COMPARING METHODS FOR QUANTIFYING THE VOICE SOURCE OF DIFFERENT PHONATION TYPES INVERSE FILTERED FROM ACOUSTIC SPEECH SIGNALS

Paavo Alku[1] and Erkki Vilkman[2]
1: Helsinki University of Technology, Acoustics Lab. Otakaari 5 A, FIN-02150 Espoo, Finland
2: Helsinki Univ. Central Hospital, Dept. Phoniatrics Haartmaninkatu 4, FIN-00290 Helsinki, Finland

ABSTRACT

This study compares quantification techniques that have been developed to parameterize the voice source obtained by inverse filtering. Quantification of the voice source of different phonation types was computed using altogether seven parameters. The results showed that phonation types could be separated from each other most effectively when quantification was based on parameters determined between the instant of the maximal glottal opening and the minimum peak of the flow derivative.

INTRODUCTION

Inverse filtering is widely applied in the analysis of voice production. Inverse filtering methods can be divided into two categories. The first category consists of techniques that are based on inverse filtering of the volume velocity signal that has been recorded at the mouth using a flow mask [1]. The resulting glottal volume velocity waveform can be calibrated in the amplitude domain and the DC-flow is also obtained. The second category of inverse filtering techniques is based on the estimation of the glottal source from the acoustic speech pressure wave that has been recorded in a free field (e.g.[2]). Glottal airflow waveforms estimated by these techniques are obtained on arbitrary amplitude scales with no indication of the DC-flow.

Glottal flows estimated by inverse filtering are usually quantified using certain parameters. Characterization of the voice source by time-based parameters that are extracted from the glottal volume velocity waveform has been widely used [3]. Time-domain quantification of the voice production using the derivative of the glottal airflow waveform has also been used [3, 4]. If flow mask is used in inverse filtering parametrization of the voice source can be done by measuring the absolute values of both the AC- and DC-flow. In the frequency domain the decay of the voice source spectrum can be parametrized using, for example, the harmonic richness factor (HRF) [5].

The aim of this research was to compare different methods that have been developed for quantification of the glottal airflow that has been inverse filtered without a flow mask. We were interested in exploring how changing the phonation type can be presented by different quantification techniques. By doing this comparison our purpose was to find the parameter that most clearly indicates changes in the phonation type.

MATERIAL AND METHODS

Time-domain quantification of the glottal airflow waveform was computed by using the following parameters [3]: open quotient (OQ), speed quotient (SQ), and closing quotient (CQ). Quantification of voice production using the derivative of the flow waveform was performed with the following time-domain parameters: return quotient (RQ) [6] and peak-to-peak quotient (PPQ) [4]. By referring to Fig. 1 these time-domain parameters can be defined as follows:

$$OQ = (t_{o1} + t_{o2}) / T$$
$$SQ = t_{o1} / t_{o2} \quad , \quad CQ = t_{o2} / T$$
$$RQ = t_{ret} / T \quad , \quad PPQ = t_{pp} / T$$

Amplitude domain quantification of the glottal source was not possible with absolute flow values because our approach was based on inverse filtering without a flow mask. However, the authors have recently presented a new amplitude-based quotient which can be used even though absolute flow values are not given be the recording apparatus [7]. This new parameter, amplitude quotient (AQ), is defined as the ratio of the AC-amplitude of the flow and the amplitude of the negative peak of the first derivative of the flow (Fig. 1):

$$AQ = A_{ac} / A_{min}$$

Frequency domain quantification of the voice source was computed using the harmonic richness factor [5]:

$$HRF = \frac{\sum_{i \geq 2} H_i}{H_1}$$

, where H_i denotes the amplitude of the *i*th harmonic computed from the spectrum of the glottal volume velocity waveform.

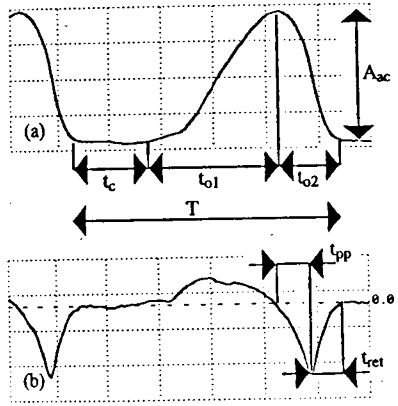

Fig. 1 (a): Glottal flow, t_c = closed phase, t_{o1} = opening phase, t_{o2} = closing phase, A_{ac} = amplitude of AC-flow, T = fundamental period
(b): Derivative of the glottal flow, A_{min} = amplitude of the negative peak, t_{ret} = return time, t_{pp} = time distance between the negative peak of the differentiated flow and the positive peak of the flow

The speech material consisted of vowels produced by five female and five male speakers. The speakers were asked to produce a sustained /a/-vowel using breathy, normal, and pressed phonation types. Recording of the signals was performed in an anechoic chamber using a condenser microphone (Brüel&Kjær 4133).

