A DYNAMIC GLOTTAL MODEL THROUGH HIGH-SPEED IMAGING

Authors: Jiangping Kong

Date: 2019-02-27T00:00:00+00:00

Abstract

This paper is a study for an improved dynamic glottal model through high-speed imaging (HSI). As is well known, speech production comprises three parts, namely speech source, speech resonance and lip radiation. Among these three parts, speech source is the most important one because it is the basis of speech. In research on speech production, acoustical models of speech source have been well established. But the physiological speech source, that is to say, the activity of glottis is seldom researched, because the vibration of vocal folds is difficult to observe and sample. A study on glottal model was established many years ago (Kong, 2007), and in that model, the static glottis was modeled by four quarters of ellipses in three modes namely normal mode, leakage mode and open mode. The dynamic glottal control function was modeled by an approximation of multiplication of sine and exponential. The problem of the dynamic glottal model is that the control parameters can' t be well explained, though the glottis can be simulated. In this study, more high-speed images were sampled, the image processing was greatly improved and the dynamic glottal control function was modeled with parameters which were significant to speech perception.

Full Text

Preamble

A DYNAMIC GLOTTAL MODEL THROUGH HIGH-SPEED IMAGING
Jiangping Kong
Center for Chinese Linguistics, Peking University
Department of Chinese Language and Literature, Peking University
Joint Research Center for Language and Human Complexity

ABSTRACT

This paper presents a study on a dynamic glottal model through high-speed imaging (HSI). As is well known, speech production comprises three components: speech source, vocal tract resonance, and lip radiation. Among these, the speech source is the most critical as it forms the foundation of speech. While acoustical models of the speech source have been well established in speech production research, physiological models of the speech source—namely, the activity of the glottis—have seldom been investigated due to the difficulty of observing and sampling vocal fold vibration.

A glottal model was established many years ago (Kong, 2007) that modeled the static glottis using four quarter-ellipses in three modes: normal mode, leakage mode, and open mode. The dynamic glottal control function was approximated by the product of a sine wave and an exponential function. However, a limitation of this dynamic glottal model was that its control parameters could not be well interpreted, even though the glottis could be simulated. In the present study, more high-speed images were sampled, image processing was substantially improved, and the dynamic glottal control function was modeled using parameters that are significant for speech perception.

Subject Keywords: High-speed imaging, Vibration of vocal folds, Dynamic glottal model

1. INTRODUCTION

Speech source is critically important as it constitutes one of the three fundamental components of speech production, alongside vocal tract resonance and lip radiation. From the perspective of speech signals, at least three types of signals can be used to model the dynamic glottis: sound pressure, airflow, and glottal area function. While glottal excitation—namely, glottal flow and sound pressure —has been studied and modeled acoustically by many researchers, glottal activities have rarely been investigated and modeled because vocal fold vibration is difficult to observe and sample. In the past decade, however, increasingly more high-quality high-speed images have been captured through high-speed video cameras. This study focuses on improving the glottal control function of the dynamic glottal model and discusses its applications.

The classical models of speech source include the one-mass model and two-mass model (Flanagan at al, 1968; Flanagan, 1969; Lucero, 1993 and 1996; Pelorson et al, 1994), which have been theoretically established. Acoustical models based on speech sounds have also been extensively studied, with seven models considered particularly important: (1) the acoustical model developed by Rosenberg (1971), (2) the acoustical model of Hedelin (1984), (3) the acoustical model of Fant (1979), (4) the acoustical model of Fant (1982b), (5) the acoustical model of Ananthapadmanabha (1984), (6) the acoustical model of Fant et al (1985a), and (7) the acoustical model of Ljungqvist et al (1985a). Brief definitions of

these acoustical models are shown in Figures 1 [Figure 1: see original paper] through 3.

