Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 4– No.4, February 2016 – www.caeaccess.org

Analysis of Pitch and Duration in Speech Synthesis using PSOLA

Kavita Waghmare
Research Student
Department of Computer Science and Information Technology
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad

Sangramsing Kayte
Research Student
Department of Computer Science and Information Technology
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad

Bharti Gawali
Professor
Department of Computer Science and Information Technology
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad

ABSTRACT
The speech synthesis system is an artificial production of speech with the help of speech synthesizers. It can be achieved using various techniques. During synthesis the smoothing of concatenating points is an important aspect to be studied. This paper attempts to find the effect of pitch-marking process using Time Domain-Pitch Synchronous Overlap and Add (TD-PSOLA) method. The database consists of 60 sentences containing various phones, syllables, phrases which provide prosodic effects in male and female voices. The analysis shows that the pitch –marking process affects the quality of speech in the synthesis process which soothes at concatenation point.

Keywords
Text-to speech (TTS), pitch, duration, PSOLA, pitch-markings.

1. INTRODUCTION
Voice communication has been the primary approach of human communication since it began to evolve at least one hundred thousand years ago. Spoken language is a complex and unique features of the human species. Speech synthesis is the process of automatic generation of speech waveforms, has been under development for several decades [1]. The speech synthesis is also referred as text-to-speech (TTS) synthesis consists of two primary stages. The foremost one is text analysis, where the input text is transliterated into a phonetic or linguistic representation, and the second one is the generation of speech waveforms, where the acoustic output is created from this phonetic and prosodic information[2][3]. These two forms are commonly called as high and low level synthesis. Figure 1 shows a simplified routine of a TTS system. The input text might be a data from a word processor, standard ASCII form, a mobile text-message, or scanned text from any text document. Then the character string is preprocessed, analyzed and transformed into a phonetic representation which is usually a string of phonemes with some additional information for correct intonation, duration, and stress. Speech sound is ultimately generated by the low-level synthesizer by the information from high-level one.

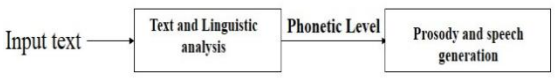

Figure 1.Text-to-Speech (TTS) Synthesis system

2. TECHNIQUES OF SPEECH SYNTHESIS
A text-to-Speech (TTS) synthesizer is a computer-based system that should be able to read any text loudly with an input written text. The goal of speech synthesis system is to develop a machine to produce natural sounding voice for conveying the information to the user in a desired accent and language. Synthesized speech can be brought out by various different methods such as articulatory synthesis, formant synthesis, concatenative synthesis and hidden Markov model based synthesis (HMM) [4]. Articulatory synthesis, attempts to model the human speech production system directly. Formant synthesis, which models the pole frequencies of speech signal or transfer function of the vocal tract based on source-filter-model. Concatenative synthesis, uses different length pre-recorded samples derived from natural language.

2.1 Concatenative Synthesis
Concatenative synthesis is based on the joining of pre-recorded words from the database. It takes small units of speech such as phones, diphones, triphone, syllables, demi syllables, phonemes and words. It is the most elementary method for synthesizing and requires a great number of words in the database. As the number of words that are used in different literatures is extremely large, storing all of those will be quite impossible. To avoid the storage of a large number of words, use of random sentences can serve to surmount this trouble. Equally most of the languages contain 30-40 phonemes at most, and then it would be easy to pull the necessary phonemes according to usage. Only it may lift the issue of discontinuity in the synthesized speech. The choice of units is generally a trade-off between longer and shorter units. Concatenation of words is relatively easy to perform and coarticulation effects within a word are captured in the stored units. This proficiency is normally confined to one speaker. The concatenation synthesis system is dependent on the choice of appropriate units and that joins those units in concert and performs close to signal processing to smooth unit transitions and to match predefined prosodic schemes [5].

