M.Tech. credit seminar report, Electronic Systems Group, EE Dept, IIT Bombay, Nov 2004

Glottography for the Diagnosis of Vocal Disorders

Priyanko Mitra (Roll: 04307022)
Supervisor: Prof. P. C. Pandey

Abstract- Glottography is a general term used for describing methods to monitor the vibration of the vocal folds. Different techniques of glottal investigation have been studied in this report, with emphasis on ultrasound, electric and electromagnetic methods. Applications of glottography, especially impedance glottography, such as detection and classification of vocal pathologies, pitch determination, speech synthesis, voiced-unvoiced classification etc. are discussed here. Finally, the report presents a brief hardware overview of the impedance glottograph developed at IIT Bombay.

1. INTRODUCTION

Speech is produced by the acoustic excitation of the vocal tract by an air stream derived from the lungs and pulsed at a rate that is determined by the vibration of the speaker's vocal folds. The frequency of vocal fold vibration determines the pitch of the voice, which is directly correlated to intonation [1]. The manner in which the vocal folds vibrate contributes to the sound quality produced by the speaker. Also, it has been hypothesized that pathological larynx manifests itself in an abnormal vibratory pattern [2].

But the larynx, due to its position in the throat, is very difficult to study. Glottis is the aperture in the larynx, which regulates the air-flow by changing its area. Glottography is a measurement of the time variation of the glottis during phonation. Glottography usually involves the transmission of a probe signal from one side of the larynx to the other, with the time variation of the glottis modulating the probe's properties [3]. The modulation is then detected and interpreted in terms of the expected geometry of the glottis, which is formed by laryngeal tissues that are in partial stages of contact during the phonation cycle. The probes used in the past have been electrical current flow, ultrasonic waves, light transmission, and airflow as generated by the speaker. Glottography thus makes possible the physical measurement of diverse vocal fold parameters such as pitch, jitter, shimmer, closed-, open-, and speed-quotients, and other perturbation measures.

This report first briefly describes some aspects of the phonatory bio-mechanics. Then it gives a comparative study of different glottographic techniques, with emphasis on impedance glottography (IGG) i.e. measurement of the varying electrical impedance of the vibrating vocal folds during phonation. In the next section, different applications of IGG are discussed. Finally, the report concludes with a discussion of some commercially available electroglottographs, and an outline of the hardware configuration of the impedance glottograph developed at IIT Bombay [4].

2. SPEECH PRODUCTION PROCESS

Lungs, vocal tract, and larynx are the main organs related with generation of sound. The lungs are the source of airflow. The vocal tract is an acoustic enclosure and acts as acoustic filter shaping the spectrum of the generated sound. The source of most speech occurs in the larynx. There are two folds of the muscular bundle known as vocal chords inside the larynx. These vocal chords obstruct the airflow from the lungs and produce audible vibrations that make the speech. The mechanism of generation of sound is known as phonation. The vibrations consist of three phases namely contact phase, separation phase and open phase [4].

During normal breathing, air passes the larynx and vocal tract unobstructed, creating little or no sound. Voicing occurs when the path is constricted or totally closed, interrupting the airflow to create turbulent pulses of air. When the vocal folds are adducted and air is expired, sub glottal air-pressure pushes the vocal folds apart, and air flows rapidly. This immediately creates a partial vacuum between the vocal folds by Bernoulli's Principle, which pulls them once again towards each other. This stops the airflow, building pressure again so that the folds again open and thus, a vibratory motion is set in.

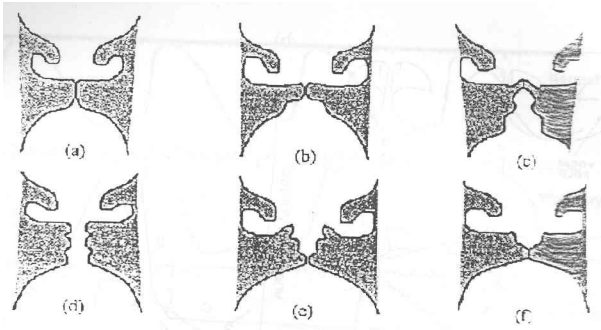

Fig. 2.1 : The vocal fold vibration cycle [5]

3. GLOTTOGRAPHIC TECHNIQUES

We can have reliable results if we can directly measure the vocal fold movements described in the previous section, instead of the speech pressure waveform. There are different glottographic methods by which this can be achieved. Laryngoscopy involves optical measurement of the vocal fold displacement. Direct laryngoscopy involves viewing by a scope, while the indirect method requires introduction of a mirror via a nostril. Thus, normal phonation is not possible during laryngoscopy [4].

In transglottal illumination, the light source is inserted through one nostril and the nasopharynx. The light obtained on the other side of the vocal folds is measured by a photocell placed externally on the wall of the throat. This approach provides a good measurement of fold displacement, but it is difficult to maintain the light source and photocell in position and its introduction is offensive to many subjects [4].

