GLOTTAL LEAKAGE STUDIED BY MEANS OF SIMULTANEOUS VIDEO-STROBOSCOPY AND FLOW MEASUREMENT

B. Cranen and F. de Jong
Nijmegen University, Dept. of Language and Speech, Phonetics section
P.O. Box 9103, 6500 HD Nijmegen, The Netherlands

ABSTRACT

Laryngoscopic studies often report that a glottal leak is quite common in females and children but not in males. This seems in contradiction with the general finding that a dc-offset component in the glottal flow is normal for both females and males. In this study this phenomenon is examined in more detail. Stroboscopic images were recorded on videotape simultaneously with oral flow. The latter signal was used to estimate glottal flow waveforms by means of inverse filtering. We come to the conclusion that the glottal flow may have a dc-offset component even though the corresponding images do not reveal a leak.

INTRODUCTION

During the last few years the insight has grown that most people, both males and females, show a dc-offset flow when producing vowels at moderate loudness levels. From data presented in the literature it can be inferred that this dc-offset flow may attain values of 10–20% of the peak-to-peak value which, in our view, can only be explained by assuming a leak. In a recent paper the acoustic consequences of glottal leakage were discussed from a theoretical point of view [1]. In that paper the concept of a parallel leak (a leak completely separated from the time-varying membranous glottis, e.g. in the form of an opening in the cartilaginous portion of the glottis) was introduced to explain certain phenomena from an acoustic point of view. At the same time, though, it was recognized that the plausibility of such a parallel leak still needs to be proven. This paper presents a first exploration in that direction.

In various laryngoscopic studies it has been reported that a leak in the posterior commissure is quite common for females and children. However, for males, these studies invariably seem to suggest complete glottal closure. The latter is in contradiction with the finding that (for moderate loudness levels) not only the flow of females but also the flow of males generally contains a dc-offset component. This contradiction can in our view only be solved by assuming that (1) the laryngoscopic examination itself causes the subject to adopt a different phonation behavior with and without viewing equipment, (2) there exists a flow generating mechanism which does not require a leak (e.g. vertical motion of the folds), or that (3) the laryngoscopic view is not complete in the sense that not every possible leak is always visible.

EXPERIMENTAL SET-UP

In order to decide whether one of the aforementioned hypotheses is more likely than the other, we carried out a pilot experiment on two male subjects of which we collected video-stroboscopic image data of the vocal folds (stored on a video tape) *simultaneously* with measurements of the following signals

- oral flow measured by means of a circumferentially vented mask
- electroglottogram (EGG)
- microphone signal of the sound outside the mask (speech produced by the subject and the ENT clinician who handled the fiberscope commenting the images)

The image data, collected by means of a Kay LS9100 Rhino Laryngeal Stroboscope of which the stroboscopic flashlight was triggered by the EGG signal, were stored on a Super VHS video recorder. The microphone signal was fed to the audio channel of this VCR. All non-image signals (oral flow, EGG, and microphone signal) were stored on an FM recorder with a frequency response that was flat upto 5 kHz. Since the microphone signal comprising the subject's speech and comments of the ENT clinician was stored on both recording devices, this signal allowed (a rather crude) synchronization of the video images on the one hand, and the signals on the FM recorder on the other. However, for the stationary type of signals we analyzed in this pilot experiment (the vocal fold behavior generally did not significantly change over intervals of one or more seconds), we considered this sufficiently accurate.

In order to be able to collect image data simultaneously with oral flow data, we made a hole in the mask approximately at the level of the nostrils through which the endoscope could be inserted under an approximately horizontal angle. To ensure an airtight seal between mask and endoscope, we clamped a soft, flexible piece of silicone rubber with a hole slightly smaller in diameter than the endoscope, between the mask and a small metal plate. Both the mask and the metal plate contained a hole slightly larger in diameter than the endoscope. In order not to render the mask useless when the endoscope is not used, the hole can be closed by a metal plug of the right diameter.