There are three types of concatenative synthesis:

i. Unit Selection synthesis

ii. Di-phone synthesis

iii. Domain Based synthesis

i. Unit Selection synthesis
The unit selection synthesis is a method in which large numbers of pre-recorded words are used in order to get

synthesized speech. The main purpose of applying this method is to employ a great number of units with varied prosodic and spectral characteristics, which will indeed synthesize more natural-sounding speech than that can be produced by a diminished set of controlled units. It is generally used method. The database contains phones in the desired language. The database should be labeled properly in order to make good quality synthesized speech.

ii. Diphone Synthesis

The Diphone synthesis method uses a relatively small size database containing all the diphones in the desired language. It consists of a single sample of each set in the desired language. The quantity of diphones in database depends on the phonotactics of the language .This method doesn't work well in the language where there is a lot of inconsequence in the pronunciation rules and in special cases where letters is pronounced differently than in general. It plays well in the languages where there are more consistencies in the pronunciation. The quality of the synthesized speech is comparatively lower than that of unit selection synthesis.

iii. Domain-specific Synthesis

Domain-specific synthesis consists of phrases and sentences related to a specific application. It can be used in the applications like railway announcement system, weather reports and many more.

iv. Pitch and Duration

There are various parameters in speech that can be used for analysis purpose. Pitch is one of the most important parameters for speech signal processing, including speech synthesis, automatic speech recognition, speech enhancement. In this paper pitch and duration is focused as it affects the concatenation synthesis process. It can provide smoothing effect on the concatenation points. Thus, it is very important to extract the pitch from the speech accurately. Pitch, in speech, is the relative highness or lowness of the tone as perceived by the ear, which depends on the number of vibrations per second produced by the vocal cords [6]. Pitch is the main acoustic correlate of tone and intonation. The tone is a variation in the pitch which serves to differentiate one word from another word. Tone is usually used for tonal languages. These languages use limited number of pitch contrasts. The domain of the tones is usually the syllable. Intonation is the melodic pattern of an utterance. It is the primarily a matter of variation in the pitch level of the voice, in some languages such as English, where stress and rhythm are also involved. It conveys differences of the expressive meaning. Duration or time characteristics are another important prosodic features that lead to the perceived naturalness of synthetic speech. Variation in duration serves as a clue to the individuality of a spoken language sound. So the primary goal in duration modeling is to model the duration pattern of natural speech, considering various features that affect the pattern. For analyzing changes in pitch and duration PSOLA method has been implemented.

3. PSOLA (PITCH SYNCHRONOUS OVERLAP AND ADD)

PSOLA (Pitch Synchronous Overlap and Add) is a digital signal processing technique used for speech processing and more specifically speech synthesis. It can be applied to change the pitch and duration of a speech signal. PSOLA is used for smooth concatenation of pre-recorded speech samples and to provide control for pitch and duration. It is essentially a method used for manipulation of pitch in speech signal. It reads the input signal's pitch and shift it upward or downwards [7] [8].

Basically, there are three versions of PSOLA namely TD-PSOLA (Time Domain-PSOLA), FD-PSOLA (Frequency Domain -PSOLA), LP-PSOLA (Linear-Predictive PSOLA). The Time-Domain version, is the most commonly used due to its computational efficiency. The LP-PSOLA is theoretically more appropriate approach for pitch-scale modifications because they provide independent control over the spectral envelope of the synthesis signal and the FD-PSOLA is used only for pitch-scale modifications. TD-PSOLA is commonly used due to its computational efficiency.

Pitch synchronous speech synthesis algorithms require the starting position of the pitch mark for every voiced segment prior to speech synthesis [9]. In Festival speech synthesis platform Linear-Predictive Coding (LPC) resynthesize is used. Pitch-marks are particularly important in prosodic modification algorithms that use a method known as pitch synchronous overlap-and add (PSOLA) to vary the time and pitch scale of a speech signal [10].

There are two major techniques for acquiring pitch-marks, these are:

1. From an electroglottograph signal, and
2. Algorithms extracting the pitch-marks directly from the speech signal.

In this work the second algorithm is applied which consists of three steps. The analysis step where the original speech signal is first separated into separate but often overlapping short-term analysis signals (ST), the alteration of each analysis, signal to synthesis, signal, and the synthesis step where these segments are recombined by means of overlapping-adding. Short term signals xm (n) are obtained from digital speech waveform xn (n) by multiplying the signal by a sequence of the pitch-synchronous analysis window hm (n).

xm (n) =hm (tm -n)x(n) ------ equ1

Where m is an index for the short-time signal. The windows, which are usually Hanning type, are centered on the successive instants tm, called pitch-marks. A pitch-mark is also referred as pitch period is defined as the location of the short-time peak of each pitch pulse in a speech signal, which means the beginning and end of the speech signal. These targets are put at a pitch-synchronous rate in the voiced parts of the signal and at a constant rate on the unvoiced parts. This pitch pulse corresponds to the glottal closure instant (GCI). These marks are set at a pitch-synchronous rate on the voiced parts of the signal and at a constant rate on the unvoiced parts .The used window length is relative to the local pitch period and the window factor is normally from 2 to 4. The segment recombination in synthesis step is performed after setting a new pitch-mark sequence. Manipulation of the fundamental frequency is attained by varying the time intervals between pitch markers. The alteration of duration is achieved by either repeating or omitting speech signal [11].

