Publication IV

Matti Airas and Paavo Alku, "Comparison of Multiple Voice Source Parameters in Different Phonation Types." In *Proceedings of the 8th Annual Conference of the International Speech Communication Association (Interspeech 2007)*, pp. 1410–1413, Antwerpen, Belgium, August 27–31, 2007.

Comparison of Multiple Voice Source Parameters in Different Phonation Types

Matti Airas, Paavo Alku

Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Finland
matti.airas@tkk.fi, paavo.alku@tkk.fi

Abstract

A large sample of vowels produced by male and female speakers were inverse filtered and parameterized using 21 different glottal flow parameters. The performance of the different parameters in expression of the phonation type was then tested using objective statistical methods. The comparison of the results revealed marked differences in the parameters' performance, and therefore, guidelines for parameter use and comparison were established.

Index Terms: voice quality, phonation type, inverse filtering, voice source, parameterization

1. Introduction

Voice quality is commonly considered to stem from laryngeal features of the voice production mechanism. The major physiological source of these changes is represented by the airflow generated by the vibrating vocal folds, the *glottal flow*. Unfortunately, direct measurement of this main source of voice quality is not possible from continuous speech due to the hidden position of the vocal folds. Hence, the only feasible means to estimate the glottal flow from speech is to use a technique called *inverse filtering*. This implies that resonances of the vocal tract are cancelled from the speech pressure signal by feeding it through anti-resonances which have been defined from the underlying speech spectrum [1].

Glottal flow estimates may be treated quantitatively by parameterizing them, and thus expressing the most important features of the original flow waveforms in a compressed numerical form. Numerous parameters concentrating on different aspects of the inverse filtered signals have been suggested. The basic parameterization approaches include time-based, frequency-based, and model fitting methods. In time-based methods, significant events such as the opening and closing instants of the glottal pulses or the maximum and minimum flow are acquired and used to compute the parameters. With special equipment such as the Rothenberg mask it is also possible to acquire absolute flow values and parameterize them [2]. In frequency-based methods, the properties of the flow magnitude spectrum such as the level difference of the harmonics or the slope of the harmonic series are used to gather the parameter values. In model-based methods, some mathematical formula representing an artificial glottal flow pulses are fitted to the flow estimate and used to represent the flow properties.

Time-based parameterization is the oldest utilized method, and still in very common use today. The open quotient (OQ) measures the relative portion of the open phase compared to the cycle length. The speed quotient (SQ) measures the ratio of the length of the opening phase to the length of the closing phase [3]. Since the opening phase of the glottal flow is commonly gradual, different automatic methods for reliable acquisition of the opening instant have been devised [4]. Variations of the OQ include the quasi open quotient (QOQ), which measures the relative amount of time during which the pulse amplitude is above a certain limit, such as 50% of the maximum level [5]. The closing quotient (ClQ) measures the ratio of the length of the closing phase to the period length [6]. Several parameters utilizing only amplitude data have been proposed as well. The amplitude quotient (AQ) and the normalized amplitude quotient (NAQ) both relate to the properties of the closing phase of the glottal pulse [7, 8]. OQ_a is a variant of OQ derived from the LF model using amplitude quantities only [9].

Numeric characterization of the spectral features of the glottal flow is widespread as well. Commonly used frequency-domain parameters include H1-H2, or ΔH_{12}, which denotes the difference of the first two harmonics [10], and the harmonic richness factor (HRF), which is the ratio between the sum of the amplitudes of the harmonics above the fundamental frequency (F0) and the amplitude of the fundamental [11]. Parabolic spectral parameter (PSP) is another proposed frequency-domain quantity, which is claimed to be less affected by changes in F0 [12].

One of the most widely used parameterization methods is the LF model fitting, in which the inverse filtered glottal flow is matched with a four-parameter synthetic glottal pulse [13]. There are at least two commonly used interchangeable sets of the parameters, which both are referred to just as "LF-parameters". The set presented in the original paper comprises t_e, t_p, t_a, and E_e, while the set often used in later papers is R_a, R_g, R_k, and OQ (which, for clarity, will be referred to as OQ_{lf}) [14]. Furthermore, a basic shape parameter R_d, which is largely identical to NAQ, is also commonly used.

A discussion of different glottal flow parameterization methods has been presented in [15]. There are several extensive studies on voice production, in which numerous time-domain parameters have been used, for example [16, 17]. However, there is a lack of studies in which the parameters are compared across parameterization types, e.g. in which time-based, LF-model and frequency domain parameters have all be used together in a comparable fashion. The current research addresses this issue and examines the effects of the phonation type (level of pressedness, i.e., hypo- vs. hyperfunction) on a large array of parameters. The aim of the research is to assess the interdependence and robustness of the different parameters and to form guidelines on their use in measurement of pressedness of speech. Reference implementations of the tested parameters are also provided as part of the HUT Aparat software package [18].

