ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

Effect of Open Quotient Parameter of Glottal Wave Excitation Signal on the 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: The availability of synthesis systems, such as the Festival Speech Synthesis System, MBROLA project, HMM-based speech synthesis system, and Vocal Tract Lab makes the cost of entering the field of speech synthesis much lower, allowing many research groups to join the development. 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. Experiment with open quotient parameter of the excitation signal was varied and vowel 'a' was synthesized and analyzed. It was observed that open quotient is inversely proportional to speech parameters.

KEYWORDS: Vowel, Synthesis, Glottal, Pitch, Excitation.

I. INTRODUCTION

Speech is a natural form of communication for human beings, and computers with the ability to understand speech and speak with a human voice are expected to contribute to the development of more natural man-machine interfaces [1]. Computers with this kind of ability are gradually becoming a reality, through the evolution of speech synthesis and speech recognition technologies. However, in order to give them functions that are even closer to those of human beings, we must learn more about the mechanisms by which speech is produced and perceived, and develop speech information processing technologies that make use of these functions [2]. We use speech every day almost unconsciously, but an understanding of the mechanisms on which it is based will help to clarify how the brain processes information and will also lead to the development of more human-like speech devices through the imitation of these functions by computers. The mechanism of speech is composed of four processes: language processing, in which the content of an utterance is converted into phonemic symbols in the brain's language center; generation of motor commands to the vocal organs in the brain's motor center; articulatory movement for the production of speech by the vocal organs based on these motor commands; and the emission of air sent from the lungs in the form of speech [3].

The speech production process has many levels, from the movement of vocal organs to the production of sounds [4]. There are four hierarchical levels in speech production: the speech sound level, vocal tract shape level, vocal organ configuration level, and muscle contraction level. There is a "one-to-many" relationship among levels starting from the sound level and moving toward the muscle contraction level. For example, as is evident in the case of ventriloquism, sounds that seem very similar can be created using different vocal tract shapes [5]. Similar vocal tract shapes can be created using differing vocal organ configurations; for example, the degree of mouth opening is determined by the relative position of both lips and the jaw. Furthermore, each individual vocal organ involves two or more competing muscles, and the vocal organ configuration is determined by their relative degree of contraction. This means that when speech with a given tone is created, the "one-to-many" relationship that exists among levels cannot be determined uniquely because there is always an excessive degree of freedom on the lower level. Another feature of articulatory movement is that this is not a simple movement, but continuous movement for the purpose of uttering continuous sounds. For example, the movement of the mouth when uttering the sound "api" is not a simple connection

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

of the individual mouth positions (articulations) for the phonemes /a/, /p/, and /i/. The articulation of the consonant /p/ is affected by the articulation of the vowel /i/ that follows it, and when it is uttering /p/, the tongue is already in position for the vowel /i/. The term "coarticulation" is used to refer to the phenomenon in which the mouth position for individual phonemes in the utterance of continuous sounds incorporates the effects of the mouth position for phonemes uttered immediately before and after. This is a very significant feature of articulatory movement [6-8].

Human speech is a combination of sounds and silences generated by the speech mechanism of the vocal tract into meaningful patterns [9]. All speech begins as a silent breath of air, created by muscular activity in the chest. The air then comes up from the lungs, via the vocal tract and exiting as a sound wave.

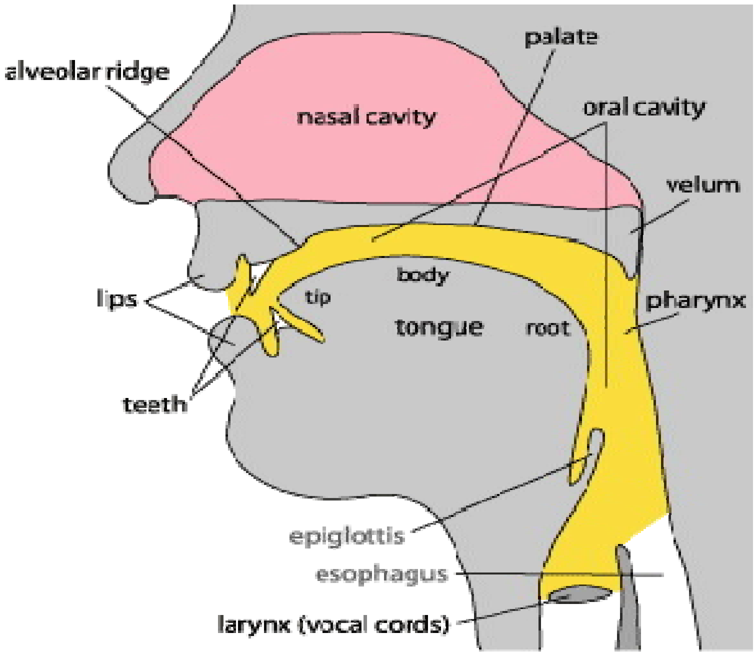

Fig. 1 Anatomy of vocal tract.

