ISCA Archive
http://www.isca-speech.org/archive

6th International Conference on Spoken
Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

ANALYSIS OF VOICE PRODUCTION IN BREATHY, NORMAL AND PRESSED PHONATION BY COMPARING INVERSE FILTERING AND VIDEOKYMOGRAPHY

Paavo Alku[1], *Jan G. Švec*[2], *Erkki Vilkman*[3], *Frantisek Šram*[2]
1: Helsinki University of Technology,
Acoustics Lab, Box 3000, Fin-02015 TKK, Finland
e-mail: Paavo.Alku@hut.fi
2: Centre for Communication Disorders, Medical Healthcom Ltd., Prague, the Czech Republic
3: University of Oulu, Oulu, Finland

ABSTRACT

The present study addresses comparison of two analysis methods of voice production, inverse filtering and videokymography (VKG). Speech data were collected from two male speakers using sustained phonation during laryngoscopy (sound corresponding approximately to the vowel /ä/). The type of phonation was varied between breathy, normal, and pressed. From the waveforms given by inverse filtering, the time length between the positive and negative peak of the glottal flow derivative and the length of the fundamental period was measured. The length of the open phase and the length of the fundamental period was measured from the VKG-images. The results suggest that the time length between the positive and negative peak of the flow derivative equals the length of the open phase of the VKG-image (i.e., the time difference between the opening and closing of the middle line of the vocal folds) for phonations with complete closure of the glottis.

1. INTRODUCTION

Different methodology has been developed during the past decades in order to analyze vibration of the vocal folds. In this area, one of the most widely used methods is inverse filtering [e.g., 1, 5, 10]. Inverse filtering is a technique that estimates the glottal volume velocity waveform by removing the filtering effects of the vocal tract. It is possible to estimate the glottal source either from the speech pressure waveform [e.g., 1] or from the oral flow recorded by a pneumotachometric mask (frequently called Rothenberg's mask) [5]. The output of inverse filtering, an estimate for the glottal waveform, can be related directly to vibration of the vocal folds. Therefore, the resulting estimate of the glottal flow is usually parameterized in order to get a quantitative representation for the function of voice production [3, 4, 7]. In addition to dealing with the flow, it is also possible to parameterize voice production by using the first derivative of the flow [3, 7].

A promising new method to get visual information of the vibrating vocal folds is to apply videokymography (VKG) [6, 8, 9]. VKG uses a modified video camera that is able to work in two modes: high-speed (approximately 8000 images per second) and standard (50 images per second). In the high-speed mode, the camera selects from the whole laryngeal image one active horizontal line that is transversal to the glottis. Consecutive line images create a VKG-image, which shows as a time-domain signal the vibratory pattern of the selected part of the vocal folds. Currently, VKG is commercially available and it is used as a low-cost high-speed imaging system in clinical practice.

To the best of authors' knowledge, there are no previous studies on the comparison of inverse filtering with videokymography. Therefore, the goal of the current study is to compare results yielded by inverse filtering to those obtained by videokymography. The study reports results that were obtained by the two techniques when analyzing function of the voice source in three different phonation types (breathy, normal and pressed).

2. MATERIALS AND METHODS

Sustained phonations produced by two male subjects (speakers EV and JS) during laryngoscopy (sound corresponding approximately to the vowel /ä/) were analyzed. Both subjects pronounced the vowel using three different phonation types (breathy, normal and pressed). The duration of each speech sample was at least 3 seconds. The same sound was repeated a sufficient number of times in order to get focused and bright VKG-images and to assure the VKG-measurement line to be placed in the middle of the vibrating part of the vocal folds, perpendicularly to glottal axis. The VKG recordings were taken using an equipment described in [9]. Simultaneously, recordings of the acoustical speech pressure waveforms were done using a condenser microphone (Brüel&Kjær 4176) and a DAT-recorder (Sony TCD-D3).

