ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

Effect of Excitation Signal Frequency on Synthesis of Vowel

Rajeev Saini[1], Randhir Singh[2], Ritika[3]

M.Tech. Student, Dept. of ECE, Sri Sai College of Engineering & Technology, Punjab, India[1]

Associate Professor, Dept. of ECE, Sri Sai College of Engineering & Technology, Punjab, India[2#]

Assistant Professor, Dept. of ECE, Sri Sai College of Engineering & Technology, Punjab, India[3]

ABSTRACT: Producing speech by means other than the natural vocal apparatus has long fascinated people. The nature of the vocal-cord excitation has long interested speech researcher. One of the most important problems has been the specification of source excitations for speech synthesizers. Research work is carried out to investigate the effects of various glottal wave parameters on the synthesis of speech. Effect of frequency of excitation signal on the synthesis of vowels were carried out. It was observed from the results that frequency has random behavior on speech parameter effect on synthesized speech signal.

KEYWORDS: Synthesis, Vowel, Excitation Signal, Glottal, Frequency.

I.INTRODUCTION

The vocalized form of human communication is speech. Speech is based upon the syntactic combination of lexicals and names that are drawn from very large (usually about 1,000 different words) vocabularies. Each spoken word is created out of the phonetic combination of a limited set of vowel and consonant speech sound units. These vocabularies, the syntax which structures them and their set of speech sound units differ, creating the existence of many thousands of different types of mutually unintelligible human languages. Most human speakers are able to communicate in two or more of them, hence being polyglots. The vocal abilities that enable humans to produce speech also provide humans with the ability to sing. A gestural form of human communication exists for the deaf in the form of sign language. Speech in some cultures has become the basis of a written language, often one that differs in its vocabulary, syntax and phonetics from its associated spoken one, a situation called diglossia. Speech in addition to its use in communication, it is suggested by some psychologists that it is also internally used by mental processes to enhance and organize cognition in the form of an interior monologue. Speech is researched in terms of the speech production and speech perception of the sounds used in vocal language. Other research topics concern speech repetition, the ability to map heard spoken words into the vocalizations needed to recreate them, that plays a key role in the vocabulary expansion in children and speech errors. Several academic disciplines study these including acoustics, psychology, speech pathology, linguistics, cognitive science, communication studies, otolaryngology and computer science [1-4].

Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech computer or speech synthesizer, and can be implemented in software or hardware products. A text-to-speech (TTS) system converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech. Synthesized speech can be created by concatenating pieces of recorded speech that are stored in a database. Systems differ in the size of the stored speech units; a system that stores phones or diphones provides the largest output range, but may lack clarity. For specific usage domains, the storage of entire words or sentences allows for high-quality output. Alternatively, a synthesizer can incorporate a model of the vocal tract and other human voice characteristics to create a completely "synthetic" voice output. The quality of a speech synthesizer is judged by its similarity to the human voice and by its ability to be understood clearly. An intelligible text-to-speech program allows people with visual impairments or reading disabilities to listen to written works on a home computer. Many computer operating systems have included speech synthesizers since the early 1990s. A text-to-speech system (or "engine") is composed of two parts: a front-end and a back-end. The front-end has two major tasks. First, it converts raw text containing symbols like numbers and abbreviations into the equivalent of written-out words. This process is often called text normalization, pre-processing, or tokenization. The front-end then

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

assigns phonetic transcriptions to each word, and divides and marks the text into prosodic units, like phrases, clauses, and sentences. The process of assigning phonetic transcriptions to words is called text-to-phoneme or grapheme-to-phoneme conversion. Phonetic transcriptions and prosody information together make up the symbolic linguistic representation that is output by the front-end. The back-end often referred to as the synthesizer then converts the symbolic linguistic representation into sound. In certain systems, this part includes the computation of the target prosody (pitch contour, phoneme durations), which is then imposed on the output speech [5-8].

Synthetic speech may be used in several applications. The application field of synthetic speech is expanding fast whilst the quality of TTS systems is also increasing steadily. Speech synthesis systems are also becoming more affordable for common customers, which makes these systems more suitable for everyday use. For example, better availability of TTS systems may increase employing possibilities for people with communication difficulties.

Applications for the Blind: Probably the most important and useful application field in speech synthesis is the reading and communication aids for the blind. Before synthesized speech, specific audio books were used where the content of the book was read into audio tape. It is clear that making such spoken copy of any large book takes several months and is very expensive. It is also easier to get information from computer with speech instead of using special bliss symbol keyboard, which is an interface for reading the Braille characters.

Applications for the Deafened and Vocally Handicapped: People who are born-deaf cannot learn to speak properly and people with hearing difficulties have usually speaking difficulties. Synthesized speech gives the deafened and vocally handicapped an opportunity to communicate with people who do not understand the sign language. With a talking head it is possible to improve the quality of the communication situation even more because the visual information is the most important with the deaf and dumb. A speech synthesis system may also be used with communication over the telephone line.