Stroboscopy produces a composite of glimpses from many cycles of vibratory motion of vocal folds. It provides a relatively accurate picture for regular vibratory pattern.

However, when there is irregular vibration, it is difficult to adjust the timing of strobe flash to actual periods of vocal fold vibration [6].

Photographic techniques work only during production of certain vowel sounds, as they require the mouth to be opened wide, the tongue to be positioned low in the oral cavity, the epiglottis to be forward, etc. In addition, the vertical phase difference of the vocal folds during low-intensity phonations in the chest may make measurements of glottal opening and closing times rather difficult. Some of the problems associated with photographic procedures are not inherent in radiographic observations of laryngeal structures during phonation [7]. But they have their own limitations. For example, coronal laminagrams of the larynx present only average configurations of the vocal folds over a number of vibratory cycles. Again, stroboscopic laminagraphy requires highly trained subjects who can hold an almost constant fundamental frequency during phonation. Most importantly, radiographic procedures expose the subjects to cumulative radiation doses.

Airflow measurement or flow glottography can be used for vocal fold vibration measurement as airflow from the lungs is associated with vibration of vocal folds, and results in a pulsating component. This pulsating component can be used for deriving laryngeal function by an inverse filtering technique applied to the speech pressure waveform [1]. However, anechoic recording and complex analyzing equipment are required.

In ultrasonic glottography, the ultrasonic Doppler frequency shift is used as a means of continuously monitoring the velocity of vocal fold motion during voice production [7]. The details are given in the next section.

Electroglottography (EGG), or impedance glottography (IGG), is the non-invasive measurement of the time variation of the degree of contact between the vocal folds during speech production. The device used to measure the vocal fold contact area (VFCA) is called the impedance glottograph, electroglottograph, or laryngograph. It is preferable not to call it an electroglottograph, since no electrical signal of biological origin is used, rather an external electrical signal is applied. The principle of operation of the device and the waveform obtained are discussed in section 5.

In recent times, glottographic sensors using high frequency propagating electromagnetic waves have been developed [3],[8]. Principle of operation of electromagnetic glottography (EMGG) will be covered in section 6.

4. ULTRASONIC GLOTTOGRAPHY

The ultrasonic Doppler velocity monitor is based on the frequency shift produced by a moving source and/or a moving observer [7]. A continuous ultrasonic beam directed into the neck is reflected from various tissue layers and the tissue-air interface at the internal laryngeal wall. If the reflecting surface is moving, then the reflected sound wave will have a frequency

$$f_r = f_t \pm 2 f_t V \cos\theta / (U + V \cos\theta)$$

where, f_t and f_r are the frequencies of the transmitted and reflected beams respectively, V is the velocity of the reflecting surface moving at an angle θ to the incident sound

beam, and U is the transmission velocity through the medium. The difference between the transmitted and reflected beams is called the difference frequency or the Doppler frequency f_d, given by

$$f_d = 2f_t V \cos\theta / (U + V\cos\theta)$$

For all ultrasonic work in the human body, V is negligible compared to U, so that the above equation reduces to

$$f_d = 2f_t V \cos\theta / U$$

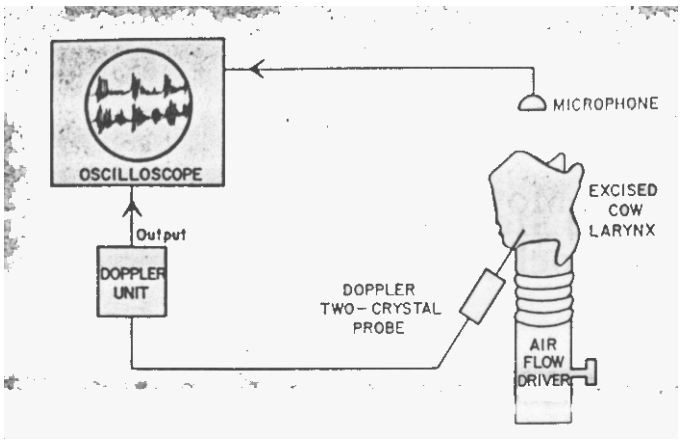

Fig. 4.1 : Simplified block diagram of the experimental apparatus [7]

The reflected signal is beat against the transmitted signal to obtain the difference frequency, f_d, which appears as the output of the ultrasonic Doppler velocity monitor and is proportional to the velocity of the reflecting surface. After plotting the vocal fold velocity as a function of time, and integrating the area under the curve, it is possible to determine the displacement of the reflecting surface as a function of time [7].