Estimation of the glottal airflow waveforms was performed with a new inverse filtering technique that is described in detail in [2]. All the estimated glottal airflow waveforms and their derivatives were analyzed using a computer cursor in order to mark time and amplitude values that were required for computation of parameters.

RESULTS

The obtained values for all the seven parameters are given in Tables 1 and 2 for female and male voices, respectively. Two relative changes were computed for each parameter by comparing normal phonation to breathy and pressed phonation to normal. In the case of OQ, for example, these relative changes were defined as follows:

$$\frac{OQ_{normal} - OQ_{breathy}}{OQ_{breathy}} 100 \%$$
$$\frac{OQ_{pressed} - OQ_{normal}}{OQ_{normal}} 100 \%$$

To compare deviation of the parameters between different speakers we expressed results using coefficient of variation (i.e. the ratio between the standard deviation and the mean). In the case of OQ, for example, coefficient of variation was defined as follows:

$$v = (sd_{OQ} / m_{OQ}) \, 100 \%$$

Female speakers:

Time-based parameters computed from glottal flows of female voices showed that the mean value of both OQ and CQ decreased while the mean of SQ increased when phonation was changed from breathy to pressed. Changing of the mean values of all these time-based parameters was monotonic when phonation was altered. The relative change was smallest for OQ. The mean value of SQ showed the largest relative change (38 %) among the three time-based quotients when breathy phonation was compared to normal. However, when phonation was further changed from normal to pressed the mean value of SQ increased only slightly (5 %). The mean value of CQ showed a clear descend when phonation was changed both from breathy to normal (-28 %) and from normal to pressed (-10 %). When these three time-based parameters were analyzed between different speakers it turned out that CQ was the only one whose value changed (decreased) monotonically for all the female subjects when phonation was changed.

RQ was the only parameter among all the analyzed quotients whose mean value did not change monotonically when phonation was altered from breathy to pressed. There were large variations between different speakers when the value of RQ was analyzed as a function of phonation. PPQ decreased monotonically for all the female subjects when phonation was changed from breathy to pressed.

AQ decreased monotonically for all the female subjects when phonation was changed. The mean value of AQ showed large relative changes (-33 % and -20 %) when different phonation types were compared. The frequency domain parameter, HRF, showed the largest relative changes when phonation was altered (135 % and 57 %). HRF-values increased for all the subjects monotonically.

Among the three time-based parameters extracted from the glottal flows the value of OQ showed the smallest deviation from one speaker to another. The value of ν averaged over the three phonation types (ν_{av}) equaled 8 % for OQ, 15 % for CQ, and 16 % for SQ. Deviation of parameter values between female speakers was largest for RQ (ν_{av} equaled 26 %). For PPQ deviation was smaller (ν_{av} equaled 20 %). The coefficient of variation for AQ gave a value that was the second smallest among all the analyzed parameters of female subjects (ν_{av} equaled 11 %). The value of HRF showed large deviations from one female voice to another (ν_{av} equaled 26 %).

Male speakers:

Mean values of time-based parameters extracted from the flow waveforms of male subjects changed monotonically (OQ and CQ decreased, and SQ increased) when phonation was altered from breathy to pressed. Relative changes between different phonation types were larger than in female voices. The relative change in the value of OQ was again smallest. SQ yielded the largest relative change for time-based parameters of male voices (87 % when phonation was changed from breathy to normal). However, the value of SQ increased only slightly (by 1 %) when phonation was further changed from normal to pressed. The value of CQ showed large changes both in breathy-to-normal (- 40 %) and normal-to-pressed (- 19 %) changes. CQ was the only time-based parameter that showed for all the male subjects a monotonic decrease when phonation was changed from breathy to pressed.

RQ showed also for male voices the largest deviation from one voice to another. The mean value of RQ did not change monotonically when phonation was altered. The mean value of PPQ decreased by 38 % and 35 % in breathy-to-normal and normal-to-pressed changes, respectively. For all the five subjects PPQ decreased monotonically when phonation was altered towards pressed.

The mean value of AQ decreased monotonically when the phonation type was changed. All the male subjects yielded the largest value of AQ in breathy phonation, the second largest in normal phonation and the smallest value in pressed phonation. The mean values of HRF increased also monotonically when phonation was changed towards pressed.