Educational Applications: Synthesized speech can be used also in many educational situations. A computer with speech synthesizer can teach 24 hours a day and 365 days a year. It can be programmed for special tasks like spelling and pronunciation teaching for different languages. It can also be used with interactive educational applications.

Applications for Telecommunications and Multimedia: The newest applications in speech synthesis are in the area of multimedia. Synthesized speech has been used for decades in all kind of telephone enquiry systems, but the quality has been far from good for common customers. Today, the quality has reached the level that normal customers are adopting it for everyday use [9-10].

II.METHODOLOGY

The research work is formulated to analyze the effect of various glottal stimulus parameters on the speech generation. During the mechanism of phonation, airflow is evicted from the lungs, arises in the trachea and is modulated by its passage through the space delimited by the vocal folds, called the glottis. Speech then results from filtering this so-called glottal flow by the vocal tract cavities, and converting the resulting velocity flow into pressure at the lips. In many speech processing applications, it is important to separate the contributions from the glottis and the vocal tract. In the first experiment effect of glottal wave frequency on synthesis of vowels were investigated with all other parameters were kept at their standard values i.e. open quotient = 0.50, spectral tilt = 0.02, and shape quotient = 3.00. Frequency from 60 Hz to 520 Hz was taken for the synthesis process. Various parameters such as minimum and maximum pitch, intensity, and first formant frequency of synthesized speech are compared with original speech. The structural parameters of the vocal tract for all the experiments are given in Table 1.

Table 1 Vocal tract size parameter specification.

Vowels	Length (cm)	Velum Position (cm)	Curvature Area (cm^2)	Teeth Position (cm)	Minimum Area (cm)
a	16.40	7.31	3.597	15.44	0.340
e	15.72	7.10	0.836	15.26	0.334
i	16.72	8.27	2.297	16.25	0.279
o	17.97	8.28	0.348	16.62	0.260
u	18.70	8.66	0.538	17.03	0.211

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

III.RESULTS AND DISCUSSION

In this experiment effect of frequency on the synthesis of vowels were investigated. Five vowels ‘a’, ‘e’, ‘i’, ‘o’, and ‘u’ were synthesis with various glottal wave frequency ranging from 60 Hz to 520 Hz frequency. The synthesis vowels were analyzed used Praat tool. Various speech parameters such as intensity, maximum and minimum pitch, first format frequency were computed. Table 2 shows the various speech parameters with respect to glottal wave frequency for vowel ‘a’. It is observed from the value observation table that with increase in glottal wave frequency speech parameter shows random behavior. Maximum pitch is seen at 478 Hz glottal wave frequency. Maximum intensity of 88 dB is observed at 520 Hz. Highest first formant frequency of 683 Hz is observed at 359 Hz.

Table 2 Speech parameters of synthesized vowel ‘a’.

Glottal Waveform Frequency (Hz)	Maximum Pitch (Hz)	Minimum Pitch (Hz)	Intensity (db)	First Formant Frequency (Hz)
60.00	277.331	267.671	62.490	537.462
120.00	121.225	77.255	64.114	524.167
179.60	181.439	158.911	84.873	537.719
239.40	242.341	213.144	73.081	505.908
299.20	302.072	265.191	79.180	597.548
359.00	364.476	314.302	71.573	683.009
418.80	424.082	374.776	73.706	621.868
478.60	484.677	423.770	78.673	536.342
520.00	265.077	232.726	88.525	597.297

Various parameters of synthesized speech are tabulated in Table 3. It is observed from the table that maximum intensity of 484 Hz is obtained 478 Hz frequency. Maximum intensity of 88 dB is observed at 520 Hz glottal wave frequency. Maximum first formant frequency of 682 Hz is observed for synthesized ‘e’ vowel.

Table 3 Speech parameters of synthesized vowel ‘e’.

Glottal Waveform Frequency (Hz)	Maximum Pitch (Hz)	Minimum Pitch (Hz)	Intensity (db)	First Formant Frequency (Hz)
60.00	277.453	267.871	62.490	537.803
120.00	121.289	105.164	64.112	524.185
179.60	181.446	157.303	84.873	537.605
239.40	242.347	210.596	73.079	506.094
299.20	302.067	263.057	79.175	597.171
359.00	364.476	314.352	71.572	682.988
418.80	424.085	374.605	73.708	621.850
478.60	484.677	423.554	78.674	536.363
520.00	265.082	232.701	88.525	597.130

Various parameters of synthesized speech are tabulated in Table 4. It is observed from the table that maximum intensity of 484 Hz is obtained 478 Hz frequency. Maximum intensity of 88.5 dB is observed at 520 Hz glottal wave frequency. Maximum first formant frequency of 682 Hz is observed for synthesized ‘e’ vowel.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

Table 4 Speech parameters of synthesized vowel 'i'.