Before the experiment started the flexible endoscope was fed some 15 cm through the hole in the mask (approximately the point where it needs to be to yield an adequate image when the mask is in place on the subjects face). Next the fiber (with the mask some 15 cm from the tip) was inserted via the nostril and gently pushed forward until the mask reached the subject's face so that he could push the mask firmly against his face. After the clinician adjusted the fiber's insertion depth to obtain a good quality image, the subject started to phonate steady vowels (/ae/) in as natural a fashion as possible.

The signals recorded on the FM-recorder were digitized at a sampling rate of 10 kHz (12 bit amplitude resolution) and stored on a digital computer. Next, the flow signals were calibrated, using the data of two calibration measurements: one set of calibration data was recorded at the beginning of the recording session and one at the end. Subsequently, each period of the flow signal was inverse filtered using formant estimates extracted from the closed glottis interval of that very period (using a pitch-synchrounous covariance LPC analysis). The analysis window (with a fixed duration of 3.6 ms) was chosen to begin at the the location of the peak in the EGG-derivative (after shifting this signal the appropriate amount of samples to account for the acoustic delay of the flow signal from glottis to sensor). The dc-offset flow value per glottal period was obtained by taking the minimum in a low-pass filtered version of the inverse filtered flow (cut-off at 1 kHz).

RESULTS

Since inverse filtering is less error prone for sounds with a high first formant, only recordings of the vowel /ae/