2. Materials and Methods

Vowel utterances of healthy, native Finnish speakers were recorded in an anechoic chamber. There were 11 speakers in total, of which 6 were women. The ages of the subjects ranged from 18 to 48 years, mean being 30 years.

The speakers were equipped with a headset microphone consisting of a unidirectional Sennheiser electret capsule. The microphone signal was routed through a microphone preamplifier and a mixer to iRiver iHP-140 digital audio recorder. Low-frequency phase distortion introduced by the digital recorder was corrected by acquiring the input impulse response of the device using an MLS measurement [19] and convolving the recorded signals using a time-reversed version of the impulse response.

The speech task consisted of uttering all eight Finnish vowels, /ɑ e i o u y æ ø/, in a randomized order in breathy, normal, and pressed phonation. The vowel strings were repeated three times. The tasks were explained to the subjects and they were asked to train the different phonation types until they were confident with them. During the recordings, the expressions were supervised, and, whenever necessary, the subjects were asked to retry the tasks with a stronger emphasis on the phonation type differences. The total size of the material was $11 \cdot 3 \cdot 3 \cdot 8 = 792$ vowels. The average duration of the utterances was 0.53 s.

The vowel recordings were segmented to separate one-vowel audio files. The files were then inverse filtered using HUT Aparat. The inverse filtering method used in this study was IAIF, the details of which are given in [20]. Parameterization was performed concurrently to the inverse filtering, with the exception of the LF-model fitting, which was performed as a batch run afterwards due to the intense computational requirements of the model optimization algorithm. The model fitting algorithm was adapted from [21]. For consistency, all LF-model parameters were computed using the fitted LF pulses. The acquired parameter set consisted of every parameter described in Section 1, or 21 parameters in total. Different algorithms for detection of the opening time instant in time-based OQ and SQ parameters are marked with subscripts [4].

Simple linear regression was performed for each parameter using the parameter itself as the dependent variable and the phonation type (breathy, normal, or pressed) as the independent variable. Then, the proportion of explained variation (R^2) was acquired from the linear model. These values were compared to measure the performance of the parameters in expression of phonation type differences. Furthermore, a Pearson cross correlation matrix of the different parameters was computed to assess the interdependence of the different parameters. All statistical treatments were performed using the R statistical software environment [22].

3. Results

In spite of the inherent difficulties in inverse filtering of vowels with a low first formant, the analyses conducted in the present study were mostly successful, with only two utterances yielding no acceptable glottal flow estimate. During the inverse filtering process, a subjective quality evaluation score on a scale of 0–3 was given for each glottal flow estimate using the general shape of the resulting glottal flow estimate as the criterion [23]. The mean value of the quality score was 2.5, hence indicating that the estimated flow waveforms could be considered reliable.

Linear regression models were computed for each of the parameters, using the phonation type as the sole independent variable. The R^2 values of the regression models are shown in Table 1. For the combined genders, NAQ was found to explain the phonation type best, with a R^2 proportion of 38.1%. AQ value was the second highest (34.3%). ClQ, ΔH_{12}, QOQ, and HRF yielded proportions over 25% as well. For the separate genders, AQ gave the highest proportion scores of 69.4% and 34.1% for males and females, respectively. Also NAQ yielded relatively high values of 54.7% and 26.6%. For males, a total of 10 other parameters attained R^2 values over 25%, while in females, besides AQ, only NAQ was able to surpass the limit of 25%.

	All	Males	Females
OQ_1	15.0	25.5	7.9
OQ_2	17.6	29.1	10.9
NAQ	38.1	54.7	26.6
AQ	34.3	69.4	34.1
ClQ	27.5	36.4	22.0
OQ_a	17.2	32.6	8.0
QOQ	26.2	32.5	22.6
SQ_1	12.1	10.4	15.2
SQ_2	3.1	1.4	4.9
ΔH_{12}	26.8	36.6	19.1
PSP	12.4	22.9	6.0
HRF	25.5	32.7	21.3
t_p	4.2	21.7	3.3
t_e	8.8	38.6	6.5
t_a	3.5	7.0	1.2
E_e	16.2	14.6	19.0
R_a	2.4	5.8	1.2
R_g	5.5	16.5	2.3
R_k	12.6	15.0	11.9
OQ_{lf}	14.5	25.0	8.0
R_d	22.2	35.8	15.1

Table 1: Proportion of variation explained by each parameters, grouped by the gender. The values are given in percents.

