught himself with a jerk	B004	BS004
5	It won't be for sale	B005	BS005
6	She was even more beautiful than when I saw her, before	B006	BS006
7	There was no answer from the other side	B007	BS007
8	Tomorrow it will be strong enough for you to stand upon	B008	BS008
9	You were going to leave after you saw me on the rock	B009	BS009
10	I want to die in it	B010	BS010

Effect Analysis of Pitch Synchronous Overlap Add (PSOLA) on Marathi and English Language

Speech is the output of a time-varying vocal tract system excited by a time- varying excitation. Pitch is one of the most important parameters for speech signal processing including speech synthesis, automatic speech recognition, speech enhancement etc.It is the most important parameter in speech synthesis. Pitch can be estimated using different methods such as Autocorrelation method, Cepstrum Pitch determination, Pitch Estimation by SIFT method and PSOLA. The pitch values contain the speaker specific information. The pitch variation carries the intonation signal associated with rhythms of words, speaking style, emotions and stress.

The gender is one of the elements which convey a voice in the enactment of the vocal tract. Randomly, the average pitch for females is about 200 Hz and for males it is about 110Hz.Pitch variation is often correlated with loudness and lowness in speech samples. Pitch modification means transposing the pitch without changing its characteristics. The discontinuity of concatenative speech synthesis samples can be improved by correcting the pitch discontinuities of speech samples. For this pitch marking is an essential step. The experimentation is done by counting on the prosodic features as pitch, duration, frequency, fundamental frequency values for a male and female speaker in English and two male speakers Marathi language. These values are calculated using Computerized Speech Lab (CSL). CSL gives values for the minimum pitch, maximum pitch, Mean frequency, Mean fundamental frequency (F0), Mean Period of the speech sample, standard deviation, median pitch, root mean square, geometric mean. In the table 4 and 5 it can be seen that there is difference in the values of male and the female speaker, it is due to the pitch difference in male and female voices.

It is clearly visible in table number 9,10,11 and 12 that in English language the values of minimum pitch and the maximum can be seen and altered in order to get good quality speech. Due to the alteration in the pitch values the duration is also altered in some of the speech samples. It is due to irregularity in English language as letter does not corresponds to sound so there is variation in the pitch values.

But on the other hand, in table 13,14,15 and 16 it is found that in Marathi being regular, there is no significant different in the original and the synthesized speech, it is due to resemblance in the written letters and their corresponding sounds as it is a regular language. The graphical representation of some sentences of English and Marathi sentences have been shown in figure 1,2,3 and 4. It consists of simple speech signal, signal after pitch marking and the pitch contour.

DOI 10.17148/IJARCCE.2016.5714

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

2. GRAPHICAL REPRESENTATION

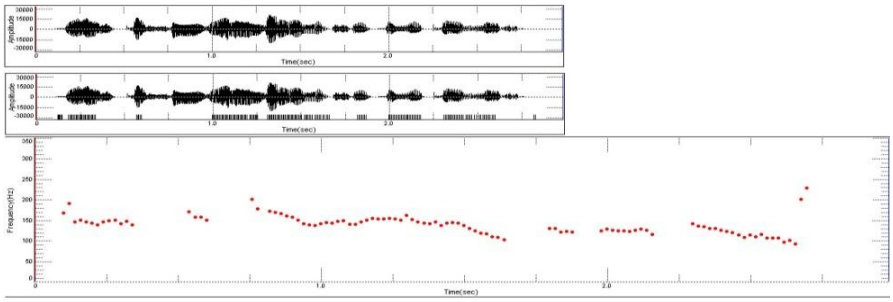

Figure 1 Graphical representation of sentence "Does that look well" for the Original speech sample