However, analysis based on taking the difference between the transmitted and reflected frequencies makes it impossible to differentiate between interfaces moving towards or away from the transducer. To solve this problem, experiments were conducted using a transillumination technique and Doppler monitor simultaneously [7]. The light sensor signal was displayed on one channel of a dual-trace oscilloscope, and the output of the Doppler unit was displayed on the other channel. The position and angle of the Doppler probe was varied until an optimum signal (steady characteristic pattern) was obtained. The output of the light sensor is a dc voltage proportional to the area of the glottal opening, so that synchronous display of both the Doppler and light sensor signals provided an indication of which part of the Doppler pattern corresponded to vocal fold opening, and which to closing.

Attempts to synchronize high-sped cinematography of vocal fold vibration with the Doppler signal have been reported [7]. On the whole, ultrasound glottography shows good correlation with other methods of measuring vocal fold motion and glottal area.

5. IMPEDANCE GLOTTOGRAPHY

5.1 Principle

Two electrodes are held in contact with the skin across the thyroid cartilage at the level of the larynx, and a high frequency ac current (300 kHz to 5 MHz) is injected through them. The supplied current is different for each particular device, but is not stronger than several milliamperes. The voltage between the electrodes depends on the tissue impedance but the typical value is about 0.5 V. In accordance a power dissipation of only several microwatts occurs at the level the subject's vocal folds [9].

The impedance between the vocal folds is a function of tissue path length. When the vocal folds are open, the tissue path length is maximum and hence impedance is maximum. When the vocal folds are closed, the tissue path length is minimum and hence impedance is minimum. Since the current is constant, the voltage across the two electrodes changes in accordance with the impedance between the vocal folds and we get an amplitude-modulated waveform across the electrodes, which is varying in direct proportion to the variation in impedance due to movement of the vocal folds.

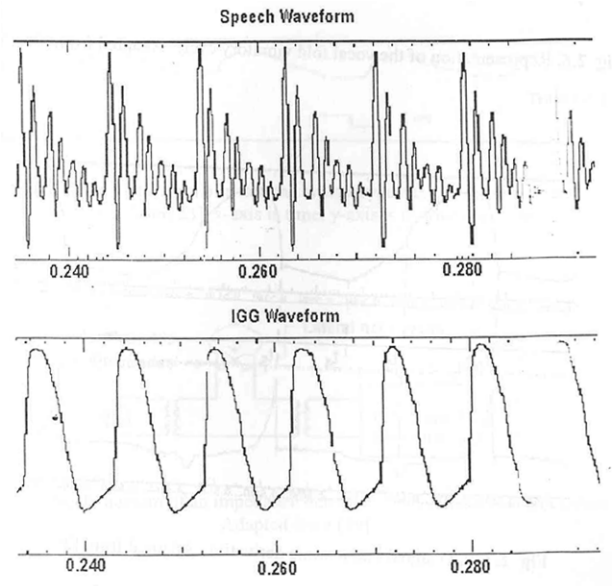

Fig. 5.1 : Speech and IGG waveforms [9]

The received signal is then demodulated by a signal detector circuit. The typical signal-to-noise ratio of the demodulator is about 40 dB. The demodulated waveform is then A/D converted and stored in a computer [9].

Fig 5.1 shows the speech and IGG waveforms. An integral part of the electroglottographic signal is the varying component generated by the vertical movement of the whole larynx. Therefore, the signal of rapid movements of the vocal folds is superimposed on the signal produced by the slower movements of the other structures. The name *Gx* was proposed for the waveform of larynx movement and the name *Lx* for the vibration component. The *Gx* component originates, for example, can be observed in swallowing, but it may also be caused by the vertical movement of the larynx which is related to the voice quality setting of the raised/lowered larynx. Fluctuations of its type are usually removed from further analysis. The DC offset changes (*Gx*) can be evened out because the effects of the varying larynx height are compensated by the use of additional electrodes or high pass filtering of the registered signal. The latter method may involve signal distortion, especially for low-pitched voices. The distortions may be caused by a too high cut off frequency of the filter (or a too wide filter transition band). This can cause the attenuation of the *Lx* signal component. The non-uniform phase response function of the filter can also change the shape of the filtered waveform. FIR (finite impulse response) filters should be used to prevent nonlinear phase shifts in the signal component. It should be noted however, that even the unfiltered output of the EGG device is not free of distortion. Particularly the demodulation circuit whose frequency transfer function may influence the frequency response of the EGG device, especially in the low frequency range constitutes an additional source of signal shape deformation [9].

While laryngoscopy, transglottal illumination and airflow measurement give information about the separation of the vocal folds and very little indication about the nature of the contact, impedance glottography does just the opposite.

The greatest advantage of this method is that it is non-invasive, and gives instantaneous outputs simple to obtain. However, some practical difficulties arise due to the need for separate adjustment for new subjects, and with time for a given speaker. It can also be affected by external movement of connecting leads, changes in surface conduction of the skin caused by perspiration, slow changes of oesophegal movement and vertical displacement of the larynx

5.2 Impedance Glottogram

Impedance glottogram helps in laryngeal function assessment by giving information about the contact phase and separation phases of the vocal folds.

