First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

INTERPRETATION OF EGG AND GLOTTAL FLOW BY MEANS OF A PARAMETRICAL GLOTTAL GEOMETRY MODEL

Bert Cranen

Dept. of Language and Speech, Nijmegen University,
P.O. Box 9103, NL-6500 HD Nijmegen, The Netherlands

ABSTRACT

We examined simultaneously measured EGG and glottal flow waveforms of 20 male subjects and found characteristic details which we tried to interpret in terms of a parametrical model of the glottal geometry [1,2]. We came to the conclusion that it is possible to simulate either EGG or glottal flow adequately, but that it is impossible to find model settings which are adequate for simulating both flow and EGG waveforms simultaneously.

The simulated glottal flow waveforms showed the same kind of details as observed in real measurements, i.e. a main flow pulse followed by a smaller one while the dip in between is a function of abduction. We could explain the non-negligible flow after the lower parts of the vocal folds have come into contact, by assuming that it consists of two components. One flow component arises when glottal leakage is combined with vertical phasing [3,4]. The other component was identified as squish flow.

I. INTRODUCTION

During speech production, the behavior of the vocal folds can be observed only indirectly. Consequently, one often has to rely on multifarious and incomplete data in studying vocal fold motion (like EGG, PGG, glottal flow, etc.). Generally spoken, there exists a highly non-linear relation between vocal fold motion and this kind of indirect measurements, so that each signal may draw attention to different aspects of vocal fold motion. Due to this non-linear behavior it becomes extremely difficult to combine insights gathered from individual signals, and to ensure that newly measured signals really add to the knowledge about the voice source system.

An approach which does allow to interpret and combine various measurements in one and the same conceptual framework makes use of a parametric model of the glottal geometry [1,2]. For each measured signal a sub-model is defined which describes how the signal relates to a common set of glottal geometry parameters. Using these sub-models, the dynamically varying parameters describing the glottal geometry can be estimated, and thus, each measured signal may be used to increase the accuracy of one or more of the parameters.

In this paper we want to investigate to what extent a model analoguous to that of Titze [2] is able to simulate certain phenomena in the "closed glottis interval" that were observed in real recordings of glottal flow. In doing so we take the position that not only the main characteristics of the waveforms themselves, but also of their derivatives are worthwhile to examine.

II. MEASUREMENTS

2.1 Method

We selected 20 Dutch male speakers without any known voice pathologies. While each speaker produced one /paepaepae.../ utterance at a constant F_0 and loudness, we measured simultaneously EGG (with a Fourcin-Abberton laryngograph), oral flow, and oral pressure (by means of a Glottal Enterprises mask). All signals were recorded on an FM-recorder. After A/D-conversion (12 bit amplitude resolution) with a sample rate of 10 kHz/signal the signals were stored onto the disk of a digital computer. From the files we selected approximately 5 syllables in the middle of the utterance. The oral pressure recordings will not be taken into consideration since they are of minor importance for the purpose at hand.

Since it takes time for the glottal flow wave to travel from the vocal folds to the transducer in the mask, events at the glottis become visible in the mouthflow with a certain delay. In order to compensate for this delay the EGG signals were first shifted +0.6 ms (corresponding to a pathway of approximately 21 cm). Next, the EGG signals were differentiated. The moments of the maxima in this differentiated EGG were assumed to indicate the moments of glottal closure. Each closure moment was used to determine the start of a rectangular analysis window (length = MIN{closed glottis interval, 3·order}) in the oral flow signal on which a a 12th-order covariance LPC analysis was carried out. Only the flow signals of subjects #7 and 20 were analyzed with a 10th-order predictor. After estimating the formants by applying a root-solving algorithm to the LPC-polynomial, each period of the oral flow signal was inverse filtered. Thus, after inverse filtering each separate period, a glottal flow estimate was obtained for the complete utterance, which was then low pass filtered with a 21st order linear phase FIR-filter with a cutoff frequency of 1000 Hz.