Correlation matrices of parameters for both genders combined and separated are given in Table 2. For both genders combined, the highest absolute correlation in the matrix is between HRF and ΔH_{12}, -0.88. Other very large correlation values (over 0.7) could be found between many of the OQ variants, NAQ and ClQ, NAQ and HRF, NAQ and ΔH_{12}, as well as ClQ and HRF. The only parameters failing to show a very large correlation with any other parameter were R_d, t_e, and AQ.

In males, the correlation of OQ_1 and OQ_{lf} was clearly highest, with a value of 0.94. Over half of the parameter pairs exhibited very high correlations (over 0.70). Only t_e fared considerably worse than others, having a very high correlation only with OQ_1.

In females, the negative correlation of HRF and ΔH_{12} was the highest, with a value of -0.91. Only a few parameter pairs exhibited very high correlations: NAQ with HRF, ΔH_{12}, ClQ, and AQ, AQ with HRF and ClQ, OQ_{lf} with t_e and OQ_1, and ΔH_{12} with HRF. Notably, most of the OQ variants and R_d did not exhibit very high correlation values with any other parameter.

4. Conclusions

Parameters focusing on the glottal closing phase (NAQ, AQ, ClQ, and to some degree R_d) were able to express the phonation type rather well. This is consistent with literature: closing phase constitutes the main excitation of the vocal tract, and ClQ and NAQ have been found to reflect phonation and vocal intensity changes [e.g. 6, 8, 24]. OQ and SQ have been found to reflect intensity changes as well [16, 17], but in the present study the two different SQ variants were able to reflect the phonation changes only weakly or not at all. The different OQ vari-

	OQ_1	OQ_2	NAQ	AQ	ClQ	OQ_a	QOQ	ΔH_{12}	HRF	t_e	OQ_{lf}
R_d	0.22	0.32	0.58	0.42	0.60	0.16	0.48	0.46	-0.41	0.11	0.25
OQ_{lf}	**0.87**	0.42	0.29	0.34	0.28	0.55	0.64	0.35	-0.32	0.56	
t_e	0.39	0.04	0.03	0.66	0.03	0.08	0.24	0.06	0.04		
HRF	-0.52	-0.61	**-0.81**	-0.47	**-0.74**	-0.60	-0.64	**-0.88**			
ΔH_{12}	0.50	0.56	**0.77**	0.53	0.69	0.51	0.66				
QOQ	**0.72**	**0.74**	0.63	0.46	0.66	0.68					
OQ_a	0.69	**0.74**	0.54	0.30	0.48						
ClQ	0.58	0.69	**0.80**	0.51							
AQ	0.36	0.33	0.66								
NAQ	0.48	0.58									
OQ_2	0.64										

	OQ_1	OQ_2	NAQ	AQ	ClQ	OQ_a	QOQ	ΔH_{12}	HRF	t_e	OQ_{lf}
R_d	0.50	0.57	**0.74**	0.67	**0.79**	0.52	**0.70**	**0.70**	-0.66	0.38	0.52
OQ_{lf}	**0.94**	0.59	0.60	0.53	0.56	0.66	**0.73**	0.56	-0.58	**0.81**	
t_e	**0.74**	0.36	0.45	0.59	0.35	0.47	0.47	0.36	-0.34		
HRF	-0.66	**-0.71**	**-0.84**	**-0.72**	**-0.79**	**-0.76**	**-0.78**	**-0.86**			
ΔH_{12}	0.63	0.67	**0.83**	**0.73**	**0.78**	0.66	**0.77**				
QOQ	**0.80**	**0.84**	**0.79**	0.65	**0.82**	**0.80**					
OQ_a	**0.73**	**0.82**	**0.77**	0.68	**0.70**						
ClQ	**0.71**	**0.80**	**0.84**	**0.73**							
AQ	0.59	0.61	**0.93**								
NAQ	0.68	**0.73**									
OQ_2	**0.71**										