Speech pressure waveforms were inverse filtered and the obtained estimates of the glottal flow were compared to the visual information given by VKG. As an inverse filtering method we used a method similar to the one described in [1]. This inverse filtering method estimates the glottal volume velocity waveform directly from the acoustic speech pressure waveform without applying the so called Rothenberg mask [5]. The inverse filtering method used in the current study applies a sophisticated all-pole modeling technique, called DAP (Discrete All-pole Modeling) [2], in estimation of the vocal tract. In comparison to conventional linear prediction, DAP yields more accurate estimates for the vocal tract especially in the case of high-pitch voices.

Unfortunately, the recording equipment did not make possible achieving exact time-synchronization between the acoustical speech pressure waveforms and the corresponding VKG-images. Therefore, the occurrence of critical time instants (e.g., opening of the glottis, instant of the maximal flow etc.) could not be compared synchronously between the glottal flows and VKG-images. Instead, the comparison between the two information signals was done by extracting from the differentiated glottal flows the following two time spans (see Fig. 8): (1) the length of the fundamental period (T_{if}), and (2) the time difference between the positive and negative peak (t_{if}). From the VKG-images, the following two values were extracted: (see Fig. 5) (1) the length of the fundamental period (T_{VKG}) and (2) the length of the open phase of the VKG -image (t_{VKG}).

3. RESULTS

The obtained results of VKG-analysis are shown for speaker EV in Fig. 1, 2, and 3 in the case of breathy, normal and pressed phonation, respectively. Similarly, VKG-images of speaker JS are shown in Fig. 4, 5, and 6 for breathy, normal and pressed phonation, respectively. Glottal flows and their derivatives computed from voices of speaker EV are shown for the three phonation types in Fig. 7, 8, and 9. Finally, results obtained by inverse filtering are shown for speaker JS in Fig. 10, 11, and 12. The extracted values are given for all the six phonations in Table 1.

VKG-images obtained from breathy voices (Fig. 1 and 4) indicated clearly that the middle part of the vocal folds did not close during the glottal cycle. In other words, there was a continuous flow though the vocal folds during the entire glottal cycle according to the videokymographic analysis. This kind of vibration was also evidenced by the corresponding glottal flows obtained by inverse filtering (upper panels of Fig. 7 and 10), because these smooth waveforms show no evidence for a clear closed phase of the glottal cycle. It is, however, worth noticing that the glottal flows of the present study were computed from speech pressure waveforms without applying the flow mask, which implies that information on the DC-flow is not available. Hence, we cannot determine from the glottal flows shown in Fig. 7 and 10, whether the minimum flow reaches zero value or not.

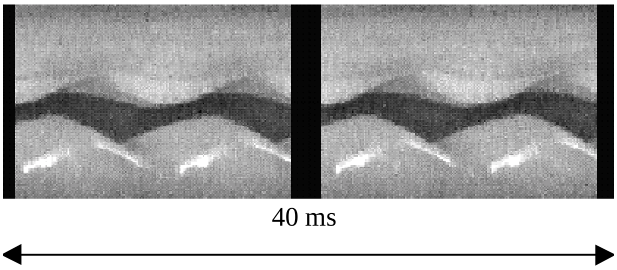

Figure 1: VKG-image, speaker EV, breathy phonation. (Position of the measurement line: middle of the vibrating vocal folds, perpendicular to glottal axis. This holds for all VKG images below.)

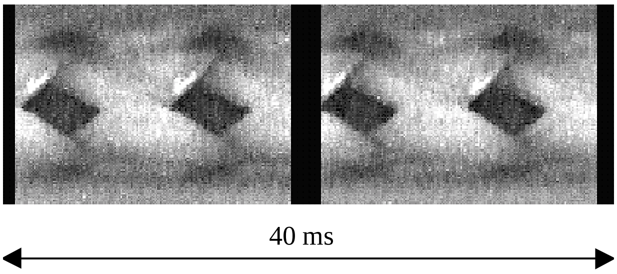

Figure 2: VKG-image, speaker EV, normal phonation.