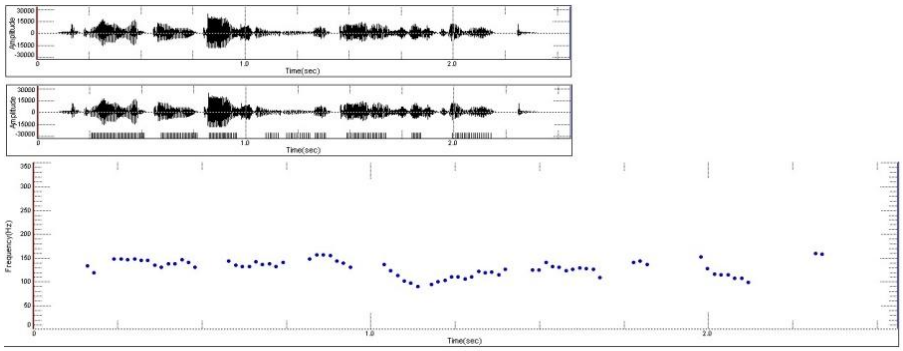

Figure 2 Graphical representation of sentence "Does that look good" for the Synthesized speech sample

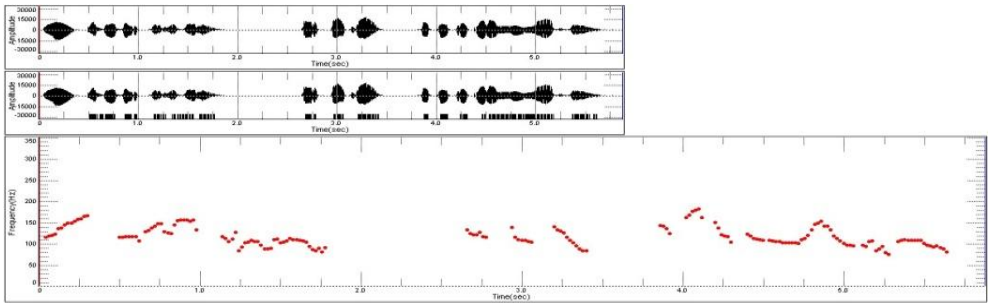

Figure 3 Graphical representation of sentence "**मी अगोदर पुढे गेलो, तर त्याचा, किती प्रिाराांनी लाभ होईल**" for original speech sample

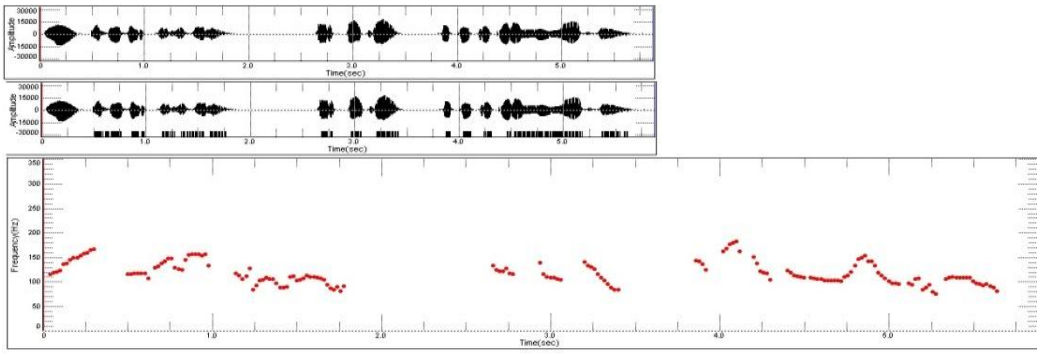

Figure 4 Graphical representation of sentence "**मी अगोदर पुढे गेलो, तर त्याचा, किती प्रिाराांनी लाभ होईल**" for synthesized speech sample

DOI 10.17148/IJARCCE.2016.5714

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

Table 9 shows the values of various statistical measures used in the pitch analysis of Female subject in Original Speech Sample