	OQ_1	OQ_2	NAQ	AQ	ClQ	OQ_a	QOQ	ΔH_{12}	HRF	t_e	OQ_{lf}
R_d	-0.06	0.13	0.46	0.44	0.48	-0.15	0.29	0.26	-0.23	0.01	-0.01
OQ_{lf}	**0.75**	0.23	-0.09	-0.01	-0.04	0.46	0.47	0.05	-0.02	**0.74**	
t_e	0.54	0.13	-0.26	0.05	-0.08	0.22	0.34	-0.13	0.17		
HRF	-0.35	-0.48	**-0.77**	**-0.73**	-0.69	-0.39	-0.48	**-0.91**			
ΔH_{12}	0.29	0.42	0.69	0.66	0.61	0.31	0.49				
QOQ	0.57	0.61	0.42	0.48	0.47	0.52					
OQ_a	0.64	0.61	0.27	0.24	0.23						
ClQ	0.43	0.58	**0.76**	**0.72**							
AQ	0.25	0.42	**0.88**								
NAQ	0.22	0.41									
OQ_2	0.55										

Table 2: Pearson cross-correlation matrices for (top to bottom) both sexes combined, males, and females. For brevity, parameters having a R^2 value smaller than 25% in every column of Table 1 have been omitted. For the sake of clarity, values over 0.7 or under -0.7 have been emphasized.

ants, however, were able to moderately reflect the phonation changes, the amplitude threshold based QOQ being the best of them. This supports claims that precise extraction of the opening instant is often difficult due to the gradual opening of the vocal folds, and therefore parameters utilizing the opening instant may be less robust than those avoiding it [8]. Frequency domain parameters, with the exception of PSP, were able to reflect the phonation changes reasonably well, a result consistent with earlier studies [e.g. 10, 11]. Interestingly, PSP has been claimed to be more robust than OQ [12], but that result was not supported in the present study.

In general, the LF model parameters did not perform very strongly. Only R_d was able to express phonation changes in a relevant manner. The definition of R_d is identical to that of NAQ, except for a scaling factor of 0.11 derived from the approximate average fundamental frequency in a study of three Swedish male speakers [14]. Thus, any differences in the R^2 of NAQ and R_d have to be attributed to the properties of the fitting of the LF-model, in which data is inherently lost. This may, however, also be somewhat dependent on the model fitting method, although the authors remain confident that the model fitting algorithm used in this study is methodologically sound. Moreover, the fitting method dependence would be yet another problem by itself.

Analysis of the cross-correlation matrices indicated that the parameters tend to correlate well with other members of the same parameter group. For example, the different OQ variants are closely coupled to each other, but not that well to the closing phase parameters. NAQ appeared to be very well correlated with many other parameters, while most of the LF model parameters showed little correlation with each other or the other groups. The correlation of NAQ and AQ was nearly perfect within the separate genders, and high even in the combined genders. This was an expected results, as in the within-gender categories the F0 normalization acts mostly as a scaling factor, while in the combined case the normalization induces differences in the value depending on speaker sex. Respectively, the correlation between HRF and ΔH_{12} ranged from very high to nearly perfect in all cases, indicating that the higher spectrum harmonics have only a minor effect on the frequency domain

parameters.

Notable gender differences can be detected both in the proportion of explained variation and in the correlation matrices. In the proportion of variation explained, the values for males are regularly twice as large as for females. This, in authors' opinion, is most likely due to the higher F0 of women, which increases the difficulty of properly estimating the locations of low formants and thus induces undesirable variation in the inverse filtering process. The only notable exception are the SQ parameters, which show somewhat higher R^2 values for females. This result supports [17], in which a negative correlation was found with voice intensity and SQ, and that the correlation was far stronger in females.

During the analysis of the data, it was also noted that the use of a wide set of vowels added considerable noise in the parameter data. Informal testing with [ɑ] vowels indicated considerably better results than with the mixed set, despite much reduced amount of data.

According to the present study, NAQ and AQ were able to express phonation type changes best of the implemented parameters, so their use in the context of phonation type changes is encouraged by the authors. The choice between these two parameters should depend on the task at hand: if the subjects are of a single sex, then AQ may yield more consistent results, but if a mixed subject setup is used, then NAQ is strongly preferred. If frequency-domain parameters are preferred, then either HRF or ΔH_{12} should give acceptable results.

The present study gave the first comparative review of a large amount of different glottal flow parameters in the context of phonation type changes. Remarkable differences among the parameters were found. Therefore, the authors wish that this study could help other researchers in the field in the selection of suitable parameterization methods.

5. References

[1] R. L. Miller, "Nature of the vocal cord wave," *J Acoust Soc Am*, vol. 31, no. 6, pp. 667–677, June 1959.

[2] M. Rothenberg, "A new inverse-filtering technique for deriving the glottal air flow waveform during voicing," *J Acoust Soc Am*, vol. 53, no. 6, pp. 1632–1645, 1973.

[3] R. Timcke, H. von Leden, and P. Moore, "Laryngeal vibrations: measurements of the glottic wave," *Archiv Otolaryngol*, vol. 68, pp. 1–19, 1958.