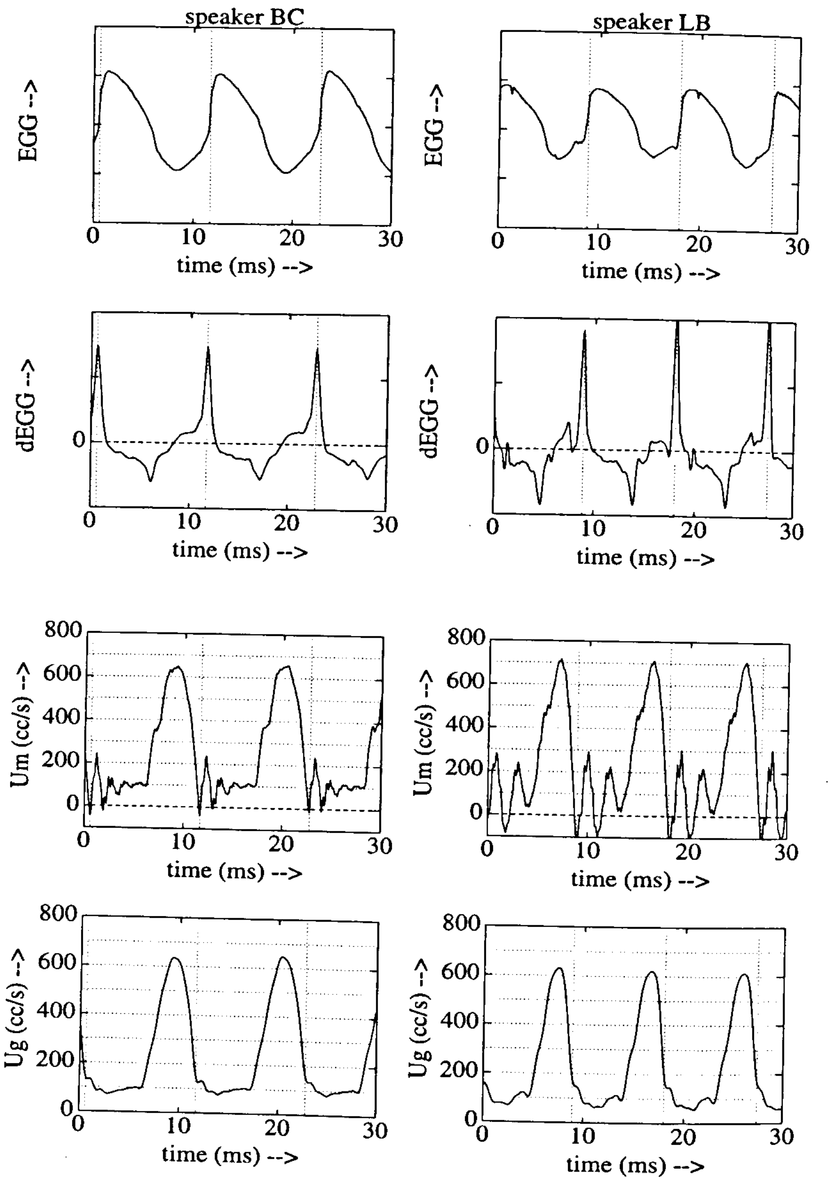

Figure 1: From top to bottom are shown the EGG, the derivative of the EGG, the oral flow signal and the inverse filtered flow signal [left: speaker BC (from segm #1); right: speaker LB (from segm #3)]. Vertical dotted lines denote the automatically detected moments of glottal closure. The videostroboscopic images that were collected simultaneously with these signals did *not* show any leak.

are considered here. From the recorded signals 3 segments per speaker of 1.5-3 seconds were selected during which, according to the information on the audio channel, glottal closure was complete. For each speaker one representative example is shown in figure 1. The corresponding image data on the video tape were inspected to verify that closure was indeed complete. For each segment, the average dc-offset (μ) and the standard deviation (σ^2) were calculated. The values are shown in table I (number of glottal periods in each segment is denoted by N).

Table I: dc-offset flow (cm^3/s)

	speaker BC			speaker LB		
segm	N	μ	σ^2	N	μ	σ^2
#1	167	77	51	140	135	19
#2	291	65	21	320	64	11
#3	183	56	20	334	57	20
total	641	66	32	794	73	33

Note that for speaker LB segment #1 comprises the /ae/ portions of a /paepae.../ utterance while the other two segments are from /ae/ vowels spoken in isolation. For speaker BC all three segments are from isolated /ae/ vowels. Unfortunately, with this speaker the fiber caused to much discomfort during /paepae.../ productions to warrant natural phonation.

DISCUSSION

Although the videostroboscopic images did not reveal any leak, both speakers show a non-negligible dc-offset value in the inverse filtered flow. For the isolated vowels this dc-offset amounts approximately 60 cm^3/s which is close to 10% of the peak value ($\approx$ 600 cm^3/s). For the vowel parts of the /paepae.../ utterance (LB: segment #1) we find almost double that value.

Regarding the hypotheses formulated in the introduction we come to the following conclusions. The fact that we observe a dc-offset flow even with an optic fiber in place suggests that it is not likely that phonation behavior is very different (at least not in a qualitative sense) whether viewing equipment is present or not. Thus, we have to explain our findings either by assuming that the laryngoscopic view is not complete in the sense that not every possible leak is always visible, or by assuming that a combination of squish flow and vertical motion of the folds is responsible for the extra flow.

The latter possibility is rather unlikely as well. In order to get a dc-offset flow of 60 cm^3/s during the entire closed glottis interval ($\approx$ 5 ms) requires that 0.3 cm^3 of air is displaced. If we estimate the area of the membranous part of the vocal folds that can take part in the vertical motion to be 0.6x1.5=0.9 cm^2 (which we think is an over-estimation), this would imply that the average vertical displacement of the folds during this interval would still have to be approximately 0.3 cm which seems irrealistically large.

As a consequence, we believe that the major part of the dc-offset flow is due to a leak. In order to account for a dc-offset value of 10% of the peak flow value, a leak opening is required which is approximately 10% of the maximum glottal opening. A reasonable estimate of the latter is 15-20 mm^2. Since it is very unlikely that an opening of 2 mm^2 would easily be overlooked, we are forced to postulate that such an opening does exist, but that the view on that opening is blocked by the arytenoids. In order to draw decisive conclusions, ways will have to be found to get reliable 3D images of the glottal region during phonation.

REFERENCES

[1] Cranen, B. and Schroeter, J. (1995) "Modeling a leaky glottis", *Journal of Phonetics*, vol. 23, pp. *unknown yet*.