The time-based quotients extracted from the glottal flows showed deviations between subjects that were smallest in CQ (ν_{av} equaled 11 %), second smallest in OQ (ν_{av} equaled 13 %) and largest in SQ (ν_{av} equaled 18 %). The value of RQ showed for male voices the most substantial deviation from one speaker to another (ν_{av} was equal to 33 %). PPQ showed also quite large deviation (ν_{av} equaled 21 %). Variation of the value of AQ from one speaker to another was larger in male voices than in female speech (ν_{av} equaled 20 %). HRF showed also for male voices great deviation between different subjects (ν_{av} equaled 27 %).

SUMMARY

Comparison of the parameters was done, first, by using as a criterion the change of the mean parameter value when phonation was altered from breathy to pressed. With this criterion the parameters could be sorted according to the following order of superiority both for female and male voices: HRF, AQ, PPQ, CQ, SQ, OQ, and RQ. Second, we analyzed for how many speakers the changing of parameters was monotonic when phonation was altered. It was found that AQ was the only parameter whose value showed a clear monotonic change (decrease) for all the subjects when phonation was altered from breathy

to pressed. Hence, the parameters could be sorted using the following order for female voices: HRF, AQ, PPQ, CQ, SQ, OQ, and RQ. For male voices the order of superiority was as follows: AQ, PPQ, CQ, HRF, SQ, OQ, and RQ. Third, quantification methods were compared by using as a criterion deviation of the parameter values between different speakers. Using this criterion the following order of superiority was obtained for female voices: OQ, AQ, CQ, SQ, PPQ, HRF, and RQ. For male voices the order was slightly different: CQ, OQ, SQ, AQ, PPQ, HRF, and RQ.

We conclude that the phonation type can be characterized most effectively by using either frequency domain parameterization with HRF or time-domain parameterization that is based on values extracted during glottal closing phase, especially during the time that spans from the instant of maximal glottal opening to the instant of the negative peak of the flow derivative. If extraction is based on the flow waveform alone, then according to our experiments the best time-domain parameter is CQ. However, if time-domain parameterization is based on the flow derivative alone, then applying PPQ is recommended. According to the results of our experiments the most effective way to characterize the voice source in the time-domain is to apply both the flow and its derivative by using parameter AQ.

REFERENCES

[1] Rothenberg, M. (1973). "A new inverse-filtering technique for deriving the glottal air flow waveform during voicing", J. Acoust. Soc. Am., Vol. 53, pp. 1632-1645.
[2] Alku, P., Vilkman, E. (1994) "Estimation of the glottal pulseform based on discrete all-pole modeling", Proc. '94 Int. Conf. on Spoken Language Processing, pp. 1619-1622.
[3] Holmberg, E.B., Hillman, R.E., Perkell, J.S. (1988). "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice", J. Acoust. Soc. Am., Vol. 84, pp. 511-529.
[4] Sundberg, J., Titze, I., Scherer, R. (1993). "Phonatory control in male singing: A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source", J. Voice, Vol. 7, pp. 15-29.
[5] Childers, D.G., Lee, C.K. (1991). "Vocal quality factors: Analysis, synthesis, and perception", J. Acoust. Soc. Am, Vol. 90, pp. 2394-2410.
[6] Price, P.J. (1989). "Male and female voice source characteristics: Inverse filtering results",Speech Communication, Vol. 8, pp. 261-277.
[7] Alku, P., Vilkman, E. (1994). "Amplitude domain quotient for characterization of the inverse filtered glottal flow", In review.

Table 1. Means (m) and standard deviations (sd) for parameters of female subjects.

phonation		OQ	SQ	CQ	RQ	PPQ	AQ	HRF
breathy	m	0.94	1.38	0.40	0.18	0.22	9.05	0.20
	sd	0.07	0.24	0.06	0.03	0.05	1.55	0.08
normal	m	0.84	1.90	0.29	0.14	0.16	6.07	0.47
	sd	0.04	0.34	0.04	0.06	0.03	0.37	0.09
pressed	m	0.78	1.99	0.26	0.16	0.12	4.85	0.74
	sd	0.10	0.28	0.04	0.03	0.02	0.46	0.13

Table 2. Means (m) and standard deviations (sd) for parameters of male subjects.

phonation		OQ	SQ	CQ	RQ	PPQ	AQ	HRF
breathy	m	0.96	1.15	0.45	0.14	0.32	22.16	0.20
	sd	0.04	0.16	0.05	0.07	0.06	4.33	0.04
normal	m	0.84	2.15	0.27	0.09	0.20	10.39	0.56
	sd	0.08	0.45	0.02	0.03	0.04	2.65	0.11
pressed	m	0.70	2.18	0.22	0.12	0.13	7.08	0.91
	sd	0.17	0.44	0.03	0.02	0.03	0.94	0.36