When the vocal folds are open and there is no lateral contact between the vocal folds, peak glottal flow occurs, and impedance is maximal (segment (e) in Fig. 5.2). The waveform in this segment is flat with small fluctuations. Next, the upper margins of the vocal folds make initial contact

Thereafter, the lower margins touch each other, and the vocal folds as a whole continue to close like a zipper. If the vocal folds close very rapidly and along their whole length, the phases (f) and (a) become indistinguishable. So, the closure phase ((f) + (a) in Fig. 5.2) has a steep slope. This knee is found in low to normal voice intensities, the slope of (f) being less than that of (a).

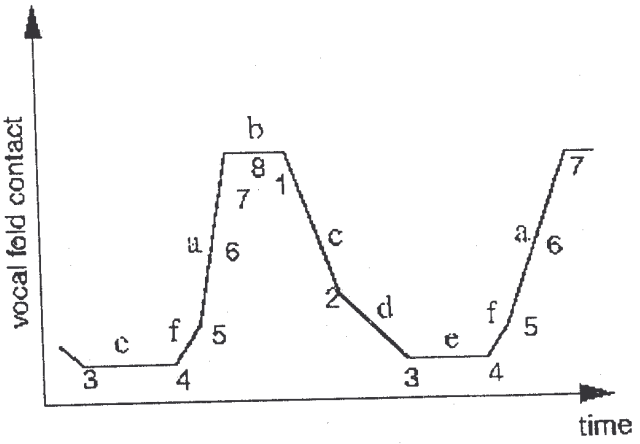

Fig. 5.2: The plot of vocal fold contact area vs. time [4]

Glottal closure is achieved during phase (a) (Fig. 5.2). Just before closure, the vocal folds are almost parallel with a narrow opening along their entire length. Closure occurs almost simultaneously along the entire midsagittal line. It has been noted that there are greater horizontal phase differences during opening than closing.

During the next phase, the vocal folds remain closed, blocking the airflow. Limited fluctuations are observed due to unstable contact between the folds, and the waveform forms a smooth hill, instead of being flat. Contact increases during this phase until maximum is reached, and then decreases again.

A similar explanation applies to the opening and open phase. The contact between the folds starts decreasing and then, the lower margins of the folds begin to separate, starting the opening. Lower margin separation occurs slowly during phase (c) (Fig. 5.2). Then the upper margins also begin to separate, accelerating the rise in impedance till full opening is attained. The glottis grows in size during this phase. As contact between the folds is not maintained anymore, the IGG waveform does not reflect the glottal width, glottal area or the glottal flow.

The time derivative of the impedance glottogram is used in the determination of periodicity of the signal [9]. It can also be used to identify the noticeable changes in the slopes during the phases of increasing and decreasing impedance, which correspond to those of the simplified model of the IGG in Fig. 5.3.

The positive peak of the derivative identifies the instant of glottal closure. This is a good, dependable marker of the pitch period for various voice qualities and intensities.

The negative peak of the DIGG (differentiated IGG) serves as an indication of glottal opening. Vocal fold movements that are reflected in the IGG have two distinct phases. First, the IGG decreases monotonically, as lateral contact between the folds decreases. During this interval, the IGG waveform is convex. Then, as the upper margin separates, the waveform turns concave.

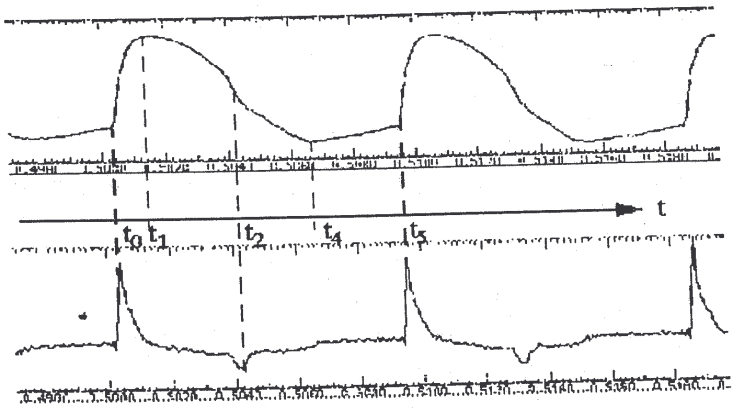

Fig. 5.3: IGG signal and its time derivative [9]

The reliability of DIGG is valid only for normal voices in modal register. Simulation of vocal fold vibratory motion that either departs from modal register or is impaired by a vocal fold nodule, suggests that the DIGG is no longer a reliable indicator of the opening and closing instants of the glottis.