Glottal Waveform Frequency (Hz)	Maximum Pitch (Hz)	Minimum Pitch (Hz)	Intensity (db)	First Formant Frequency (Hz)
60.00	277.472	267.676	62.490	537.412
120.00	121.225	77.255	64.114	524.167
179.60	181.439	158.927	84.873	537.723
239.40	242.346	210.466	73.080	506.121
299.20	302.067	263.057	79.175	597.171
359.00	364.476	314.352	71.572	682.988
418.80	424.085	374.605	73.708	621.850
478.60	484.677	423.554	78.674	536.363
520.00	264.849	230.120	88.525	605.135

Various parameters of synthesized speech are tabulated in Table 5. It is observed from the table that maximum intensity of 485 Hz is obtained 478 Hz frequency. Maximum intensity of 88.52 dB is observed at 520 Hz glottal wave frequency. Maximum first formant frequency of 683 Hz is observed for synthesized 'o' vowel.

Table 5 Speech parameters of synthesized vowel 'o'.

Glottal Waveform Frequency (Hz)	Maximum Pitch (Hz)	Minimum Pitch (Hz)	Intensity (db)	First Formant Frequency (Hz)
60.00	277.388	267.672	62.490	537.448
120.00	121.290	105.199	64.109	524.179
179.60	181.439	158.911	84.873	537.719
239.40	242.347	210.544	73.080	506.111
299.20	302.072	265.191	79.180	597.548
359.00	364.464	316.992	71.575	683.212
418.80	424.085	374.605	73.708	621.850
478.60	485.060	427.987	78.674	535.979
520.00	265.082	232.701	88.525	597.130

Various parameters of synthesized speech is tabulated in Table 6. It is observed from the table that maximum intensity of 485.06 Hz is obtained 478 Hz frequency. Maximum intensity of 88.525 dB is observed at 520 Hz glottal wave frequency. Maximum first formant frequency of 683.009 Hz is observed for synthesized 'o' vowel.

Table 6 Speech parameters of synthesized vowel 'u'.

Glottal Waveform Frequency (Hz)	Maximum Pitch (Hz)	Minimum Pitch (Hz)	Intensity (db)	First Formant Frequency (Hz)
60.00	277.388	267.672	62.490	537.448
120.00	121.225	77.255	64.114	524.167
179.60	181.445	157.273	84.873	537.603
239.40	242.341	213.144	73.081	505.908
299.20	302.072	265.169	79.181	597.538
359.00	364.476	314.302	71.573	683.009
418.80	424.043	370.259	73.702	622.146
478.60	485.068	427.871	78.674	535.994
520.00	264.849	230.120	88.525	605.135

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

IV.CONCLUSION

Speech processing technology continues to fascinate engineers and researchers to incorporate natural speech in human-machine interaction. The dynamic model is based on a mathematical formulation of the human vocal system from the glottis to the lips. The research work is carried out to investigate the effect of glottal waveform parameters on the synthesis of vowels. Frequency of excitation signal is an important factor in the synthesis process. Investigations were carried out with range 60 Hz to 520 Hz of excitation frequency and vowels were synthesized at different frequency. While frequency is varied and other parameters such as open quotient = 0.50, spectral tilt = 0.02, and shape quotient = 3.00 of glottal wave was kept at standard values. It is observed form the results that frequency has same effect on every synthesized vowel and maximum value of pitch, intensity, and first formant frequency is obtained at 478 Hz glottal wave frequency when various speech parameters were analyzed.

REFERENCES

[1] R. D. Kent, "The MIT encyclopedia of communication disorders," *Choice Reviews Online*, vol. 41, no. 10, pp. 41–5649–41–5649, Jun. 2004.
[2] R. P. Lippmann, "Speech recognition by machines and humans," *Speech Communication*, vol. 22, no. 1, pp. 1–15, Jul. 1997.
[3] E. Moulines and F. Charpentier, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Speech Communication*, vol. 9, no. 5–6, pp. 453–467, Dec. 1990.
[4] P. Taylor, A. Black, and R. Caley. "The architecture of the Festival speech synthesis system", 1998.
[5] D. H. Klatt, "Review of text-to-speech conversion for English," *J. Acoust. Soc. Am.*, vol. 82, no. 3, p. 737, 1987.
[6] M. R. Schroeder, "A brief history of synthetic speech," *Speech Communication*, vol. 13, no. 1–2, pp. 231–237, Oct. 1993.
[7] M. Honda, "Speech synthesis technique based on speech production mechanism", *Acoustical Society of Japan*, pp.777-82,1999
[8] S. Saito and K. Nakata, "Fundamentals of speech signal processing", Academic Pr, 1985.
[9] J. Flanagan, "Speech Analysis, Synthesis, and Perception," Springer-Verlag, 1972.
[10] W.T. Fitch, J. Giedd, "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," *The Journal of the Acoustical Society of America*, vol. 106, pp. 1511-1522, 1999.