4. SPEECH CORPUS

The speech database collected for the analysis consists of 60 sample speech signal collected from a male speaker and a female speaker. Each speaker has uttered 30 sentences. These sentences are collected from stories. The database collection

has been exercised in a recording studio in the afternoon session with a complete noise-free environment. After collecting speech samples, these sentences have undergone through 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 technical specification is drawn in a table 1and the sentences used has been evinced in the table 2 and table 3 with their corresponding labels [12][13][14]..

Table 1 shows 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	Male:1 Female:1

Table 2 shows English sentences used for Pitch Estimation in a 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
3	He was a head shorter than his companion, of almost delicate physique	R003	RS003
4	There was a change now	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 favor	R008	RS008
9	I had faith in them	R009	RS009
10	She turned in at the hotel	R010	RS010
11	I was the only one who remained sitting	R011	RS011
12	We'll have to watch our chances	R012	RS012
13	The ship should be in within a week or ten days	R013	RS013
14	I suppose you wonder why she is coming up here	R014	RS014
15	He moved away as quietly as he had come	R015	RS015
16	It was a curious coincidence	R016	RS016
17	Suddenly his fingers closed tightly over the handkerchief	R017	RS017
18	There was nothing on the rock	R018	RS018
19	There has been a change, she interrupted him	R019	RS019
20	Each day she became a more vital part of him	R020	RS020
21	It was a temptation, but he resisted it	R021	RS021
22	Does that look good	R022	RS022
23	Very early in my life, I separated from my mother	R023	RS023
24	This is no place for you	R024	RS024

25	They only lifted seven hundred and fifty	R025	RS025
26	I'll go over tomorrow afternoon	R026	RS026
27	But it won't continue, she said with easy confidence	R027	RS027
28	The night was calm and snowy	R028	RS028
29	He had comparatively no advantages at first	R029	RS029
30	Without them he could not run his empire	R030	RS030

Table 3 shows English sentences used for Pitch Estimation 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
11	The journey was continued at dawn	B011	BS011
12	She saw the answer in his face	B012	BS012
13	In it was the joy of life	B013	BS013
14	That is the strange part of it	B014	BS014
15	Then he shouted, Shut up	B015	BS015
16	Also, I want information	B016	BS016
17	Let them go out and eat with my boys	B017	BS017
18	You were destroying my life	B018	BS018
19	You read the quotations in today's paper	B019	BS019
20	Let us talk it over and find a way out	B020	BS020
21	It is a good property, and worth more than that	B021	BS021
22	Again he had done the big thing	B022	BS022
23	Daylight was tired, profoundly tired	B023	BS023
24	Please do not think that I already know it all	B024	BS024
25	You have all the advantage	B025	BS025
26	They are not biologists nor sociologists	B026	BS026
27	I want to know how all this is possible	B027	BS027
28	You have all the advantage	B028	BS028
29	And as never before, he was ready to obey	B029	BS029
30	My age, in years, is twenty two	B030	BS030

5. STATISTICAL ANALYSIS

The pitch values contain the speaker specific information. The pitch variation carries the intonation signal associated with rhythms of speech, speaking manner, emotions and accent. The gender is one of the factor which convey a part in characterization of vocal tract. Randomly, the average pitch for female is about 200 Hz and for male it is about 110Hz.Pitch variation is often correlated with loudness and

lowness in speech samples. The experimentation is done by counting on the prosodic features as pitch, duration, frequency, fundamental frequency values for a male and a female speaker

These values are calculated using Computerized Speech Lab (CSL). For Pitch analysis, statistical techniques such as mean, standard deviation, root mean square (rms), geometric mean are used. It is clearly visible in table number 4, 5, 6, & 7 where 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. According to PSOLA algorithm, the values of mean frequency and fundamental frequency F0 is either increased or decreased according the necessity. From the table it can also be seen that the frequency is reduced while mean period is increased and vice versa. The standard deviation indicates the dispersion of pitch values from its mean value whereas median shows the central values of the respective speech samples. Root mean square (RMS) is a useful measure when the data are negative and positive such as sinusoids. It is also a quadratic measure and calculated using values of the samples. A geometric mean is often used when comparing different items finding a single "figure of merit" for these items when each item has multiple properties that have different numeric ranges [12][13][14].