The contact area between the vocal folds can be viewed laterally and transversely. Assuming that the depth of contact does not change during vocal fold vibration, the lateral contact area depends on the length of contact area along the upper margins of the vocal folds. There is, thus, a strong correlation between the IGG and the length of vocal fold contact.

The description of IGG waveform fails for voices with a continuously open glottis. The open phase of the fold movement is better represented by other means, especially by inverse filtering and invasive visual inspection.

Also, a closed glottis cannot be completely determined from IGG alone, because the IGG waveform of a breathy voice with an open glottal chink may look essentially the same as the IGG waveform with complete glottal closure. Other confirming means are to be used if determination of the occurrence of complete glottal closure is very important.

5.3 Laryngeal pathology classification with the help of IGG and PGG

Impedance glottography and Photoglottography together can produce a complete description of a patient's vibratory movement of vocal folds. Some quantitative measures extracted from IGG and PGG signals help in the classification of vocal fold disorders [6].

Speed Quotient is a ratio of the opening time per cycle to the closing time per cycle. The opening time is defined as the time the upper edges open until they reach the peak glottal aperture. The closing time is defined as the time from peak glottal aperture until the vocal folds close. From Fig. 5.4,

$$\mathrm{SQ} = t_A / t_B$$

Open Quotient is the ratio of the time the glottis is open to the time for an entire glottal cycle. The time of glottal opening is defined as the time when the upper edges open till the vocal folds close. It is expressed as

$$\mathrm{OQ} = t_C / t_D$$

Shift Quotient is the ratio of the opening time per cycle to the entire time of glottal opening, calculated as

$$\mathrm{ShQ} = t_A / t_C$$

Jitter Ratio is the absolute value of the difference in time between consecutive cycles over the mean glottal period, quantified as

$$\mathrm{JR} = (t_{D(n)} - t_{D(n+1)}) / t_{D(ave)}$$

where, $t_{D(n)}$ is the time for the nth glottal cycle, $t_{D(n+1)}$ is the time for the (n+1)st glottal cycle, and $t_{D(ave)}$ is average glottal cycle time.

Shimmer Ratio is a measure of the absolute differences in the PGG amplitude to the mean amplitude, simply calculated from two consecutive amplitudes as

$$\mathrm{SR} = 20\log_{10}\left|\frac{A_i}{A_{i+1}}\right|$$

The above ratios are calculated purely from the PGG signal. IGG in conjunction with the PGG provides more details about the glottal cycle, and some new quotients are extracted.

Close Quotient is defined as ratio of the complete closure time to the glottal cycle time, calculated as

$$\mathrm{CQ} = t_F / t_D$$

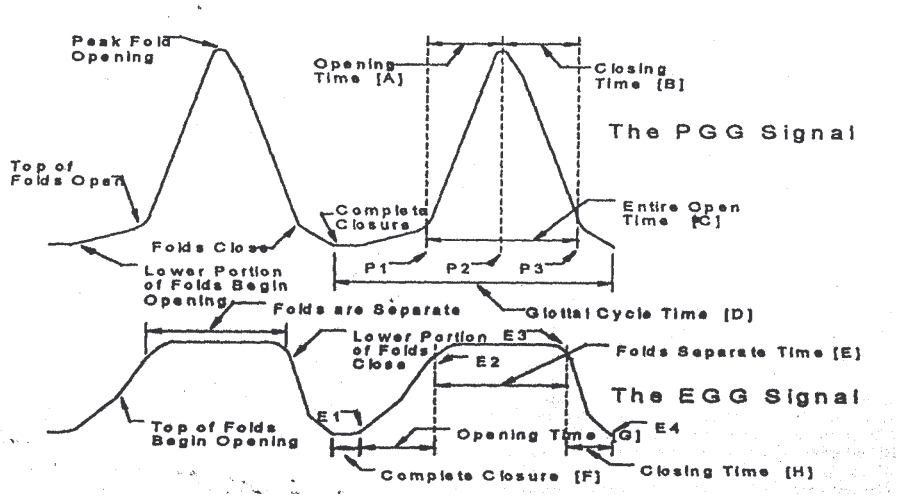

Fig. 5.4: Times for the calculation of ratios and quotients [6]

Relative Open Quotient is the ratio of the time the folds are separated on the IGG signal to the entire open time on the PGG signal, given by

$$\mathrm{ROQ} = t_E / t_C$$

Relative Shift Quotient is a signed quotient that reflects the shift in time between the open times defined IGG and PGG. The equation is

$$\mathrm{RShQ} = (t_{C(mid)} - t_{E(mid)}) / t_E$$

where, $t_{c(mid)}$ is the time at which the midpoint of the entire open time occurs and $t_{E(mid)}$ is the time at which the midpoint of the folds' separate time occurs. The midpoint times are the times with respect to the start of the signal.