Label	End of Analysis	Min pitch	Max pitch	Mean Frequency	Mean F0 (Hz)	Mean Period (msec)	Standard Deviation (Hz)	Median Pitch (Hz)	Root Mean square (Hz)	Geometric Mean (Hz)
B001	1.77	135.54	306.62	221.35	211.44	4.73	46.19	216.65	226.03	216.46
B002	2.45	129.65	293.82	200.8	213.31	4.69	39.21	219.01	224.22	217.17
B003	2.31	146	326.94	250.85	243.77	4.1	39.41	254.49	253.89	247.49
B004	2.19	131.03	310.48	225.56	216.68	4.62	43.04	224.11	229.56	221.26
B005	1.80	149.09	295.45	236.44	230.89	4.33	34.63	239.69	238.91	233.77
B006	4.65	115.86	323.09	221.98	215.37	4.64	35.85	224	224.83	218.86
B007	2.75	125.54	285.08	216.56	208.03	4.81	39.4	225	220.08	212.56
B008	3.56	130.13	305.94	207.35	201.68	4.96	33.61	205.43	210.03	204.57
B009	4.10	121.29	294.09	225.42	220.99	4.53	29.97	226.35	227.39	223.31
B010	1.68	116.6	310.93	209.8	204.14	4.9	33.09	212.86	212.34	207.09

Table 10 shows the values of various statistical measures used for the pitch analysis of Female subject in Synthesized Speech Sample

Label	End of Analysis	Min pitch	Max pitch	Mean Frequency	Mean F0(Hz)	Mean Period (msec)	Standard Deviation (Hz)	Median Pitch (Hz)	Root Mean Square (Hz)	Geometric Mean (Hz)
BS001	3.78	126.95	255.9	210.07	206.54	4.84	24.29	214.71	211.46	208.45
BS002	2.45	137.5	302.55	224.74	217.75	4.59	38.71	223.21	228.01	221.31
BS003	2.29	143.6	328.59	228.17	217.69	4.59	46.26	243.23	232.75	223.12
BS004	1.53	126.92	299.96	231.36	223..56	4.47	39.5	230.92	234.63	227.69
BS005	1.53	123.92	299.96	231.36	223.56	4.47	39.5	230.92	234.63	227.69
BS006	4.21	85.81	318.45	224.14	217.22	4.6	34.87	220.71	226.82	221.07
BS007	2.75	125.54	285.08	216.56	208.03	4.81	39.4	225	220.08	212.56
BS008	3.76	130.6	312	205.64	200.16	5	32.68	205.74	208.2	202.97
BS009	4.08	122.51	287.04	218.82	213.64	4.68	30.91	222.47	220.98	216
BS010	1.57	161.83	246.95	211.84	210.15	4.76	18.53	214.43	212.63	211.01

Table 11 shows the values of various statistical measures used for the pitch analysis of Male subject in Original Speech Sample

Label	End of Analysis	Min pitch	Max pitch	Mean Frequency	Mean F0(Hz)	Mean Period (msec)	Standard Deviation (Hz)	Median Pitch (Hz)	Root Mean Square (Hz)	Geometric Mean(Hz)
R001	1.83	100.48	164.77	132.92	130.97	7.54	16.4	130.32	133.9	131.93
R002	3.47	93.49	200.74	128.62	126.39	7.87	17.65	126.26	129.81	127.48
R003	4.06	93.9	178.63	129.4	127.12	7.84	17.21	126.81	130.53	128.26
R004	1.62	100.31	148.6	124.64	122.8	8.15	15.07	126.09	125.53	123.73
R005	4.21	97.09	196.29	125.89	123.87	8.09	16.44	124.07	126.95	124.86
R006	3.35	91.68	266.16	130.43	126.46	7.92	27.38	123.28	133.25	128.23
R007	2.81	102.1	211.86	131.29	129.24	7.68	17.59	128.49	132.22	130.22
R008	3.96	94.08	175.41	128.03	126.45	7.84	14.48	125.89	128.84	127.23
R009	1.54	89.19	150.97	122.08	119.12	8.39	18.59	125.18	123.46	120.63
R010	2.04	98.32	263.96	142.32	136.4	7.33	34.13	138.14	146.29	139.07