Table 4 shows the values of various statistical measures used in the pitch analysis for Female subject for 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
B011	2.64	100.18	285.36	213.25	203.86	4.91	40.65	223.3	217.05	208.89
B012	2.71	103.57	304.38	225.94	219.2	4.56	34.03	228.46	228.45	222.93
B013	2.68	92.59	272.84	201.22	191.99	5.21	37.48	205.72	204.64	197.09
B014	2.10	155.79	289.11	219.9	215.23	4.65	32.32	215.9	222.21	217.56
B015	2.06	121.72	289.47	217.04	210.45	4.75	35.96	218.58	219.94	213.88
B016	2.35	122.56	270.84	212.08	207.78	4.81	28.41	208.18	213.95	210.04
B017	3.18	95.77	288.73	206.42	198.23	5.04	38.48	202.71	209.93	202.55
B018	2.23	87.28	296.5	215.28	205.5	4.87	40.05	214.04	218.93	210.93
B019	3.55	72.87	279.23	211.87	203.95	4.9	31.74	212.94	214.21	208.69
B020	3.45	102.01	240.97	199.33	195.37	5.12	24.22	201.22	200.78	197.57
B021	3.94	132.89	278.78	205.11	200.41	4.99	30.43	204.57	207.34	202.8
B022	2.48	142.97	304.3	215.27	208.74	4.79	36.8	219.73	218.36	212.06
B023	3.34	134.85	300.74	226.59	221.36	4.52	32.87	227.06	228.93	224.07
B024	4.08	143.85	320.99	233.31	227.94	4.39	34.96	230.44	235.89	230.67
B025	3.55	143.5	309.9	221.19	216.7	4.61	31.47	221.14	235.89	218.96
B026	2.96	156.15	280.04	213.4	210.75	4.74	24.41	210.37	223.4	212.06
B027	4.20	103.44	372.71	218.27	212.15	4.71	33.68	221.65	214.77	215.42
B028	2.31	106.48	298.51	217.19	208.94	4.79	39.13	225.11	220.83	213.32
B029	4.12	145.74	288.12	215.09	211.55	4.73	27.05	214.96	216.77	213.35
B030	2.33	129.03	290.84	228.29	222.8	4.49	32.71	235.25	230.59	225.72

Table 5 shows the values of various statistical measures used for the pitch analysis for Female 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)
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
BS011	2.54	88.37	282.28	209.78	197.13	5.07	44.87	224.43	214.48	204.08
BS012	2.69	126.56	287.62	221.68	216.66	4.62	30.74	221.68	223.77	219.33
BS013	2.32	92.06	271.04	201.83	192.83	5.19	37.19	206.18	205.18	197.8
BS014	2.02	137.56	304.85	221.51	215.86	4.63	34.9	218.72	224.19	218.73
BS015	2.06	138.8	288.22	221.81	217.06	4.61	31.57	220.32	224	219.5
BS016	2.33	123.1	273.3	210.58	205.88	4.86	29.66	208.37	212.63	208.35
BS017	3.15	85.42	293.75	203.21	191.78	5.21	43.09	201.18	207.69	198.02
BS018	2.21	81.77	297.06	213.98	203.72	4.91	40.42	217.09	217.71	209.44
BS019	3.53	73.19	292.13	209.13	201.81	4.96	31.8	210.61	211.75	206.29
BS020	3.45	102.01	240.97	199.33	195.37	5.12	24.22	201.22	200.78	197.57
BS021	3.94	132.89	278.78	205.11	200.41	4.99	30.43	204.57	207.34	202.8
BS022	2.48	142.97	304.3	215.27	208.74	4.79	36.8	219.73	218.36	212.06
BS023	3.15	103.13	299.06	222.04	214.04	4.67	38.67	224.04	225.35	218.31
BS024	3.93	133.49	333.64	229.3	220.95	4.53	41.32	231.5	232.96	225.31
BS025	3.52	143.54	312.15	220.78	216.34	4.62	31.57	217.4	223.01	218.56
BS026	2.96	156.15	280.04	213.4	210.75	4.74	24.41	210.37	214.77	212.06
BS027	3.73	80.63	314.7	222.28	216.43	4.62	31.98	223.17	224.55	219.67
BS028	2.29	98.78	321.3	216.18	206.24	4.85	41.57	226.36	220.08	211.64
BS029	3.77	153.66	285.34	214.77	212.41	4.71	22.6	213.61	215.94	213.59
BS030	2.07	126.98	287.75	228.14	221.5	4.51	36.28	236.61	230.96	225