IGG Speed Quotient is derived from the SQ defined earlier. It is designed to reflect changes in opening and closing time. ISQ is the ratio of the opening time of the IGG signal to the closing time of the IGG signal. It is defined as

$$\mathrm{ISQ} = t_G / t_H$$

By taking derivative of the IGG and PGG signals, important transition points can be determined. However, as IGG and PGG are both noisy signals, the first derivatives will be very noisy. By employing a low pass filter, key transition points are well retained and noise eliminated effectively.

After the transition points in IGG and PGG are marked independently using percent-of-slope method, the sequence of transition points are checked automatically by a finite state algorithm machine and the location of the marking error is reported if error encounters so that marking can be modified manually.

The final step is the classification of the signals by using the extracted features to determine whether or not the patient has an abnormality in the vocal folds. The designed classifier measures the match probability between the input signals and an existing knowledge database. The database stores the statistical data (assumed to follow normal distribution, for simplicity) of the features for a number of groups representing different types of diseases. One of the groups contains people with normal voice. The probability of match between an input signal's feature and a feature from one of the groups in the classifier is defined as the area of overlap between the two distributions of that feature.

Though the available training data was very limited, it was found that there were noticeable differences in the features among the different pathological groups. The system was able to classify correctly with an accuracy of 64%. The classification accuracy can be greatly improved by with a larger training set, more elaborate grouping methods, more effective control methods of data acquisition and more effective feature extraction method.

5.4 Comparison between high-speed filming, IGG and other glottographic techniques

High-speed films are most commonly used to monitor details of the glottal cycle. However, the technique is difficult, expensive and cannot be performed during natural conditions due to the need of using a laryngeal mirror. In comparison with high speed filming, transillumination can be performed more easily and under more natural conditions of speech. In combination with other measures such as impedance glottography, transillumination is useful for examining the relation between vibratory performance and acoustic output [10].

A comparison between glottal width measures obtained by transillumination and from simultaneous fibre-optic filming showed that temporal information supplied by the two methods was virtually identical. However, to compare smaller faster movements during phonation, a high-speed filming system is required in place of the fibre-optic endoscope.

Contrary to the belief that IGG and PGG provide information about exactly complimentary parts of the glottal cycle, actually they are likely to overlap, because the glottis rarely either opens or closes abruptly over its entire length. Rather, for part of the cycle, the folds are likely to be in contact or separated over only part of their length. The validity of these assertions was tested by Baer et al [10] to assess comparable information available in high-speed films.

High-speed laryngeal films were taken of one male and one female subject producing a steady phonation of the vowel /i/ [10]. High quality acoustic readings were obtained at the time of the filming. PGG and IGG signals were also obtained simultaneously with the high-speed films. Using a computer, frame-by-frame measurements were made from the films during the portions where film speed was constant. Audio and glottographic signals were sampled, digitized and aligned with the film measurements.

The results show that PGG and film measures give the same information about peak glottal opening and glottal closure in normal phonation. There is more uncertainty about the moment of glottal opening as opening is slower than closure. The IGG signal is also consistent with this notion.

In conclusion, films not only give measures of glottal area, but also distribution of width along the glottis. On the other hand, glottographic techniques cannot detect the width distribution along the glottis but can be used to detect the presence of horizontal phase differences during opening and closing.

6. ELECTROMAGNETIC GLOTTOGRAPHY

Electromagnetic glottography (EMGG) is a new technique, whereby a transverse electromagnetic wave in the GHz range is propagated and then detected to obtain information on the condition of the larynx tissue interfaces [3],[8]. A transducer arrangement for combined EMGG and EGG measurements described herein is shown in Fig. 6.1. The EM antennae are placed on the front of the neck near the thyroid prominence (Adam's apple). In this experiment, EGG electrodes were placed additionally on both sides of the neck

Electromagnetic devices can operate either in forward scattering (diffraction) or backward scattering mode (reflection). No improvement in signal strength was obtained in spite of imposing a favourable angle of receiving the reflected signal. So the diffraction mode is favoured.

The transmitted transverse EM signal is a wave packet containing 10 wave cycles at 2.3 GHz, with a wavelength of 13 cm in air, and 1.8 cm in tissue. Average power emitted was 300mW and energy per pulse was less than 1 nJ. The pulses formed a train with a repetition rate of 2 MHz. Absorption co-efficients at 1-4 GHz was 5-10 per cm, allowing about 10 cm penetration into the body and back to the sensor. Reception was accomplished using a homodyne mode detector, signal integration and bandpass filtering. The system could detect motion in the near, intermediate or far field, with one antenna being used as the transmitter and the other as the receiver. In Fig. 6.1, the electric field patterns are shown for the EGG electrodes from right to left across the glottis and the propagating *E* field is shown with arrows near the sending EMGG electrode.The signal is integrated and filtered (analog, single pole, high pass at 70 Hz, and low pass at 7 kHz) so that only the tissue interface motions in the voice harmonic frequency range are detected. Sensor signals associated with movement of vocal folds and related air pressure induced tissue motions occur at pitch frequencies between 70 and 250 Hz. So they are easily distinguishable from signals returning from stationary tissue interfaces by using suitable bandpass filters. Sensors used for these experiments were called GEMS sensors (McEwan 1994, Burnett 1999) because they were optimized for Global ElectroMagnetic Sensing.