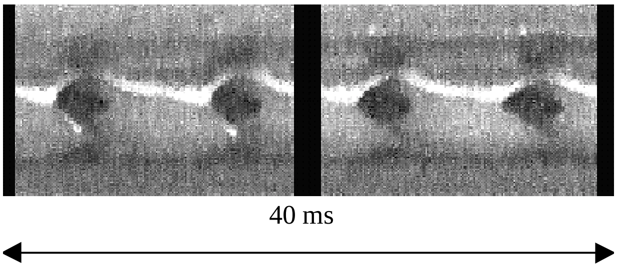

Figure 3: VKG-image, speaker EV, pressed phonation.

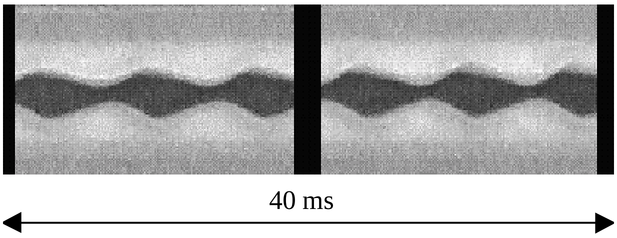

Figure 4: VKG-image, speaker JS, breathy phonation.

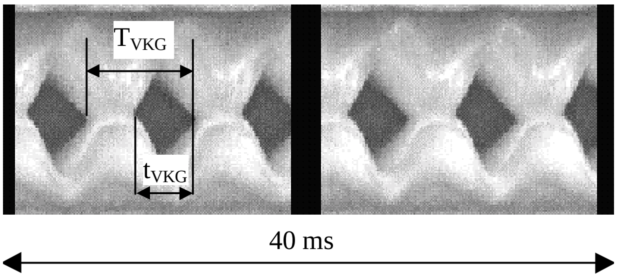

Figure 5: VKG-image, speaker JS, normal phonation.

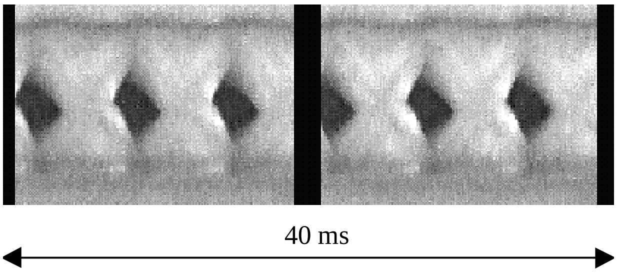

Figure 6: VKG-image, speaker JS, pressed phonation

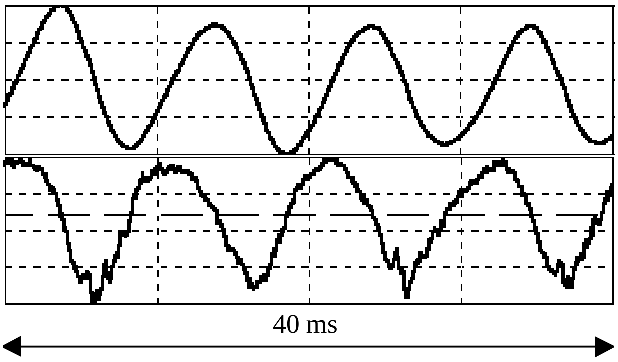

Figure 7: Glottal flow (upper panel) and its derivative (lower panel, zero level shown by a dashed line), speaker EV, breathy phonation.

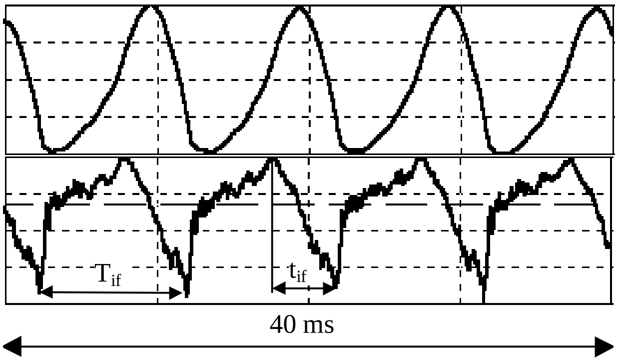

Figure 8: Glottal flow (upper panel) and its derivative (lower panel, zero level shown by a dashed line), speaker EV, normal phonation.

