

Effect of Spectral Tilt Parameter of Excitation Signal on the Synthesis of Vowel /a/

Rajeev Saini¹, Randhir Singh², Ritika³M.Tech. Student, Dept. of ECE, Sri Sai College of Engineering & Technology, Punjab, India¹Associate 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, India³

ABSTRACT: 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. It is known that the shape and periodicity of the vocal-cord excitation are subject to large variations. Investigations were carried out the effects of various glottal wave parameters on the synthesis of speech. Spectral tilt parameter of the excitation signal was varied and vowel 'a' was synthesized and analyzed. Spectral tilt is also inversely proportional to the speech parameter i.e. intensity, pitch, and first formant frequency of the synthesized vowel

KEYWORDS: Vowel, Synthesis, Glottal, Pitch.

I. INTRODUCTION

Voice is one of the most important instruments of human communication and it is such a complex phenomenon that, despite being investigated over the years in different areas as engineering, medicine and singing, not all of its attributes seem to be known. As sound identity of human beings, the voice reflects individual characteristics such as age, sex, race, social status, personal characteristics and even emotional state. It is maybe in the nuances and inflections of the voice that lies the expressive power of human language. Claudius Galen, a Greek second century physician, physiologist and philosopher, studied the voice production and believed that voice was the mirror of the soul [1-3]. Voice is the unique signal generated by the human vocal apparatus and is perceived as the sounds originated from a flow of air from the lungs, which causes the vocal folds to vibrate, and that are subsequently modified by the vocal tract. Very briefly, voice is the result of a balance between two forces: the force of the air leaving the lungs and the muscle strength of the larynx, where the vocal folds are located. As a physical phenomenon, voice is defined as a complex sound, whose voiced regions, i.e., those resulting from a vibration of the vocal folds consist of a fundamental frequency and a large number of harmonics. Several mathematical models have been proposed over the years both to model the voice production system or to estimate the flow of air passing through the glottis (i.e., the space between the vocal folds) called the glottal flow or glottal source [3].

According to the anatomy and physiology of speech production, the glottal flow is the airflow velocity waveform that comes out of the glottis and enters the vocal tract, also called the glottal source. The shape of the glottal source waveform in a single period is denoted by glottal pulse. The glottal flow and the glottal pulse represent the airflow that passes through the glottis and that the glottal source is the glottal pressure wave or, in some cases, the glottal aperture, i.e., the derivative of the glottal pulse [4-5].

As the vocal folds open and close the glottis at identical intervals, the frequency of the sound generated is equal to the frequency of vibration of the vocal folds [6]. Each cycle consists of four glottal phases, as it can be seen in Fig. 1 closed, opening, opened and closing (or return).

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 9, September 2016

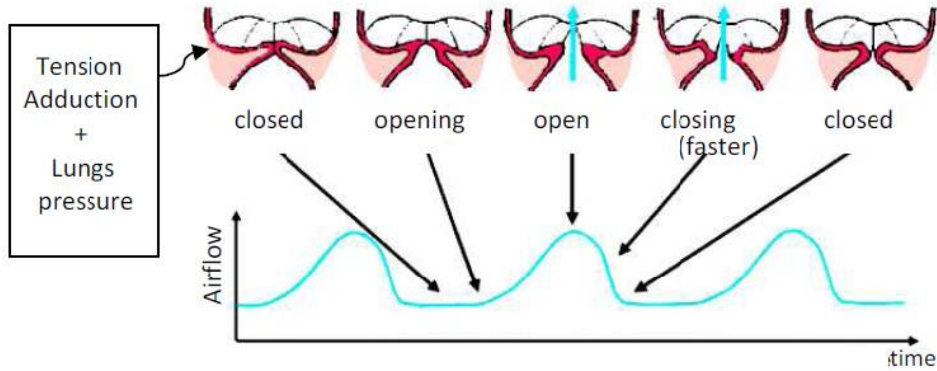


Fig. 1 Glottal phases.