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

Table 12 shows the values of various statistical measures used for the pitch analysis for Male subject for Synthesized Speech Sample

Label	End of Analysis	Min pitch	Max pitch	Mean Frequency	Mean F0(Hz)	Mean Period (msec)	Standard Deviation (Hz)	Median Pitch (Hz)	Root Mean Square (Hz)	Geometric Mean(Hz)
RS001	1.81	92.93	203.8	135.9	132.58	7.54	22.27	130.78	137.68	134.2
RS002	3.45	99.59	274.17	130.49	127.11	7.87	24.89	125	132.83	128..64
RS003	4.05	94.73	178.03	124.67	127.61	7.84	16.52	127.4	130.76	128.67
RS004	1.59	98.93	157.63	124.67	122.75	8.15	15.5	123.99	125.61	123.71
RS005	4.20	92.52	161.13	125.48	123.66	8.09	15.26	123.6	126.4	124.57
RS006	3.33	91.77	269.53	130.68	126.27	7.92	29.26	122.83	133.89	128.22
RS007	2.80	102.51	245.37	132.8	130.21	7.68	20.87	130.45	134.41	131.42
RS008	4.07	92.99	164.62	129.34	127.55	7.84	15.32	127.64	130.24	128.44
RS009	1.52	80.77	151.49	122.05	118.5	8.44	19.67	127.38	123.59	120.35
RS010	1.86	84.79	185.86	130.73	126.83	7.88	22.12	133.35	132.56	128.82

Table 13 shows the values of various statistical measures used for the pitch analysis of Male subject 1 for Original Speech Sample in Marathi language

Label	End of Analysis	Min pitch	Max pitch	Mean Frequency	Mean F0 (Hz)	Mean Period (msec)	Standard Deviation (Hz)	Median Pitch (Hz)	Root Mean Square(Hz)	Geometric Mean(Hz)
M001	5.12	82.67	202.41	142.42	136.56	7.32	28.36	137.74	145.19	139.53
M002	5.04	83.74	291.3	139.64	133.28	7.5	33.76	132.11	143.62	136.25
M003	5.86	76.96	209.26	146.45	141.83	7.05	26.61	139.74	148.83	144.12
M004	6.79	85.84	223.64	139.02	134.15	7.45	26.06	138.77	141.42	136.59
M005	5.88	73.08	225.88	135.41	129.28	7.73	29.78	129.44	138.62	132.3
M006	7.7	75.16	221.76	151.52	144.35	6.93	32.45	144.24	154.93	147.97
M007	6.20	84.4	224.83	141.2	134.37	7.44	30.34	144.31	144.39	137.84
M008	5.83	85.73	228.32	108.48	105.89	9.46	20.81	104.06	110.2	106.78
M009	5.07	89.34	275.24	147.89	139.97	7.14	35.86	137.71	152.14	143.83
M010	5.54	97.11	265.81	160.14	150.64	6.64	40.71	142.62	165.2	155.23

Table 14 show the values of various statistical measures used for the pitch analysis of Male subject in Synthesized Sample in Marathi language

Label	End of Analysis	Min pitch	Max pitch	Mean Frequ-ency	Mean F0(Hz)	Mean Period (msec)	Standard Deviation (Hz)	Median Pitch (Hz)	Root Mean Square (Hz)	Geometric Mean (Hz)
MS001	5.12	82.67	202.41	142.42	136.56	7.32	28.36	137.74	145.19	139.53
MS002	5.04	83.74	291.3	139.64	133.28	7.5	33.76	132.11	143.62	136.25
MS003	5.86	76.96	209.26	146.45	141.83	7.05	26.61	139.74	148.83	144.12
MS004	6.79	85.84	223.64	139.02	134.15	7.45	26.06	138.77	141.42	136.59
MS005	5.88	73.08	225.88	135.41	129.28	7.73	29.78	129.44	138.62	132.3
MS006	7.7	75.16	221.76	151.52	144.35	6.93	32.45	144.24	154.93	147.97
MS007	6.20	84.4	224.83	141.2	134.37	7.44	30.34	144.31	144.39	137.84
MS008	5.83	85.73	228.32	108.48	105.89	9.44	20.81	104.19	110.44	107.02
MS009	5.07	89.34	275.24	147.89	139.97	7.14	35.86	137.71	152.14	143.83
MS010	5.54	97.11	265.81	160.14	150.64	6.64	40.71	142.62	165.2	155.23