Figure 1 shows three models developed for the glottal flow pulse: the Rosenberg model (proposed in 1971), the Hedelin model (developed in 1984), and the Fant model (established in 1979). Figure 2 [Figure 2: see original paper] displays models developed for the differential form of glottal flow—that is, sound pressure —including the Fant model (established in 1982) and the Ananthapadmanabha model (established in 1984). Figure 3 [Figure 3: see original paper] shows models established through sound pressure, including the Fant model (established in 1985) and the Ljungqvist et al model (established in 1985).

Comparing these models reveals that the three models in Figure 1 are based on glottal flow, while those in Figures 2 and 3 are based on sound pressure. Scholars initially established their models using glottal flow pulses with few parameters, but subsequently found these models insufficiently effective and flexible for simulating different phonation types. Researchers then tended to develop models based on sound pressure with relatively more parameters, which proved more flexible and effective. Among these models, the LF-model established by Fant et al (1985a) is the most effective and flexible.

Fant' s earlier model employed three parameters and allowed for changes in open quotient (Fant, 1979a, 1979b, 1980). The LF-model reported by Fant et al in 1985 was a composition of the L-model (Liljencrants) and F-model (Fant) (Fant et al. 1985a). Figure 4 [Figure 4: see original paper] shows the basic definition of the LF-model, which has two phases: the first phase from 0 to 'Te' is created by multiplying sine and exponential functions, and the second phase from 'Te' to 'Tc' is created by an exponential function. The formulas in Figure 4 show the relationships among the parameters. With this model, a glottal pulse of sound pressure can be specified by four parameters: 'Tp, Te, Ta, and Ee' . Additionally, the parameter 'Tc' equals 'T0' , which equals 1/F0. To date, the LF-model remains one of the best models for simulating different glottal sources. The purpose of introducing the LF-model here is to provide a reference for comparative study with the physiological glottal model presented in this paper.

Various methods exist for studying speech source, including acoustical and physiological approaches. From a signal perspective, speech source can be investigated through acoustical signals, electroglottography (EGG) signals, air-pressure signals, high-speed image signals of vocal folds, and so on. Among these signals, high-speed digital imaging of vocal vibration most directly and accurately reflects the nature of vocal vibration and explains the relationship between vocal fold movement and speech sound characteristics. The earliest samples of vocal vibration were captured by Bell Telephone Laboratories in the 1930s, and since then, high-speed motion pictures have been used to study vocal fold vibration. Current systems for studying vocal fold vibration include the high-speed digital imaging system at the University of Tokyo, the kymography system by Kay, the Weinberger Speedcam system, the Kodak Ektapro

system, and others. The development of high-speed digital imaging systems has provided a solid foundation for studying vocal fold vibration.

2. GLOTTAL DETECTING

The methodology and procedures of this research include sampling high-speed digital images, image rotation and cropping, image motion compensation, image contrast adjustment, and parameter extraction.

2.1. Sampling of High-Speed Images

In this study, samples of vocal fold vibration were captured simultaneously with EGG and speech sound signals using a high-speed imaging system produced by KAY with an endoscope. The sampling rates were 2000 Hz and 4500 Hz. The samples included different phonation types, sustained vowels with low, middle, and high pitches, vowel samples with gliding voice from low to high and high to low, and samples of the four basic tones in Mandarin.

2.2. Image Rotation and Cropping

The samples were pre-processed through rotation and cropping. Images could be rotated automatically or manually depending on the specific samples, as it was sometimes very difficult to rotate samples into the correct position, especially for disordered glottis. The samples captured by the Kay system were 128×256 pixels in size and could be cropped to desired dimensions; for example, the image on the left in Figure 5 [Figure 5: see original paper] was cropped to 100×100 pixels as shown on the right.

Figure 5 illustrates the process: the first image is original, the second has been rotated using a bi-cubic algorithm, the third is a windowed image, and the fourth has been cropped to 100×100 pixels.