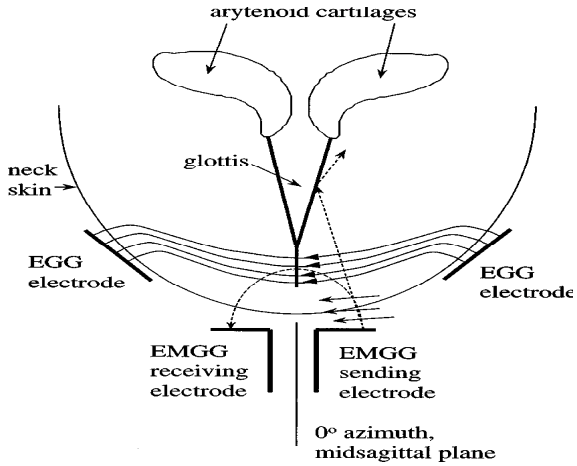

Fig. 6.1: EMGG and EGG electrode placement in a horizontal plane, with a cross section through the neck and larynx. [3]

A digital inverse filter is used later to restore the wave shape. Thus, reflections from stationary or slowly moving tissues, such as slow artery blood flow pulsation, are not detected. This high-pass filtered mode is called the ‘‘field disturbance’’ mode and is particularly useful for survey work, where absolute tissue interface locations relative to the antennae are uncertain, but movement is easily detected [3].

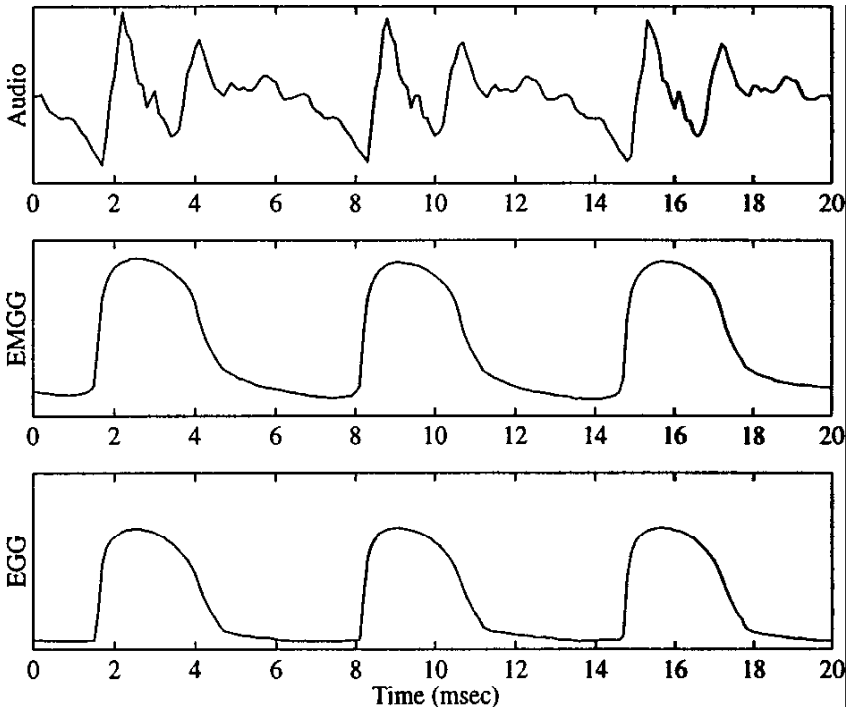

Fig. 6.2: An acoustic microphone signal, the EMGG signal, and the corresponding EGG signal recorded simultaneously from a subject for three pitch periods, and vowel /a/. [3]

Fig 6.2 compares the EMGG signal with the acoustic signal and the EGG signal for normal, low pitch voice from subject 1. In this example, as in other figures to follow, the acoustic signal was shifted 1.4 ms in time to correct for the slower sound speed from the glottis to the microphone, in contrast to the near instantaneous light speed of the EGG and EMGG. The negative peak on the acoustic signal is associated with the closure of the glottis. Both sensor signals show agreement on glottal closure and glottal opening, as well as the general shape of the contact pattern. This is a case for best agreement between the two waveforms.