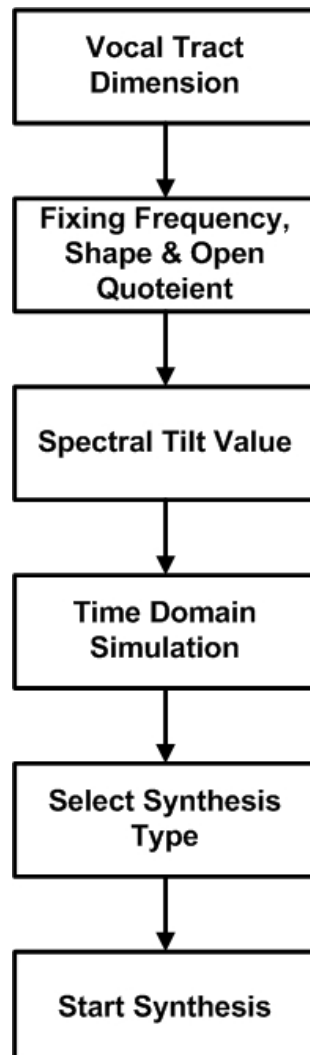


Fig. 2 Synthesis Process.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 9, September 2016

Typically, when the folds are in a closed position, the flow begins slowly, builds up to a maximum, and then quickly decreases to zero when the vocal folds abruptly shut [5]. However, studies have indicated that total closure of the glottis is an idealistic assumption and that the majority of the individuals exhibit some sort of glottal leakage during the assumed closed phase [3]. The time interval during which the vocal folds are closed and no flow occurs is referred to as the glottal closed phase. The next phase, during which there is nonzero flow and up to the maximum of the airflow velocity is called the glottal opening phase, and the time interval from the maximum to the time of glottal closure is referred to as the closing or return phase. Many factors can influence the rate at which the vocal folds oscillate through a closed, open and return cycle, such as the vocal folds muscle tension, the vocal fold mass and the air pressure below the glottis [7-10].

II. RELATED WORK

Research work is carried out using Vocal Tract Lab (VTL) tool. VTL implements various models, for example models for the vocal tract, the vocal folds, the acoustic simulation etc. 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. It has currently 23 degrees of freedom (vocal tract parameters) that control the shape and position of the model articulators. For the simulation of acoustics, the area function of the vocal tract is calculated, i.e., the variation of the cross-sectional area along the center line of the model. The area function is then transformed into a transmission line model of the vocal tract and can be analyzed in the frequency domain or simulated in the time domain. Figure 2 shows the flowchart of synthesis process.

III. RESULTS AND DISCUSSION

In this experiment effect of spectral tilt parameter of glottal pulse is varied from 0.00 to 0.20 with rest of parameter fixed at standard values. Vowel ‘a’ is synthesized in this experiment. 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 random effect on the speech parameters. Maximum pitch of 121.239 Hz, maximum intensity of 65.658 dB, and maximum first formant frequency of 536.287 Hz is observed for 0.00 spectral tilt of glottal pulse. Figure 3 and 4 shows the intensity and first formant frequency of synthesized vowel with respect to spectral tilt.

Table 1 Speech parameters of synthesized vowel ‘a’ w.r.t spectral tilt variation.

A. Spectral Tilt	B. Pitch C. Maximum D. (Hz)	E. Pitch F. Minimum G. (Hz)	H. Intensity I. (dB)	J. First Formant Frequency K. (Hz)
L. 0.00	R. 121.216	X. 77.210	DD. 65.658	JJ. 536.287
M. 0.04	S. 121.329	Y. 105.240	EE. 61.089	KK. 499.910
N. 0.08	T. 121.318	Z. 105.212	FF. 57.820	LL. 462.352
O. 0.12	U. 121.212	AA. 106.034	GG. 56.227	MM. 430.529
P. 0.16	V. 121.239	BB. 106.041	HH. 55.453	NN. 405.356
Q. 0.20	W. 121.199	CC. 104.833	II. 55.083	OO. 388.853

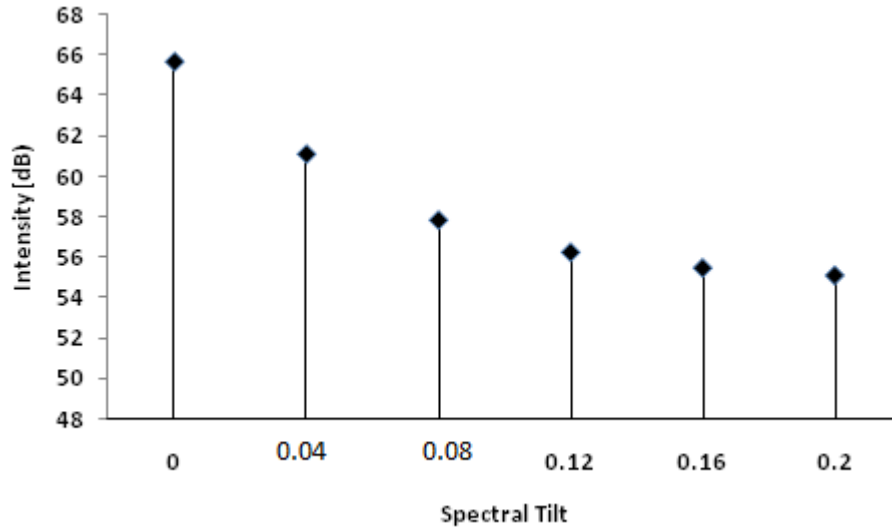


Fig. 3 Intensity of Synthesized vowel.