Basically, the speech mechanism as four components: muscular activity, air, some type of resistance or obstruction to the air which causes some sort of sound to be made, and amplification to make the sound loud enough to be heard. Changes to the air flow between the lungs and mouth and nose produce different sounds. Air starts off in the lungs, flows up through the trachea (windpipe), through the larynx, past the epiglottis and through the pharynx. From there, the air can go either through the mouth or nose. The Vocal tract is the channel of air flow between the larynx and the mouth and nose [10-12].

The vocal tract is the cavity in human beings and in animals where sound that is produced at the sound source (larynx in mammals; syrinx in birds) is filtered. In birds it consists of the trachea, the syrinx, the oral cavity, the upper part of the esophagus, and the beak. In mammals it consists of the laryngeal cavity, the pharynx, the oral cavity, and the nasal cavity. The estimated average length of the vocal tract in adult male humans is 16.9 cm and 14.1 cm in adult females [13-15].

II. METHODOLOGY

The research work is formulated to analyze the effect of various glottal stimulus parameters on the speech generation. In this research work Vocal Tract Lab (VTL) is implemented. Vocal Tract Lab (VTL) is an articulatory speech synthesizer and a tool to visualize and explore the mechanism of speech production with regard to articulation, acoustics, and control. Experiment is carried out with open quotient parameter is varied from 0.20 to 0.80 with other glottal waveform parameters kept at standard values i.e. frequency = 120 Hz, spectral tilt = 0.02, and shape quotient =

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

3.00. In this experiment only vowel 'a' is synthesized. The synthesized speech is compared with original speech signal. Figure 2 shows the schematics of block diagram of research methodology.

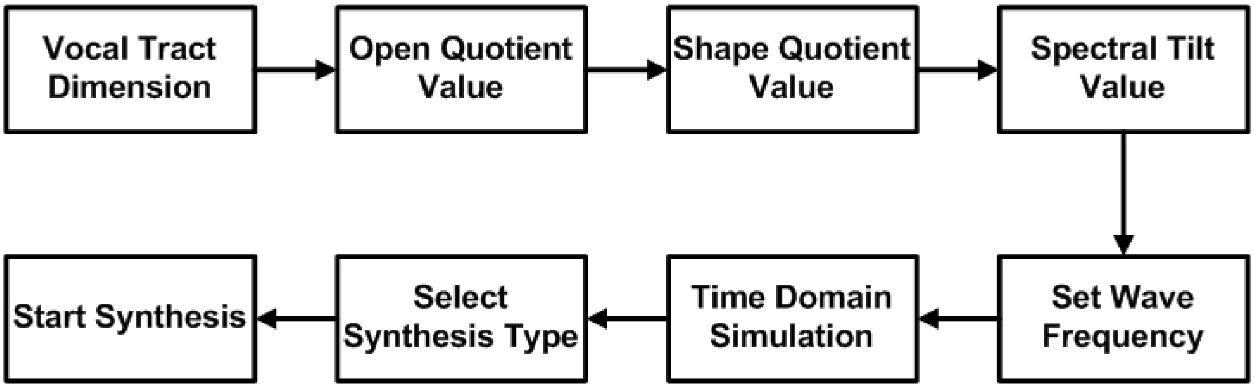

Fig. 2 Block diagram of speech synthesis process using VTL.

III RESULTS AND DISCUSSION

In this experiment effect of open quotient parameter of glottal pulse is varied from 0.20 to 0.80 with rest of parameter fixed at standard values. Vowel 'a' is synthesized in this experiment. Figure 5.1 shows the vocal tract shape for vowel 'a'. Synthesized speech is analyzed in Praat and various speech parameters were computed. Estimated parameters are tabulated in Table 1. It is observed from the parameters that open quotient has inverse effect on the speech parameters i.e speech parameters are inversely proportional to open quotient parameter of glottal pluse. Maximum pitch of 276 Hz, maximum intensity of 72 dB, and maximum first formant frequency of 550 Hz is observed for 0.20 open quotient of glottal pulse.