It should be stressed that the chosen inverse filtering procedure is the result of a careful study in which the interpretability of the glottal flow estimates in terms of models like described in [1,2] played a major role.

Finally, we averaged all glottal periods of a given duration within an utterance for each signal of each subject. The chosen period duration was the one which occurred most often (the number of periods varied from 10 to 30). In fact we did not only average the selected periods, but data segments of 2048 data points which had the target period in the center. Thus we obtained characteristic waveforms of oral flow, EGG, EGG-derivative (dEGG), glottal flow and glottal flow derivative for each subject.

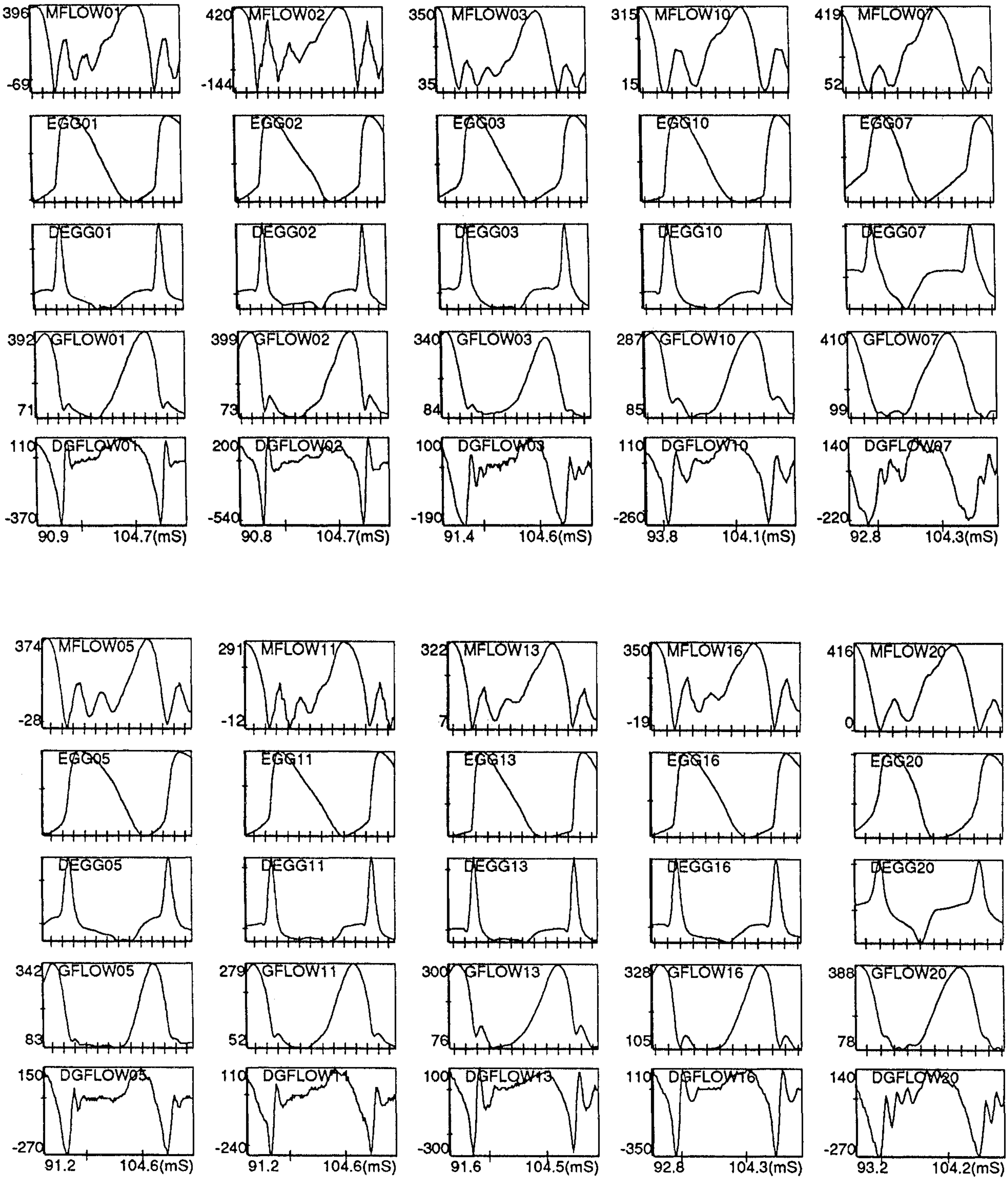

Figure 1: Ensemble averages calculated from measurements of oral flow (MFLOW), electroglottogram (EGG), EGG-derivative (DEGG), glottal flow (GFLOW), and glottal flow derivative (DGFLOW) as derived from 10 different male subjects (cf. section 2.1). Figures along the flow axes denote cc/s; figures along the flow derivative axis denote $cc/s^2 \cdot 1000$. The time scales are chosen so that 1.5 glottal period is visible.

2.2 Results

Examples of such speaker characteristic waveforms are shown in Fig. 1. Note that the glottal flow waveforms are in cc/s and, that they are all scaled with respect to their minimum and maximum, and that the base line does not correspond to zero flow.

Regarding the glottal flow waveforms, the most interesting observation is that our automatic inverse filtering procedure more often than not yields a relatively small secondary flow modulation after the main pulse. This flow modulation sometimes assumes the shape of a small second pulse completely separated from the main pulse; sometimes it is merely a rounding of the back slope; most of the time it is something in between. Also remarkable is that sometimes a distinct change in the rising slope of the glottal flow can be observed (subjects #2, 11). Furthermore, it may be observed that the start of the flat portion in the closed glottis interval of the glottal flow coincides more or less with the start of the flat portion in the dEGG (if present). Finally, note that the symmetry of the EGG-waveforms of subjects #7 and 20 suggests that the vocal folds were relatively abducted [1,2], but that surprisingly, the dc-flow is not significantly larger than for the other subjects. This suggests that part of the dc-flow in normal phonation may be caused by a leak which is more or less independent of the degree of abduction of the membraneous part of the folds.