DOI 10.17148/IJARCCE.2016.5714

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

Table 15 show the values of various statistical measures used for the pitch analysis of Male subject in Original Speech Sample in Marathi language

Label	End of Analysis	Min pitch	Max pitch	Mean Frequency	Mean F0 (Hz)	Mean Period (msec)	Standard Deviation (Hz)	Median Pitch (Hz)	Root Mean Square (Hz)	Geometric Mean(Hz)
M101	5.74	94.06	164.05	115.32	113.67	8.80	14.22	112.76	116.18	114.48
M102	3.35	78.44	131.45	110.98	109.36	9.14	13.37	107.26	111.78	110.18
M103	10.73	90.42	150.58	111.53	109.95	9.10	13.93	108.42	112.39	110.72
M104	8.35	89.14	136.66	109.85	109.11	9.17	9.18	109.18	110.23	109.48
M105	4.26	92.13	133.78	108.39	107.51	9.30	9.91	108.42	108.84	107.94
M!06	11.31	85.86	208.94	107.01	105.91	9.44	12.38	105.74	107.72	106.42
M!07	6.68	85.85	150.01	113.02	110.28	9.07	18.40	107.82	114.50	111.61
M108	3.80	91.59	182.42	110.12	108.27	9.24	15.59	103.70	111.20	109.15
M109	8.34	87.36	269.63	111.56	109.00	9.17	21.29	107.85	113.57	110.12
M110	10.77	87.06	195.13	112.16	110.20	9.07	16.01	107.81	113.29	111.13

Table 16 show the values of various statistical measures used for the pitch analysis of Male subject in Synthesized Speech Sample in Marathi language

Label	End of Analysis	Min pitch	Max pitch	Mean Frequency	Mean F0 (Hz)	Mean Period (msec)	Standard Deviation (Hz)	Median Pitch (Hz)	Root Mean Square (Hz)	Geometric Mean (Hz)
MS101	5.74	86.41	183.65	114.86	112.50	8.89	17.24	111.80	116.14	113.65
MS102	3.35	73.29	137.05	109.56	107.58	9.30	14.55	108.50	110.51	108.50
MS103	10.73	71.45	155.97	111.41	109.35	9.14	15.61	109.59	112.49	110.36
MS104	8.35	82.47	146.04	109.58	108.50	9.22	11.02	108.52	110.13	109.03
MS105	4.26	75.57	144.38	108.83	107.26	9.32	13.14	107.20	109.20	109.61
MS106	11.31	79.79	225.35	107.08	105.49	9.48	14.34	105.75	108.03	106.25
MS107	6.68	76.71	157.40	113.02	109.80	9.11	19.73	107.82	114.72	111.38
MS108	3.80	75.70	190.07	110.46	107.94	94.26	18.25	104.27	111.95	109.14
MS109	8.34	82.28	271.08	112.09	108.86	9.19	23.06	108.19	114.43	110.31
MS110	10.77	79.74	317.45	113.91	110.59	9.04	24.44	108.44	116.49	112.05

3. EXPERIMENTAL ANALYSIS

To evaluate the comparative difference between pitch of English and Marathi Languages Euclidean Distance is used. This difference is used as a measure of correlation between user preferences. The smaller the difference, the closer the correlation between user preferences.