2.3. Image Motion Compensation (MC)

One adverse factor affecting the accuracy and validity of high-speed video (HSV) quantitative assessment is the motion of the endoscope' s lens relative to the larynx. Endoscopic motion makes it difficult to track the dynamic characteristics of laryngeal anatomical structures when dividing the glottis into left and right parts, as shown in Figures 6 [Figure 6: see original paper] and 7 [Figure 7: see original paper]. Therefore, we must perform "motion compensation" (MC), which involves detecting and removing endoscopic motion from HSV images. We primarily employed the approach and method published by Dimitar D. Deliyski (2005). Figure 7 shows the results after MC, and comparing the parameters in Figures 6 and 7 demonstrates that the results after MC are very good.

2.4. Image Contrast Adjusting

Another adverse factor affecting HSV quantitative assessment is contrast adjustment of HSV images when binarizing them to obtain the glottal shape, which depends on different operators. This influences quantitative estimation of glottal area and other parameters. Therefore, we automatically adjust the contrast of HSV images using the method shown in Figure 8 [Figure 8: see original paper]. From top to bottom: (1) the accumulated histogram when the glottis is fully opened in the first 100 frames of the video; (2) the accumulated histogram when the glottis is fully closed in the first 100 frames; (3) their subtraction yields the histogram of the glottal area, which shows a peak in the low gray region reflecting the gray value of the glottal region; (4) we smooth this histogram and use the gray values at the left and right sides of the first peak to automatically adjust the contrast of all frames in this HSV to obtain binarized images showing the glottal shape.

Figure 9 [Figure 9: see original paper] shows the results of contrast adjustment: the left image is an original image of modal female voice, the middle image shows the window used to limit the glottal area, and the right image shows the glottis automatically detected by the system.

2.5. Parameter Extraction

After the glottis has been detected and its area obtained, definitions must be established to extract glottal parameters. Figure 10 [Figure 10: see original paper] provides two graphs to help describe the glottis and establish definitions. The left graph shows a glottis where 'ABCD' represents the whole glottis, 'AB' represents the anterior glottis, 'CD' represents the posterior glottis, 'AC' represents the left glottis, and 'BD' represents the right glottis. Additionally, 'o' is the center of the glottis, 'lo' is the left width, 'ro' is the right width, 'ao' is the anterior length, and 'po' is the posterior length. The right graph illustrates one period of the glottal area function, where 'a' represents glottal opening instant, 'b' represents the local maximum of glottal area, 'c' represents glottal closing instant, and 'd' represents the next glottal opening instant.

The definitions of F0, OQ, and SQ are as follows: (1) fundamental frequency is defined as 1/ 'ad' (Hz); (2) open quotient is defined as the ratio of 'ac' over 'ad' ; (3) speed quotient is defined as the ratio of 'ab' over 'bc' . These definitions are shown on the right side of Figure 10.

Figure 11 [Figure 11: see original paper] shows the 13 parameters extracted by our system. From top to bottom: dynamic glottal area with glottal opening instant, glottal closing instant, and local maximum; left and right glottal areas; anterior and posterior areas; left and right widths; anterior and posterior lengths; and ratio of length over width. In this study, these parameters were used for dynamic glottal modeling.

3. MODELING ON DYNAMIC GLOTTIS

In the dynamic glottal model (Kong, 2007), the static glottis was modeled by four quarters of ellipses in three modes—normal mode, leakage mode, and open mode—while the dynamic glottal control function was approximated by the product of a parabola and sinusoid. Although this approach was effective, the control parameters could not be well interpreted despite successfully simulating the glottis. In this study, the static glottis was again modeled by four quarters of ellipses, with improvements focused on the dynamic glottal control function.

3.1. Model of Static Glottis

In this study, the static glottis was modeled by four quarters of ellipses, and the normal mode of the static glottal model was used to explain the new dynamic glottal control function, as shown in Figure 12 [Figure 12: see original paper]. The static glottis is modeled by four quarter-ellipses: 'left-posterior' (lp) ellipse, 'right-posterior ellipse' (rp), 'right-anterior ellipse' (ra), and 'left-anterior ellipse' (la). These four quarter-ellipses are calculated using two elliptical semi-major axes and two elliptical semi-minor axes, respectively.