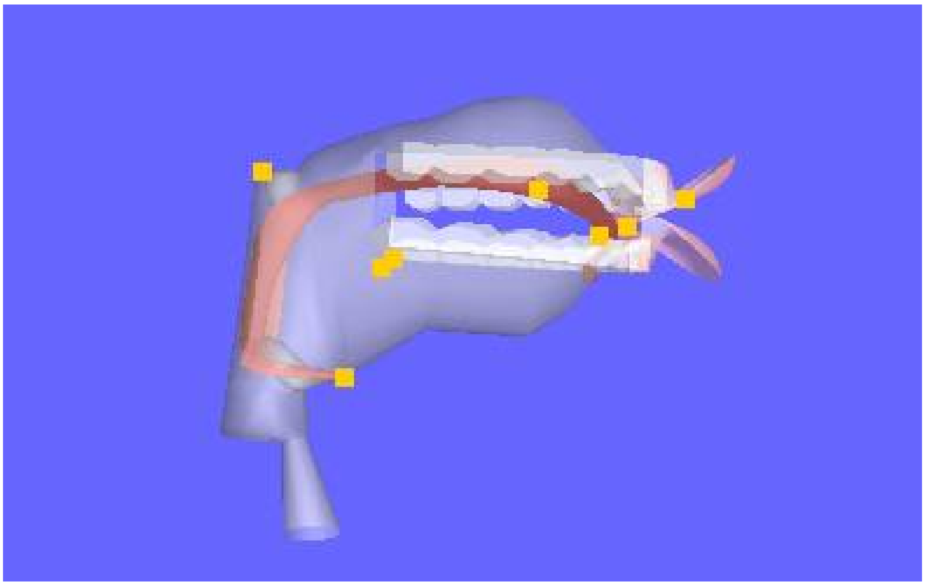

Fig. 3 Vocal tract shape for vowel 'a'.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

Table 1 Speech parameters of synthesized vowel ‘a’ w.r.t open quotient variation.

Open Quotient	Pitch Maximum (Hz)	Pitch Minimum (Hz)	Intensity (dB)	First Formant Frequency (Hz)
0.20	276.173	119.166	72.388	550.811
0.40	275.881	119.183	65.832	529.051
0.60	121.265	105.290	62.857	525.193
0.80	121.404	106.211	60.734	523.824

IV. CONCLUSION

VTL implements various models. The core of the synthesizer is a 3D articulatory model of the vocal tract that defines the shape of the airway between the glottis and the lips. The research work is carried out to investigate the effect of glottal waveform parameters on the synthesis of vowels. ‘a’ was synthesized with open quotient is varied from 0.20 to 0.80 and remaining parameters were at constant values i.e. frequency = 120 Hz, shape quotient = 3.00, and spectral tilt = 0.02 It is observed that the open quotient is inversely proportional to speech parameters. Also maximum value of parameter was obtained at 0.20 value of open quotient.

REFERENCES

[1] J. L. Obenour and D. M. Levy, “A Digital-Computer Simulation of the Human Speech-Production Mechanism,” *IEEE Trans. Biomed. Eng.*, vol. BME-19, no. 2, pp. 114–119, Mar. 1972.

[2] B. Guerin, M. Mrayati, and R. Carre, “ A voice source taking account of coupling with supraglottal cavities,” *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1, pp. 47-50, 1976.

[3] S. Terepin and F. Fallside, “A polynomial vocal tract model for speech synthesis”, *IEEE International Conference on Acoustics, Speech, and Signal Processing*. May, Vol. 7, pp. 919-922, 1982.

[4] Man Sondhi and J. Schroeter, “A hybrid time-frequency domain articulatory speech synthesizer,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 7, pp. 955–967, Jul. 1987.

[5] H. Wu, P. Badin, and B. Guerin, “Vocal tract simulation: Implementation of continuous variations of the length in a Kelly-Lochbaun model, effects of area function spatial sampling”, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 12, pp. 9-12, 1987.

[6] K. Adamson, N. Black, and G. Donnam, “A concurrent acoustical model of the human vocal tract”, *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 6, pp.2568-2569, 1992.

[7] Q. Lin and G. Fant. “An articulatory speech synthesizer based on a frequency-domain simulation of the vocal tract”, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Mar 23, Vol. 2, pp. 57-60, 1992.

[8] K. Jones and J. Harris, “The silicon vocal tract”, *International Conference on Neural Networks*, Vol. 2, pp. 902-907, 1996.

[9] A. Alwan, “Modeling speech production and perception mechanisms and their applications to synthesis, recognition, and coding”, *Fifth International Symposium on Signal Processing and its Applications*, Vol. 1, pp. 7, 1999

[10] P. Gray, M. P. Hollier, and R. E. Massara, “Non-intrusive speech-quality assessment using vocal-tract models,” *IEE Proc., Vis. Image Process.*, vol. 147, no. 6, p. 493, 2000.

[11] D. O’Shaughnessy, “Interacting with computers by voice: automatic speech recognition and synthesis,” *Proc. IEEE*, vol. 91, no. 9, pp. 1272–1305, Sep. 2003.

[12] P. Birkholz and D. Jackel, “Influence of temporal discretization schemes on formant frequencies and bandwidths in time domain simulations of the vocal tract system”, *INTERSPEECH*, Oct., 2004.

[13] P. Birkholz and B. Kroger, “Vocal Tract Model Adaptation Using Magnetic Resonance Imaging”, *International Seminar on Speech Production*, pp. 493-500, 2006.

[14] H. Deng, R. K. Ward, M. P. Beddoes, and M. Hodgson, “A New Method for Obtaining Accurate Estimates of Vocal-Tract Filters and Glottal Waves From Vowel Sounds,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 2, pp. 445–455, Mar. 2006.

[15] S. Prom-on, Y. Xu, and B. Thipakorn, “Modeling tone and intonation in Mandarin and English as a process of target approximation,” *J. Acoust. Soc. Am.*, vol. 125, no. 1, p. 405, 2009.