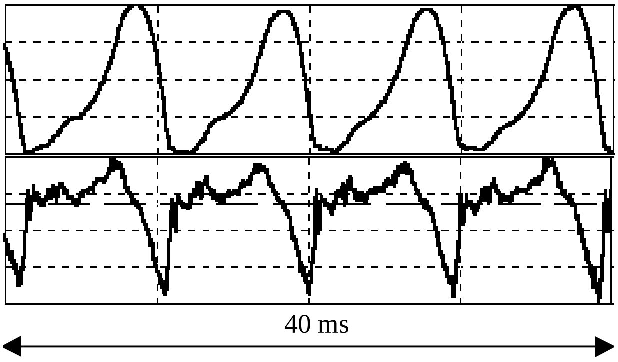

Figure 9: Glottal flow (upper panel) and its derivative (lower panel, zero level shown by a dashed line), speaker EV, pressed phonation.

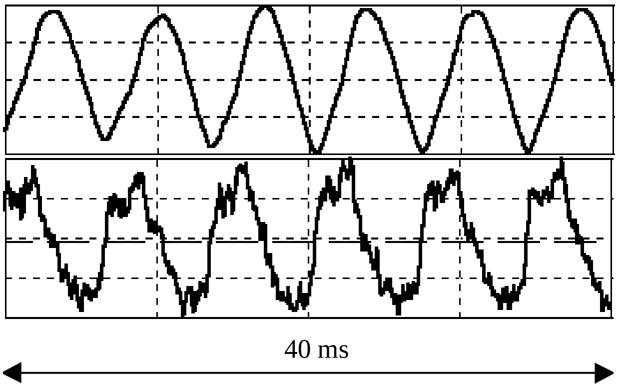

Figure 10: Glottal flow (upper panel) and its derivative (lower panel, zero level shown by a dashed line), speaker JS, breathy phonation.

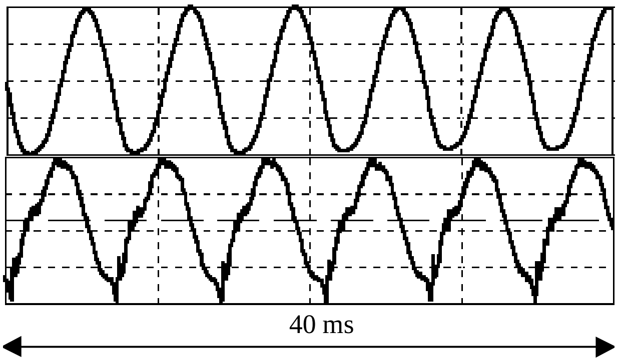

Figure 11: Glottal flow (upper panel) and its derivative (lower panel, zero level shown by a dashed line), speaker JS, normal phonation.

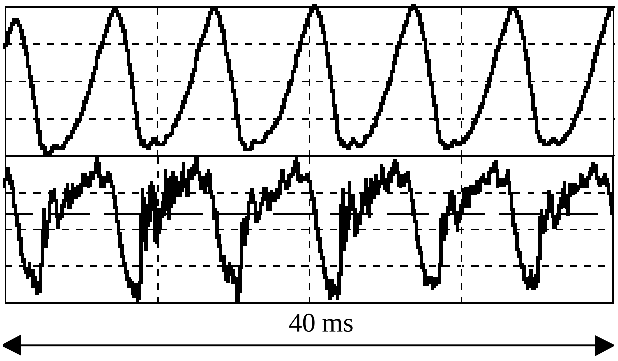

Figure 12: Glottal flow (upper panel) and its derivative (lower panel, zero level shown by a dashed line), speaker JS, pressed phonation.