3.2. Glottal Properties of Dynamic Glottis

The left and right dynamic widths and the anterior and posterior dynamic lengths in one glottal period are regarded as dynamic glottal control functions. In the dynamic model, they represent the contours of the two elliptical semi-major axes and two elliptical semi-minor axes in one period, which drive the dynamic glottal model and synthesize a glottal pulse. Based on parameters extracted from high-speed images of different phonations, the dynamic glottal control function can be classified into six basic types that approximate different portions of a sinusoid, as shown in Figure 13 [Figure 13: see original paper].

Figure 13 displays six images: image 'a' shows two dynamic glottal control functions approximating two sinusoids; image 'b' shows two functions approximating a portion of a sinusoid from 300° of the first sinusoid to 240° of the second; image 'c' shows two functions approximating a portion from 0° to 180°; image 'd' shows two functions with different local maxima, leading to different SQs; image 'e' shows two functions approximating a portion from 270° of the first sinusoid to 180° of the second; and image 'f' shows two functions with different lengths, leading to different OQs.

3.3. Modeling on Dynamic Glottis in Open Phase

To model the dynamic glottal control function more precisely, four function portions were selected from two periods of a sinusoid, as shown in Figure 14 [Figure 14: see original paper]. The angles of these two sinusoids range from 0 to 720 degrees. The selected sinusoid portion in this study was defined by the angle of the first sinusoid period and the angle of the second sinusoid period, respectively.

In these two sinusoid periods, only the portion between 270° of the first period and 270° of the second period was used for modeling the dynamic control function. Based on the properties of the real glottis, four typical sinusoid portions were chosen to model the dynamic glottal control function.

Figure 15 [Figure 15: see original paper] shows a sinusoid portion from 270° to 630° of the two periods, with values between -1 and 1. These values are normalized from 0 to 1 for the y-axis and 100 points for the x-axis in the right image, which will be used as part of the dynamic glottal control function. This is called 'type 1'.

Figure 16 [Figure 16: see original paper] displays a sinusoid portion from 360° to 540° of the two periods, with values from 0 to 1. The values are normalized from 0 to 1 for the y-axis and 100 points for the x-axis in the right image, to be used as part of the dynamic glottal control function. This is called 'type 2'.

Figure 17 [Figure 17: see original paper] shows a sinusoid portion from 270° to 540° of the two periods, with values between -1 and 1. The right image shows the normalized period with values from 0 to 1 for the y-axis and 100 points for the x-axis, to be used as part of the dynamic glottal control function. This is called 'type 3'.

Figure 18 [Figure 18: see original paper] displays a sinusoid portion from 360° to 630° of the two periods, with values between -1 and 1. The right image shows the normalized period with values from 0 to 1 for the y-axis and 100 points for the x-axis, to be used as part of the dynamic glottal control function. This is called 'type 4'.

3.4. Modeling on Dynamic Glottal Control Function

Based on the four types described above, parameters for F0, OQ, and SQ were added to produce a complete dynamic glottal control function. For clarity, the angles selected from the two sinusoid periods will be set and explained separately, as shown in Figure 19 [Figure 19: see original paper], which contains four plots displaying the four typical types of dynamic glottal function: plot 'a' shows type 1, plot 'b' shows type 2, plot 'c' shows type 3, and plot 'd' shows type 4.

In Figure 19, plot 'a' displays the type 1 dynamic glottal control function with F0 = 100 Hz (sampling rate 10k), OQ = 50%, SQ = 100%, and the open-phase pulse selected from 270° of the first sinusoid period to 270° of the second. Plot 'b' shows the type 2 function with F0 = 100 Hz, OQ = 50%, SQ = 100%, and the open-phase pulse selected from 360° of the first period to 180° of the second. Plot 'c' shows the type 3 function with F0 = 100 Hz, OQ = 50%, SQ = 100%, and the open-phase pulse selected from 270° of the first period to 180° of the second. Plot 'd' shows the type 4 function with F0 = 100 Hz, OQ = 50%, SQ = 100%, and the open-phase pulse selected from 360° of the first period to 270° of the second.