The table number 6 and 7 shows the total time of analysis, the minimum and the maximum value of pitch, mean value of frequency, mean value of the fundamental frequency, mean period, the standard deviation of the pitch values, the median value of pitch, root mean square value of pitch and the geometric mean of the male subject of the original speech sample and the synthesized one [12][13][14].

Table 6 shows the values of various statistical measures used for the pitch analysis for Male subject for 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
R011	2.52	85.62	182.49	132.93	130.53	7.66	17.83	129.01	134.11	131.74
R012	2.26	80.35	160.23	130.65	128.62	7.77	14.93	131.91	131.49	129.7
R013	3.12	78.15	169.69	124.25	121.84	8.21	16.86	125.71	125.37	123.08
R014	3.12	92.09	251.31	132.94	130.26	7.68	21.58	127.86	134.66	131.5
R015	2.79	83.62	172.62	133.75	131.06	7.63	18.2	136.26	134.97	132.45
R016	1.93	75.73	200.95	139.66	134.27	7.45	25.8	140.82	141.82	137.1
R017	3.52	78.34	162.77	129.41	126.75	7.89	17.69	130.68	130.6	128.13
R018	2.38	102.26	171.71	131.42	129.71	7.71	15.41	129.33	132.3	130.55
R019	3.00	105.36	161.17	134.46	133.58	7.49	10.91	132.96	134.9	134.02
R020	2.56	73.62	257.96	131.86	127.89	7.82	24.71	132.6	134.13	129.83
R021	1.26	80.41	151.82	129.32	126.62	7.9	16.65	134.75	130.36	128.08
R022	2.98	93.49	230.14	140.61	137.15	7.29	22.85	142.38	142.44	138.85
R023	3.12	98.96	176.27	139.88	137.97	7.25	16.11	138.99	140.79	138.93
R024	1.82	94.8	285.32	133.87	129.08	7.75	31.66	130.8	137.5	131.18
R025	2.59	107.33	256.43	139.05	136.62	7.32	21.99	134.13	140.76	137.71
R026	2.17	82.87	170.37	131.71	128.13	7.8	20.05	132.72	133.2	130.03
R027	3.76	88.52	172.97	135.42	133.17	7.51	16.76	135.52	136.44	134.33
R028	2.15	81.37	169.36	130.9	127.06	7.87	20.94	131.28	132.55	129.08
R029	2.13	102.94	26924	146.79	142.04	7.04	29.52	141.34	149.69	144.27
R030	4.22	98.13	251.4	142.69	139.87	7.15	22.56	140.2	144.45	141.19