Regarding the EGG waveforms it appears that especially during the closed glottis interval the derivative can be quite different among subjects. Fig. 1 is organized so that each column shows two examples of the same type. The peak in the dEGG, however, invariably coincides with the minimum in the glottal flow derivative. The behavior of the minima in the dEGG is more variable. Sometimes, one clear global minimum can be detected in which case it coincides with the moment of glottal opening. Very often, however, the minimum in the dEGG is much less clear, or there are various minima. One of the local minima (not necessarily the global minimum) in the dEGG, however, corresponds to the moment where the glottal flow derivative shows a (sudden) rise.

III. STATEMENT OF THE PROBLEM

From signals like shown in Fig. 1 it can be observed that many changes in the (d)EGG coincide with changes in the glottal flow (derivative) (see also [3]).

However, not all changes in the EGG waveform and its derivative coincide consistently with an event in the glottal flow waveform. For events in the closed glottis interval this would be perfectly understandable if the glottis was really closed, since then one would expect the vocal fold contact area to change without a corresponding change in the flow. However, as can be seen from Fig. 1, the glottal flow waveform is neither zero, nor entirely flat in the closed glottis interval.

One cannot claim to understand how EGG and glottal flow are related to glottal geometry before one is able to simulate observations as mentioned above by means of a model. In this paper we want to concentrate on the question to what extent Titze's model (if adapted to incorporate the "effective glottal area" concept as described in [3,4]) is capable of simulating the smaller flow modulation after the main closure gesture.

IV. MODELING

4.1 Method

For simulating the EGG we have used the same approach as in [1,2] and assumed that EGG is proportional to vocal fold contact area. For simulating flow we adopted a somewhat different method which is described below.