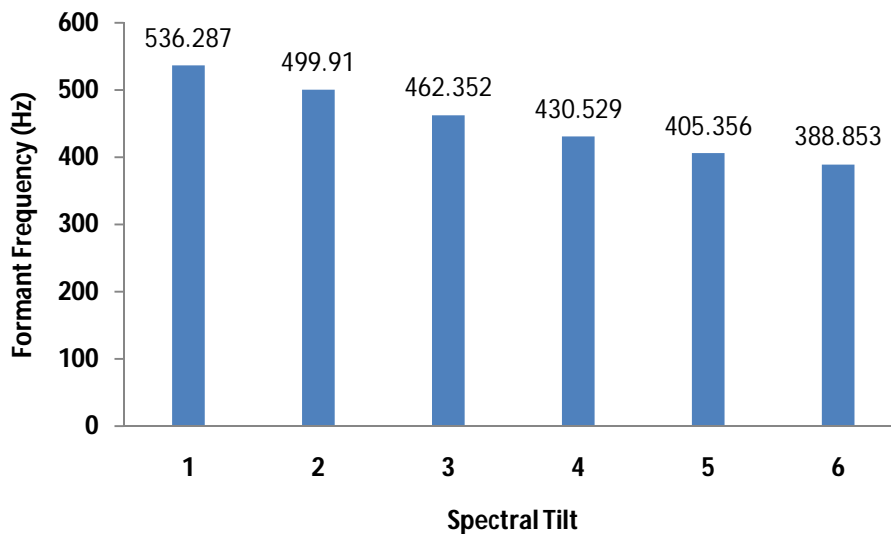


Fig. 4 First formant frequency of Synthesized vowel.

IV. CONCLUSION

The dynamic model is based on a mathematical formulation of the human vocal system from the glottis to the lips. The dynamic model of the articulatory synthesis system has three physiologically derived components: the anatomical/geometric structure, the acoustic wave propagation, and the gestural control. VTL implements various models. The research work is carried out to investigate the effect of glottal waveform parameters on the synthesis of vowels. Vowel 'a' was synthesized with spectral tilt is varied from 0.00 to 0.20 with other parameters kept at standard values. It is observed that the spectral tilt has random effect on the speech parameters. Maximum value of parameter was obtained at 0.00 value of spectral tilt.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 9, September 2016

REFERENCES

- [1] 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.
- [2] D. O'Shaughnessy, "Interacting with computers by voice: automatic speech recognition and synthesis," *Proc. IEEE*, vol. 91, no. 9, pp. 1272–1305, Sep. 2003.
- [3] 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.
- [4] P. Birkholz and B. Kroger, "Vocal Tract Model Adaptation Using Magnetic Resonance Imaging", *International Seminar on Speech Production*, pp. 493-500, 2006.
- [5] 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.
- [6] 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.
- [7] K. S. Rama Murty, B. Yegnanarayana, and M. Anand Joseph, "Characterization of Glottal Activity From Speech Signals," *IEEE Signal Process. Lett.*, vol. 16, no. 6, pp. 469–472, Jun. 2009.
- [8] J. Kane and C. Gobl, "Identifying regions of non-modal phonation using features of the wavelet transform", *INTERSPEECH*, Aug., pp. 177–180, 2011.
- [9] P. Birkholz, B. Kroger, and C. Neuschaefer-Rube, "Synthesis of breathy, normal, and pressed phonation using a two-mass model with a triangular glottis", *INTERSPEECH*, pp. 2681-2684, 2011.
- [10] P. Birkholz, B. J. Kroger, and C. Neuschaefer-Rube, "Model-Based Reproduction of Articulatory Trajectories for Consonant–Vowel Sequences," *IEEE Trans. Audio Speech Lang. Process.*, vol. 19, no. 5, pp. 1422–1433, Jul. 2011.