[4] H. Pulakka, "Analysis of human voice production using inverse filtering, high-speed imaging, and electroglottography," Master's thesis, Helsinki University of Technology, Espoo, Finland, 2005.

[5] C. Dromey, E. T. Stathopoulos, and C. M. Sapienza, "Glottal airflow and electroglottographic measures of vocal function at multiple intensities," *J Voice*, vol. 6, no. 1, pp. 44–54, 1992.

[6] R. B. Monsen and A. M. Engebretson, "Study of variations in the male and female glottal wave," *J Acoust Soc Am*, vol. 62, no. 4, pp. 981–993, October 1977.

[7] P. Alku and E. Vilkman, "Amplitude domain quotient of the glottal volume velocity waveform estimated by inverse filtering," *Speech Commun*, vol. 18, pp. 131–138, 1996.

[8] P. Alku, T. Bäckström, and E. Vilkman, "Normalized amplitude quotient for parametrization of the glottal flow," *J Acoust Soc Am*, vol. 112, no. 2, pp. 701–710, August 2002.

[9] C. Gobl and A. Ní Chasaide, "Amplitude-based source parameters for measuring voice quality," in *Proc ISCA VOQUAL'03 Workshop on Voice Quality: Functions, Analysis and Synthesis*, Geneva, 2003, pp. 151–156.

[10] I. R. Titze and J. Sundberg, "Vocal intensity in speakers and singers," *J Acoust Soc Am*, vol. 91, no. 5, pp. 2936–2946, May 1992.

[11] D. G. Childers and C. K. Lee, "Vocal quality factors: Analysis, synthesis, and perception," *J Acoust Soc Am*, vol. 90, no. 5, pp. 2394–2410, November 1991.

[12] P. Alku, H. Strik, and E. Vilkman, "Parabolic spectral parameter – A new method for quantifiction of the glottal flow," *Speech Commun*, vol. 22, pp. 67–79, 1997.

[13] G. Fant, J. Liljencrants, and Q. Lin, "A four-parameter model of glottal flow," *STL-QPSR*, no. 4, pp. 1–13, 1985.

[14] G. Fant, "The LF-model revisited. transformations and frequency domain analysis," *STL-QPSR*, no. 2–3, pp. 119–156, 1995.

[15] P. Alku, "Parameterisation methods of the glottal flow estimated by inverse filtering," in *Proc ISCA Workshop on Voice Quality: Functions, analysis and synthesis (VOQUAL03)*, Geneva, Switzerland, August 2003, pp. 81–87.

[16] E. B. Holmberg, R. E. Hillman, and J. S. Perkell, "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice," *J Acoust Soc Am*, vol. 84, no. 2, pp. 511–1787, August 1988.

[17] A. M. Sulter and H. P. Wit, "Glottal volume velocity waveform characteristics in subjects with and without vocal training, related to gender, sound intensity, fundamental frequency, and age," *J Acoust Soc Am*, vol. 100, no. 5, pp. 3360–73, 1996.

[18] M. Airas, H. Pulakka, T. Bäckström, and P. Alku, "A toolkit for voice inverse filtering and parametrisation," in *9th Eur Conf Speech Comm and Tech*, Lisbon, Portugal, September 4-8 2005, pp. 2145–2148.

[19] D. D. Rife and J. Vanderkooy, "Transfer-function measurement with maximum-length sequences," *J Audio Eng Soc*, vol. 37, pp. 419–444, 1989.

[20] P. Alku, B. Story, and M. Airas, "Estimation of the voice source from speech pressure signals: Evaluation of an inverse filtering technique using physical modelling of voice production," *Folia Phoniatr Logo*, vol. 58, no. 2, pp. 102–113, 2006.

[21] H. Strik, B. Cranen, and L. Boves, "Fitting a LF-model to inverse filter signals," in *Proc 3rd Eur Conf Speech Comm Tech*, vol. 1, Berlin, Germany, 1993, pp. 103–106.

[22] R. Ihaka and R. Gentleman, "R: A language for data analysis and graphics," *J Computat Graphical Stat*, vol. 5, no. 3, pp. 299–314, 1996.

[23] L. Lehto, M. Airas, E. Björkner, J. Sundberg, and P. Alku, "Comparison of two inverse filtering methods in parameterization of the glottal closing phase characteristics in different phonation types," *J Voice*, 2006, accepted for publication.

[24] G. Fant, "Some problems in voice source analysis," *Speech Commun*, vol. 13, pp. 7–22, 1993.