The greater the distance, the less correlation exists between user preferences. We have used the Euclidean distance to find the similarity percentage of the original speech sample with respect to synthesized speech. The observations found are discussed below:

For English Female database, the similarity percentage ranges between 52.59 to 100 %.
For English Male database, the similarity percentage ranges 64.81 to 100 %.
For Marathi database subject 1, the similarity percentage is 100 %.
For Marathi database subject 2, the similarity percentage lies between 65.51 to 100%.

Table 17 Euclidean Distance and Similarity Percentage for Female Speaker in English Language

Label	Standard Deviation (Hz)	Label	Standard Deviation (Hz)	Euclidean distance	Similarity Percentage
B001	46.19	BS001	24.29	21.9	52.59
B002	39.21	BS002	38.71	0.5	98.72

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

B003	39.41	BS003	46.26	6.85	85.19
B004	43.04	BS004	39.5	3.54	91.78
B005	34.63	BS005	39.5	4.87	87.19
B006	35.85	BS006	34.87	0.98	97.27
B007	39.4	BS007	39.4	0	100.00
B008	33.61	BS008	32.68	0.93	97.23
B009	29.97	BS009	30.91	0.94	96.96
B010	33.09	BS010	18.53	14.56	56.00

Table 18 Euclidean Distance and Similarity Percentage for Male Speaker in English Language

Label	Standard Deviation (Hz)	Label	Standard Deviation (Hz)	Euclidean Distance	Similarity Percentage
R001	16.4	RS001	22.27	5.87	73.64
R002	17.65	RS002	24.89	7.24	70.91
R003	17.21	RS003	16.52	0.69	95.99
R004	15.07	RS004	15.5	0.43	97.23
R005	16.44	RS005	15.26	1.18	92.82
R006	27.38	RS006	29.26	1.88	93.57
R007	17.59	RS007	20.87	3.28	84.28
R008	14.48	RS008	15.32	0.84	94.52
R009	18.59	RS009	19.67	1.08	94.52
R010	34.13	RS010	22.12	12.01	64.81

Table 19 Euclidean Distance and Similarity Percentage for Male for Subject 1 Speaker in Marathi Language

Label	Standard Deviation (Hz)	Label	Standard Deviation(Hz)	Euclidean Distance	Similarity Percentage
M001	28.36	MS001	28.36	0	100
M002	33.76	MS002	33.76	0	100
M003	26.61	MS003	26.61	0	100
M004	26.06	MS004	26.06	0	100
M005	29.78	MS005	29.78	0	100
M006	32.45	MS006	32.45	0	100
M007	30.34	MS007	30.34	0	100
M008	20.81	MS008	20.81	0	100
M009	35.86	MS009	35.86	0	100
M010	40.71	MS010	40.71	0	100

Table 20 Euclidean Distance and Similarity Percentage for Male for Subject 2 Speaker in Marathi Language

Label	Standard Deviation (Hz)	Label	Standard Deviation (Hz)	Euclidean Distance	Similarity Percentage
M101	14.22	MS101	17.24	3.02	82.48
M102	13.37	MS102	14.55	1.18	91.89
M103	13.93	MS103	15.61	1.68	89.24
M104	9.18	MS104	11.02	1.84	83.30
M105	9.91	MS105	13.14	3.23	75.42
M!06	12.38	MS106	14.34	1.96	86.33
M!07	18.40	MS107	19.73	1.33	93.26
M108	15.59	MS108	18.25	2.66	85.42
M109	21.29	MS109	23.06	1.77	92.32
M110	16.01	MS110	24.44	8.43	65.51

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

4. RESULTS AND CONCLUSION

To evaluate the Performance for naturalness Euclidean Distance has been calculated. It is observed that:

- For English Female database, the similarity percentage ranges between 52.59 to 100 %. The 10% sentences are found to be of 100% similarity. This suggests that the original speech is equal to synthesized speech. Out of 10 sentences 50 % sentences are found in the range of 90-99. The similarity percentage of some sentences is 20% which are in the range of 80-89. Few sentences, similarity ranges in between 51-59, which indicates that these sentences are not similar to original sentences. It is due to the nasal sounds present in those sentences. As English is an irregular language.
- For English Male database, the similarity percentage ranges 64.81 to 99 %. The 6 sentences are found to be in range of 90-99 percent similarity. Out of 10 sentences 1 sentences are found to be in the range 80-89 and 2 sentences are in the range of 70-79. And one in the range of 60-69 percentages.
- For Marathi database subject 1, the similarity percentage is 100 % for 10 sentences. This is due to regular language pronunciation.
- For Marathi database subject 2, the similarity percentage lies between 65.51 to 93.26%. The similarity percentage of 3 sentences are found to be in the range of 90-99%. The 5 number of sentences are in the range of 80-89 % similar and 1 in the range of 70-79. The 1 number of sentences are in the range 60-69.

From these observations it can be concluded that language used is also very important in the application of synthesized speech generation. The regular languages bring more naturalness in the synthesized speech than irregular languages.

REFERENCES

[1] https://en.wikipedia.org/wiki/Language
[2] Santosh K.Gaikwad, Bharti W.Gawali, Pravin Yannawar, "A Review on Speech Recognition Technique", International Journal of Computer Applications (0975 – 8887) Volume 10– No.3, November 2010
[3] Prof. Preeti S.Rao, "Review of methods of Speech Synthesis", M-Tech Credit Seminar Report, Electronic Systems Group, EE Dept, IIT Bombay, Nov 2011.
[4] Mohammed Waseem, C.N.Sujatha, "Speech Synthesis System for Indian Accent using Festvox", International Journal of Scientific Engineering and Technology Research, Vol.03,Issues.34,November-2014.
[5] Archana Balyan,S.S Agrawal,Amita Dev, "Speech Synthesis: A Review", International Journal of Engineering Research & Technology,Vol.2,June 2013.
[6] Shurti Gupta,Parteek Kumar, "Comparative study of text to speech system for Indian Language", International Journal of Advances in Computing and Information Technology, April 2012.
[7] A.Indumathi, Dr.E.Chandra, "Survey On Speech Synthesis", Signal Processing: An International Journal(SPIJ),Vol (6) 2012.
[8] Sangram Kayte,Kavita Waghmare,Dr.Bharti Gawali, "Marathi Speech Synthesis: A review", International Journal on Recent and Innovation Trends in Computing and Communications, Vol 3,June 2015.
[9] Mahwash Ahmed,Shibli Nisar,"Text-to-Speech Synthesis using Phoneme Concatenation", International Journal Of Scientific Engineering and Technology, Vol No.3,Issue No.2,1 Feb 2014.
[10] Dr.K.V.N Sunitha , P.Sunitha Devi, "BHAASHIKA: Telugu TTS system", International Journal of Engineering Science and Technology, Vol.2(11),2010.
[11] Amin Shadravan Lalezari, Jalil Shirazi," Pitch Detection of Speech Signal using Wavelet Transform", International Journal of Scientific Engineering and Technology, Vol No.4,01 May 2015.
[12] Ashwini Songar,Mrs B.Harita,"MATLAB based Voice Conversion Model using PSOLA Algorithm", International Journal of Digital Application & Contemporary research, Vol 1,Issue 8,March 2013.
[13] Vivek Vijay Nar, Alice N.Cheeran, Souvik Banerjee, "Verification of TD-PSOLA for Implementing Voice Modification", International Journal of Engineering Research and Application,Vol.3,Issue 3,May-June 2013.
[14] http://festvox.org/databases/iiit_voices/
[15] Anant Bhatt, "A PSOLA based Approach for Voice Morphing", International Journal of Digital Applications & Contemporary Research, Vol 3, Issue 7, February 2015
[16] Ulrich Germann," An Iterative Approach to Pitch-marking of speech signals without Electroglottographic Data,CiteSeer 5M,2006
[17] JodoP.Cabra,LuisC.Oliveria,"Pitch-Synchronous Time-Scaling for Prodosic and VoiceQuality bhaTransformations",INTESPEECH 2005.