4. MODAL VOICE SYNTHESIS

Based on the static glottal model and the four types of glottal control functions, different phonation type pulses can be synthesized using parameters for F0, OQ, SQ, and glottal dimensions.

4.1. Modal Voice Synthesis

Modal voice is the most commonly used phonation type in spoken language, typically with F0 around 70-300 Hz, SQ around 100-300%, and OQ around 50-60%. The synthesis parameters are listed in Table 1 .

Table 1. Synthesis Parameters for Modal Voice

Parameter	Value
F0	100 Hz
OQ	50%
SQ	300%
Angle 1	360°
Angle 2	180°
Left width	1.3 mm
Right width	1.0 mm
Anterior length	8 mm
Posterior length	4 mm

For synthesizing modal voice, the parameters were set as follows: F0 = 100 Hz, OQ = 50%, SQ = 300%, angle 1 = 360°, and angle 2 = 180°. The left width is 1.3 mm, right width 1.0 mm, anterior length 8 mm, and posterior length 4 mm. The dynamic parameters and synthesized dynamic glottal area are displayed in plot 'a' of Figure 20 [Figure 20: see original paper].

Figure 20 contains four plots: plot 'a' shows left width, right width, anterior length, posterior length, and glottal area from top to bottom; plot 'b' shows dynamic glottises overlapped together; plot 'c' shows the spectrum of the glottal area function at -14.8929 dB/oct; and plot 'd' shows the spectrum of the glottal area function in differential form at -11.074 dB/oct.

4.2. Modal Voice Synthesis with Different SQ

In human vocal fold vibration, the left and right vocal folds often do not abduct and adduct simultaneously within one vocal period. The basic parameters for synthesizing such a voice are listed in Table 2 .

Table 2. Basic Parameters for Modal Voice with Different SQs

Parameter	Value
F0	100 Hz
OQ	50%
SQ (left width & anterior length)	300%
SQ (right width & posterior length)	75%
Angle 1	360°
Angle 2	180°
Left width	1.5 mm
Right width	1.5 mm
Anterior length	8 mm
Posterior length	4 mm

The SQs for the dynamic glottal control functions of left width and anterior length are 300%, while those for right width and posterior length are 75%. The angles for selecting sinusoid portions are 360° and 180°. The widths of left and right glottis are both 1.5 mm, with anterior length 8 mm and posterior length 4 mm. The synthesized parameters and spectra are shown in Figure 21 [Figure 21: see original paper].

Figure 21 contains four plots: plot 'a' displays left width, right width, anterior length, posterior length, and glottal area from top to bottom; plot 'b' shows dynamic glottises overlapped; plot 'c' shows the glottal area function spectrum at -14.6635 dB/oct; and plot 'd' shows the differential glottal area function spectrum at -10.912 dB/oct. These parameters and synthesized glottal areas demonstrate that the acoustical models discussed above cannot synthesize such glottal pulses, which are more flexible than those produced by acoustical models.

4.3. Falsetto Synthesis

Falsetto is a high-pitched voice type not commonly used in normal speech but frequently employed in oral performance such as singing and opera. In Chinese oral cultures like Kunqu and Peking opera, falsetto is often used by Dan (young female) performers. The parameters for synthesizing falsetto are listed in Table 3 .

Table 3. Falsetto Synthesis Parameters

Parameter	Value
F0	400 Hz
OQ	100%
SQ	100%
Angle 1	315°
Angle 2	225°
Left width	0.7 mm
Right width	0.7 mm

Parameter	Value
Anterior length	7 mm
Posterior length	5 mm

The F0 of 400 Hz is very high for a male speaker. The OQ of 100% is the maximum for vocal fold vibration, and the SQ of 100% is very small, indicating low power at high frequencies. The sinusoid portion was selected from 315° of the first period to 225° of the second period. The left and right glottal widths are 0.7 mm, with anterior length 7 mm and posterior length 5 mm, indicating a very narrow and long glottis. The synthesized dynamic control function, glottal area function, dynamic glottis, and spectra are shown in Figure 22 [Figure 22: see original paper].