The model described in [1] uses the minimum glottal area in conjunction with the subglottal and supraglottal areas to control glottal flow. We feel that a more precise modeling of the acoustic impedance is needed if one desires to understand all details in the measured waveforms. Therefore, we used the *effective glottal area* (A_g) based on the work of Ishizaka and Flangan [5]. It is described in more detail in [4] and is defined as

$$A_g \approx \sqrt{\frac{1.37 \cdot A_{g1}^2 \cdot A_{g2}^2}{0.37 \cdot A_{g1}^2 + A_{g2}^2}} \qquad (1)$$

where A_{g1} is the area of the glottal inlet, and A_{g2} the area of the glottal outlet.

The glottal flow (U_g) is related to this area as

$$U_g = A_g \cdot \sqrt{\frac{2 \cdot P_{tr}}{k \cdot \rho}} \qquad (2)$$

where P_{tr} denotes transglottal pressure, k a constant, and ρ the mass density of the air. If the lung pressure (P_{lung}) is substituted for P_{tr}, Eq. 2 describes the so called short circuit flow (U_{sc}), i.e. the flow that would result if no vocal tract is present [6].

As was shown in [4], these equations can explain why a "hump" after the main glottal pulse may exist: It arises as a result of the combination of glottal leakage and vertical phasing. Dips directly following the main pulse (cf. Fig. 1), however, cannot be modeled this way.

If the vocal folds have closed at the lower margins, the air remaining in the upper part is sqeezed out during the rest of the closing gesture. This effect is not accounted for by Eqs. 1 and 2. Since we estimated the volume of air displaced by the squishing effect to be of the same order of magnitude as the volume of air involved in the hump caused by glottal leakage and vertical phasing, we decided to incorporate ***squish flow*** and extended the model accordingly. We have assumed that the glottal flow from Eq. 2 is augmented by a component U_{squish} which equals

$$U_{squish} \approx -dV_g/dt \qquad (3)$$

where V_g is that part of the glottal volume which formes a dead wake for the air flow when the lower parts of the folds are in contact. It is assumed that the air particles show no lateral motion and that the inertive and compliance effects due to the extra volume are negligible.

4.2 Results

The columns in Fig. 2 represent simulation results for two different settings of the model. From top to bottom the following signals are shown: the short circuit flow (U_{sc}), the squish flow (U_{squish}), the sum of these two flows (U_{no_load}), and the real glottal flow (U_g) which results when the model excites a vocal tract (producing an /a/-vowel). The prephonatory shape of the glottis was chosen uniform with a moderate bulging ($Q_s = 0$; $Q_b = 1$) and the vertical phasing coefficient was assumed to be $Q_p = 0.1$.

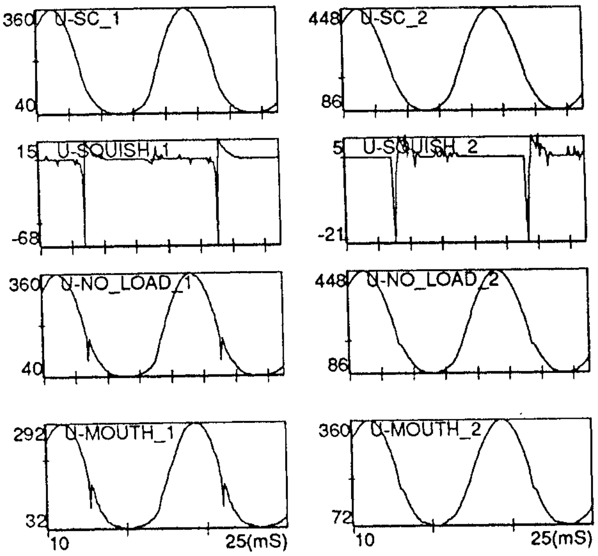

Figure 2: Simulation results for two different degrees of abduction

In the left column the vocal folds were more adducted ($Q_a = 0.9$) than in the right column ($Q_a = 1.4$).

The simulations in Fig. 2 look very much like real measurements. Also the fact that the dip after the main pulse decreases when the abduction increases corroborates our findings that in real speech the dip normally disappears in breathy voices and towards voiced consonants.