VKG-images obtained from the vowels produced using normal (Fig. 2 and 5) and pressed (Fig. 3 and 6) phonation types showed for both of the subjects clearly distinguished closed and open phases. For these phonations also the glottal flow given by inverse filtering indicated closure of the glottis. When comparing glottal flow waveforms obtained by inverse filtering to VKG-images it is worth emphasizing that the open phase of the flow is not conceptually equal to the open phase computed from the corresponding VKG-image. This comes from the fact that the VKG-image describes behavior of the glottis only on a single line in the middle of the vocal folds. Therefore, it is possible that there is a flow through the glottis even though the VKG-image shows a closed phase.

Quantitative results on the comparison of the two analysis methods is given in Table 1. The main result given by these numerical data is the similarity between the time length of the open phase of the VKG-image and the time length between the positive and negative peak of the glottal flow derivative in those phonation types (normal and pressed), where the middle line of the vocal folds closes during the glottal cycle.

4. CONCLUSIONS

The preliminary results obtained in this study show that visual information on vocal fold vibration given by videokymography correlates with glottal flow waveforms estimated by inverse filtering. It was found that the length of the open phase of the VKG-image (i.e., the length of time during which the middle line of the vocal folds is open) equals the time difference between the positive peak of the differentiated glottal flow (i.e., the instant of maximal acceleration of the flow) and the negative peak of the differentiated glottal flow (i.e., the instant of maximal deceleration of the flow). This result sounds logical because it implies that opening/closing of the middle line of the vocal folds corresponds to the largest change of the flow within the glottal cycle. However, due to the lack of time-synchronous analysis between the glottal flow and VKG, the current study cannot confirm that opening and closing of the VKG-image coincides exactly with the positive and negative peak of the flow derivative, respectively.

Speaker	Phonation type	T_{if} (ms)	t_{if} (ms)	T_{VKG} (ms)	t_{VKG} (ms)
EV	breathy	10.2	4.5	10.2	10.2
EV	normal	9.8	4.0	9.8	4.5
EV	pressed	9.8	3.2	10.0	3.2
JS	breathy	6.9	3.4	7.2	7.2
JS	normal	6.9	3.8	7.1	3.9
JA	pressed	6.5	2.7	6.3	2.8

Table 1: Fundamental period (T_{if}) and the difference between the positive and negative peak (t_{if}) of the differentiated glottal flow given by inverse filtering. Fundamental period (T_{VKG}) and length of the open phase (t_{VKG}) computed from videokymography.

5. ACKNOWLEDGEMENTS

This study was supported by the Academy of Finland.

6. REFERENCES

1. Alku, P. "Glottal wave analysis with Pitch Synchronous Iterative Adaptive Inverse Filtering," Speech Communication 11: 109-118, 1992.
2. El-Jaroudi, A., and Makhoul, J. "Discrete all-pole modeling," IEEE Transactions on Signal Processing 39: 411-423, 1991.
3. Fant, G. "Some problems in voice source analysis," Speech Communication 13:7-22, 1993.
4. Holmberg, E.B., Hillman, R.E., and Perkell, J.S. "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice," Journal of the Acoustical Society of America 84:511-529, 1988.
5. Rothenberg, M. "A new inverse-filtering technique for deriving the glottal air flow waveform during voicing," Journal of the Acoustical Society of America 53:1632-1645, 1973.
6. Schutte, H.K., Švec, J.G., and Šram, F. "First results of clinical application of videokymography," The Laryngoscope 108:1206-1210, 1998.
7. Sundberg, J., Titze, I., and Scherer, R. "Phonatory control in male singing: A stusy of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source," Journal of Voice 7:15-29, 1993.
8. Švec, J.G. "On vibration properties of human vocal folds: voice registers, bifurcations, resonance characteristics, development and application of videokymography," (Doctoral Dissertation). University of Groningen, the Netherlands, 2000. [http://www.ub.rug.nl/eldoc/dis/medicine/j.svec]
9. Švec, J.G., and Schutte, H.K. "Videokymography: High-speed line scanning of vocal fold vibration," Journal of Voice 10: 201-205, 1996.
10. Wong, D., Markel, J., and Gray, A. "Least squares glottal inverse filtering from acoustic speech pressure waveforms," IEEE Transactions on Acoustics, Speech and Signal Processing 27: 350-355, 1979.