In Figure 22, the synthesized glottal area function closely resembles a sinusoid, and the glottis appears narrow and long. The glottal area function spectrum is -23.5172 dB/oct, indicating small high-frequency power, while the differential glottal area function spectrum is -17.6235 dB/oct, showing that power remains small in human speech production.

4.4. Vocal Fry Synthesis

Vocal fry is a phonation type with very low pitch and high-frequency power, sometimes appearing in the middle of the low tone (tone 3) in Mandarin. This phonation closely resembles creaky voice, which typically has irregular periods. The basic parameters for synthesizing vocal fry are listed in Table 4 .

Table 4. Vocal Fry Synthesis Parameters

Parameter	Value
F0	40 Hz
OQ	15%
SQ (left width, anterior & posterior length)	300%
SQ (right width)	100%
Angle 1	360°
Angle 2	180°
Left width	1.1 mm
Right width	1.1 mm
Anterior length	1.3 mm
Posterior length	1.2 mm

The F0 of 40 Hz is very low for both male and female speakers. The OQ of 15% is very small, while SQ is 300% for left width, anterior length, and posterior length, and 100% for right width. Angle 1 is 360° of the first sinusoid period and

angle 2 is 180° of the second period. The synthesized parameters and spectra are shown in Figure 23 [Figure 23: see original paper].

In Figure 23, the synthesized glottal area function resembles a sawtooth waveform, and the glottis appears round and small. The glottal area function spectrum is -18.2591 dB/oct, indicating moderate high-frequency power, while the differential glottal area function spectrum is -8.7638 dB/oct, showing very high power in human speech production.

4.5. Synthesis of Diplophonia

Diplophonia is a voice type with different fundamental frequencies for left and right vocal folds. While not a normal voice in human speech, it often appears in disordered voices and is sometimes used in singing performance. The basic parameters for synthesizing diplophonia are listed in Table 5 .

Table 5. Basic Parameters for Diplophonia Synthesis

Parameter	Left/Anterior	Right/Posterior
F0	200 Hz	180 Hz
OQ	100%	100%
SQ	100%	100%
Angle 1	360°	360°
Angle 2	180°	180°
Width	1.5 mm	1.5 mm
Length	6 mm (anterior)	4 mm (posterior)

The F0 of the left glottis is 200 Hz while the right glottis is 180 Hz—a 20 Hz difference. The anterior glottis F0 is 200 Hz and the posterior glottis F0 is 180 Hz, also with a 20 Hz difference. The OQ is 100% (very large) and SQ is 100% (very small). The synthesized parameters and spectra are shown in Figure 24 [Figure 24: see original paper].

In Figure 24, plot 'a' displays the dynamic glottal control functions, dynamic glottis, and spectra of glottal area functions, showing different F0s for left and right glottal widths and a super-period covering approximately 10 vocal periods. Plot 'b' shows that the glottis vibrates asymmetrically. The glottal area function spectrum is -16.0798 dB/oct and the differential glottal area function spectrum is -10.8986 dB/oct, indicating substantial high-frequency power despite small SQs. This sample cannot be synthesized by any acoustical models, demonstrating that the glottal model operates at a deeper level that is more effective and flexible for source production, while acoustical models remain at the surface level.

5. CONCLUDING REMARKS

As is well known, voice models can be studied from perspectives including speech science, phonetics, and speech engineering, and can be established for various purposes: acoustical models for speech synthesis, physiological models for medical applications, phonetic models for studying the linguistic significance of phonation types, and singing models for synthesizing special songs or teaching singing. This research improved dynamic glottal control functions based on the physiological model first established by Kong (2001) and published in 2007. The improvements include three categories of basic parameters: (1) F0, OQ, and SQ —the most common parameters in speech science and phonetics research, whose properties are closely related to human perception and linguistic significance; (2) four types of dynamic glottal control functions selected from portions of a sinusoid defined by two angles; and (3) glottal size parameters including left and right glottal widths and anterior and posterior glottal lengths. With this improved physiological model, various phonation types can be simulated and further studied for multiple purposes across many fields.