According to Eq. 3 the squish flow is proportional with the rate of change in the volume V_g. In other words, to obtain a squish flow of a reasonable amplitude, the vocal folds have to close over a reasonable length at approximately the same instant (no horizontal phasing) and reasonably far down. This is achieved best with adducted vocal folds and a glottis which has a uniform crossection over at least part of its depth.

Note that the minimum flow in the simulations is rather low in comparison with the real measurements. The minimum flow could be increased by increasing the abduction. However, increasing abduction would decrease the squish flow and consequently, the dip would vanish. This may indicate that the quantitative aspects of the model are not quite right. On the other hand we have already noted that the real measurements seem to indicate that dc-flow has not necessarily a direct relation to abduction.

The EGG waveforms generated by the settings mentioned above are not realistic (cf. [1,2]). They are too symmetric, although the extrema of the dEGG have the correct position in comparison with measurement data. More realistic waveforms can be obtained by more convergent pre-phonatory shapes.

The problem is that when using a larger Q_s (e.g. $Q_s = 3$: the nominal value according to [2]), flow is almost exclusively controlled by the upper part of the folds (cf. Eq. 1) and that the squish flow decreases dramatically. The fact that it is impossible to find one setting of the model which is suited to simulate both EGG and glottal flow, together with the observation that the model is unable to simulate the variability in real dEGG waveforms suggests that more investigations are needed into how the glottal shape varies as a function of time.

V. CONCLUSIONS

We examined simultaneously measured EGG and glottal flow waveforms for 20 male subjects. Both the glottal flow and EGG waveforms and their derivatives show characteristic details which are worthwile to simulate, in order to test our knowledge of the voice source.

We concentrated on the simulation of glottal flow and succeeded to generate waveforms which show the same kind of details as observed in real measurements, i.e. a main flow pulse followed by a smaller one while the dip in between is a function of abduction. The phenomenon was interpreted as a result of two combined effects. The first effect was described in earlier work [3,4]: a "hump" on the back slope in the glottal flow must be expected when glottal leakage is combined with vertical phasing. The second effect was identified as a consequence of squish flow. Also the dependence of the dip on abduction could be explained.

The settings of the model needed to generate reasonably looking flow waveforms are not suited to generate adequate EGG waveforms; reversely, the settings needed to generate reasonably looking EGG waveforms do not allow the simulation of the desired details in the glottal flow. The inability of the model to generate both credible EGG and glottal flow waveforms with one and the same setting of the parameters suggests that the model still needs some refinements in the way the time varying glottal shape is modeled.

An interesting observation is that the duration of the second pulse in our interpretation apparently has something to do with vertical phasing. The fact that we were able to give a phsiological explanation for this second pulse gives us extra confidence that our method of inverse filtering is correct. Apparently not all "ripple" in the glottal flow estimates has to be interpreted as formant ripple that was not compensated for. This may be important information for those who do the inverse filtering by hand.

REFERENCES

[1] Titze, I.R. "Parameterization of the glottal area, glottal flow, and vocal fold contact area", J. Acoust. Soc. Amer., Vol. 75, pp. 570–580, 1984.

[2] Titze, I.R. "A four-parameter model of the glottis and vocal fold contact area", Speech Comm., Vol. 8, pp. 191–201, 1989.

[3] Cranen, B. "Simultaneous modeling of EGG, PGG and glottal flow", To be published in Proc. of Vocal Fold Physiology Conference, Stockholm 1989.

[4] Cranen, B., Boves, L., "On subglottal formant analysis", J. Acoust. Soc. Amer., Vol. 81, pp. 734–746, 1987.

[5] Ishizaka K. and Flangan, J.L. "Synthesis of voiced sounds from a two mass model of the vocal cords", Bell Syst. Techn. Journ., Vol. 51, pp. 1233–1268, 1972.

[6] Ananthapadmanbha, T.V. and Fant, G. "Calculation of true glottal flow and its components", Speech Comm Vol. 1, pp. 167–184, 1982.