Table 7 shows the values of various statistical measures used for the pitch analysis for Female 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
RS011	2.50	85.88	273.35	134.9	130.98	7.63	26.25	129.5	137.4	132.82
RS012	2.24	92.02	233.27	136.68	133.28	7.5	24.42	133.45	138.81	134.87
RS013	3.10	78.8	178.87	124.51	122.13	8.19	17.25	125.45	125.68	123.33
RS014	3.10	109.86	160.44	131.04	129.93	7.7	12.44	127.3	131.63	130.48
RS015	2.77	98.38	177.05	135.42	133.47	7.49	16.41	134.92	136.4	134.44
RS016	3.35	88.45	182.49	137.17	134.84	7.42	17.33	137.29	138.25	136.04
RS017	1.89	86.49	186.05	142.59	139.03	7.19	21.39	141.2	144.15	140.89
RS018	3.50	78.22	338.41	133.34	128.38	7.79	32.77	131.23	137.27	130.57
RS019	1.77	100.63	283.25	139.04	134.28	7.45	32.17	131.96	142.64	136.35
RS020	2.57	82.33	210.66	133.85	130.25	7.68	22.15	132.9	135.64	132.05
RS021	2.97	105.67	161.33	135.53	134.66	7.43	10.97	132.68	135.97	135.1
RS022	2.56	90.51	160.62	129.96	127.67	7.83	16.83	132.14	131.03	128.84
RS023	1.24	74.37	259.59	135.29	129.08	7.75	30.71	135.21	138.65	132.18
RS024	3.05	70.8	220.68	140.47	136.22	7.34	23.55	142.56	142.41	138.48
RS025	2.95	104.64	176.15	141.69	139.95	7.15	15.59	140.27	142.54	140.83
RS026	1.80	96.08	150.38	128.59	126.91	7.88	14.26	130.74	129.36	127.77
RS027	2.63	103.73	324.19	142.06	138.23	7.23	31.2	133.47	145.4	139.82
RS028	2.15	81.37	169.36	130.9	127.06	7.87	20.94	131.28	132.55	129.08
RS029	1.88	102.34	240.9	140.69	137.96	7.25	22.05	139.46	142.38	139.24
RS030	2.92	100.69	196.48	147.15	143.84	6.95	22.04	146.34	148.78	145.5

6. CONCLUSION

From the analysis, it is found that pitch and duration affects the synthesis process. In the table 4, 5, 6 and 7 the significance difference can be seen in the original speech and the synthesized speech. It is found that some values of pitch are raised and some are lowered in order to improve the quality of speech, which is useful in smoothly concatenating the words. A good quality speech can in turn provide good quality synthesized speech. Thus it can be said that pitch marking process is really very essential to get good quality synthesized speech.

7. REFERENCES

[1] Archana Balyan, S. S. Agrawal, Amita Dev,"Speech Synthesis: A review", IJERT, vol.2 Issue 6, June 20013.

[2] A. Indumathi, Dr. E. Chandra," Survey on speech synthesis", Signal Processing: An International Journal (SPIJ), Volume (6): Issue (5): 2012.

[3] Shruti Gupta, Prateek Kumar, "Comparative study of text to speech system for Indian Language", International Journal of Advances in Computing and Information Technology ISSN 2277-9140 April 2012.

[4] D.Sasirekha, E.chandra," Text to Speech:A Simple Turorial", International Journal of Soft Computing and Engineering(IJSCE),ISSN:2231-2307,Volume-2,Issue-1,March 2012.

[5] Mahwash Ahmed,Shibli Nisar,"Text-to-Speech using Phoneme Concatenation", International Journal of Scientific Engineering and Technology,Vol 3,Feb 2014.

[6] Allum Mousa,"Voice Conversion Using Pitch shifting algorithm by time stretching with PSOLA and Re-Sampling",Journal of Electrical Engineering Vol.61.No1,2010.

[7] JodoP.Cabra,LuisC.Oliveria,"Pitch-Synchronous Time-Scaling for Prodosic and VoiceQuality bhaTransformations",INTESPEECH 2005.

[8] R.Muralishankar,A.G.Ramakrishana and P.Prathibha, "Modification of Pitch using DCT in the Source Domain",Elsevier-speech communication,vol-42,Feb 2004.

[9] Ulrich Germann,"An Iterative Approach to Pitch-marking of speech signals without Electroglottographic Data,CiteSeer 5M,2006

[10] H.Hussien,M.Wolff,O.Jokisch,F.Duckhorn,G.Strecha and R.Hoffmann,"A Hybrid Speech Signal Based Algorithm for Ptich Marking Using Finite State Machines,INTERSPEECH 2008.

[11] Anant Bhatt,"A PSOLA based Apporach for Voice Morphing",IJDACR,Feb-2015

[12] Kavita Waghmare, Reena H. Chaudhari, Bharti W. Gawali, "Accent identification using MFCC for Hindi Language", Advances in Computational Research, Volume 7, Issue 1, 23 January 2015.

[13] Reena H. Chaudhari, Kavita Waghmare, Bharti W. Gawali, "Accent Recognition using MFCC and LPC with Acoustic Features", International Journal of Innovative Research in Computer and Communication Engineering , Vol. 3, Issue 3, 9 March 2015.

[14] Sangramsing Kayte, Kavita Waghmare, Dr. Bharti Gawali "Marathi Speech Synthesis: A review" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 3 Issue: 6 3708 – 3711, 24 June 2015 (Impact Factor 5.837