Although the model can now easily synthesize many different phonation types, several aspects warrant further research and improvement. First, because the sampling rate of high-speed imaging systems is not high enough, parameters extracted from high-pitched voice samples are not very accurate, particularly for SQ and F0, as the peak position in one period of the glottal area function is unstable. Second, motion compensation and automatic contrast adjustment in the image processing system can be further improved. Third, due to sampling rate limitations, studying the spectrum of the glottal area function and its relationship with other signals remains difficult. We believe that with continued development of high-speed imaging systems, good samples with higher sampling rates—perhaps even 3D samples—can be captured for modeling a 3D dynamic glottal model to more accurately simulate vocal fold vibration and synthesize different phonations.

NOTES

1. This research is funded by the National Natural Sciences Foundation of China (No: 61073085). We would also like to thank Prof. Edwin Yiu at the University of Hong Kong for high-speed image sampling, all the subjects, and Wang Gaowu for program improvement of the high-speed image system.

REFERENCES

KONG Jiangping, 2007. *Laryngeal Dynamics and Physiological Models*, Peking University Press.

FLANAGAN J.L. and Landgraf L.L. (1968). Self-oscillating source for vocal-tract synthesizers. *IEEE Trans.* 16, March 1968, 57-64.

LUCERO J. C. 1993. Dynamics of the two-mass model of the vocal folds: equilibria, bifurcations, and oscillation region. *J. Acoust. Soc. Am.* 94(6), December.

LUCERO J. C. 1996. Chest- and falsetto-like oscillations in a two-mass model of the vocal folds. *J. Acoust. Soc. Am.* 100 (5), November.

PELORSON X., Hirochberg A., Hassel van R.R., Wijnands A.P.J., Auregan Y. 1994. Theoretical and experimental study of quasi-steady flow separation within the glottis during phonation. Application to a modified two-mass model. *J. Acoust. Soc. Am.* 96(6), December.

ROSENBERG A.E. (1971). Effect of glottal pulse shape on the quality of natural vowels. *Journal of the Acoustical Society of America*, 49, 583-98.

HEDELIN P. (1984). A glottal LPC-vocoder. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1.6.1-1.6.4. San Diego.

ANANTHAPADMANABHA T.V. (1984). Acoustic analysis of voice source dynamics. *Speech Transmission Laboratory –Quarterly Progress and Status Report*, 2-3, 1-24. Royal Institute of Technology, Stockholm.

LJUNGQVIST M. and Fujisaki H. (1985). A comparative study of glottal waveform models. *Technical Report of the Institute of Electronics and Communications Engineers*, Japan, EA85-58, 23-9.

FANT G. (1979a). Glottal source and excitation analysis. *STL-QPSR*, No. 1, pp. 85-107.
.1979b. Voice source analysis –a progress report. *STL-QPSR*, Nos. 3-7, pp. 31-54.
.1980. Voice source dynamics. *STL-QPSR*, Nos. 2-3, pp. 17-37.
.1982b. The voice source, acoustic modeling. *STL-QPSR*, No. 4, pp. 28-48.
.Liljencrants J. and Lin Q. (1985a). A four parameter model of glottal flow. *STL-QPSR*, No. 4, 1985, pp. 1-13.
.and Lin Q. (1988). Frequency domain interpretation and derivation of glottal flow parameters. *STL-QPSR*, Nos. 2-3, pp. 1-21.

DIMITAR D. Deliyski, (2005). "Endoscope Motion Compensation for Laryngeal High-Speed Video endoscopy" , *Journal of Voice*, Vol. 19, No. 3, pp. 485-496.

KONG Jiangping, (2001), "Study on Dynamic Glottis: through High-Speed Digital Imaging" , (in English), Ph.D. dissertation, City University of Hong Kong, Hong Kong, China.

Note: Figure translations are in progress. See original paper for figures.