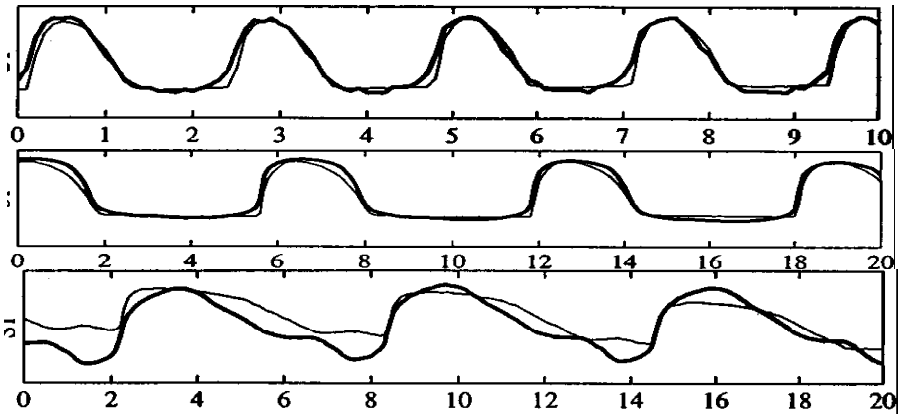

Fig 6.3: Examples of falsetto, breathy and pressed voice; EMGG in bold lines, EGG in fine. [3]

Fig 6.3 compares the EGG and EMGG signals for falsetto, breathy and pressed voices. In spite of slight differences, the two curves show more or less the same trend.

Signal amplitude in the homodyne sensor is proportional to both a change in distance of the target tissue interface from the sensor and to scattering cross-sections of the reflecting locations. Hence, there is an ambiguity in separating targets with large frontal areas but small position changes from those targets with small frontal areas but relatively large position changes.

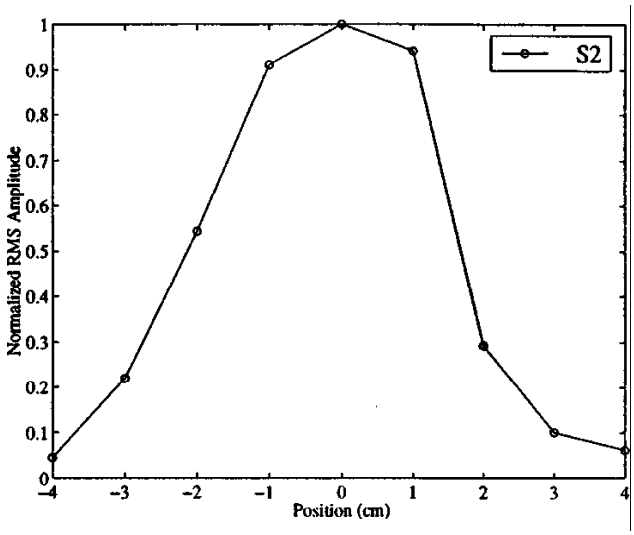

Fig 6.4: RMS value of EMGG signal vs. antenna azimuth angle [3]

Fig 6.4 shows that the root mean square value of the EMGG signal decreases with increasing azimuth angle from the mid-sagittal plane on both sides, which clearly speaks in favour of the diffraction mode.

Since the antennas used in EMGG are non-focusing and several EM cycles were transmitted per pulse, it was not possible to find which oscillating tissue interfaces were directly responsible for the signals. However the position of the sensor and longitudinal range limit of the sensor restricted the site of oscillations to the vocal folds or tracheal walls. Focused time-gated EM sensor signals are being developed to accurately measure tissue interface motions and determine exact sources of EM wave reflections.

Comparison between EMGG, IGG and PGG has been done by Titze et al. [3]. The waveforms of EMGG were found to be so consistent with IGG that it is natural to associate EM scattering site with changes in glottal tissue configuration. Both IGG and EMGG showed agreement on instants of glottal opening, closure as well as general shape of contact pattern. Timing between EMGG and PGG were validated by high-speed photographic systems, coupled to laryngoscopes. Closer agreements were found between EGG and EMGG when the anterior glottis was the primary probing region, suggesting that the sensor operated in the diffraction mode than in a reflection mode in this near to mid-field region.

The greatest advantage of EMGG is that it provides sensing at a distance, thus leaving space for simultaneous use of other sensors. In future with sensor arrays and more focusing antennas, improved spatial discrimination can be obtained.

EM sensor data combined with corresponding acoustic data enable robust, accurate onset of voiced speech, pitch detection, acoustic signal denoising with -20dB reduction, narrowband speech compression and speaker verification.

7. APPLICATIONS OF THE IGG

7.1 Laryngeal function assessment

Pathological conditions of the vocal folds affect the impedance variation during the movement of the vocal folds, and hence are reflected in the impedance glottogram [4]. This has been established by viewing the impedance glottogram of a normal voice, whispery voice, creaky voice, breathy voice and tense voice.

Normal voice is characterized by three distinct phases – a relatively sharp rise during rapid closing of the vocal folds, a slow fall indicating the separation of the vocal folds, and a flatter base denoting the open interval of the vocal folds. The duty cycle of the signal is approx. 50%.

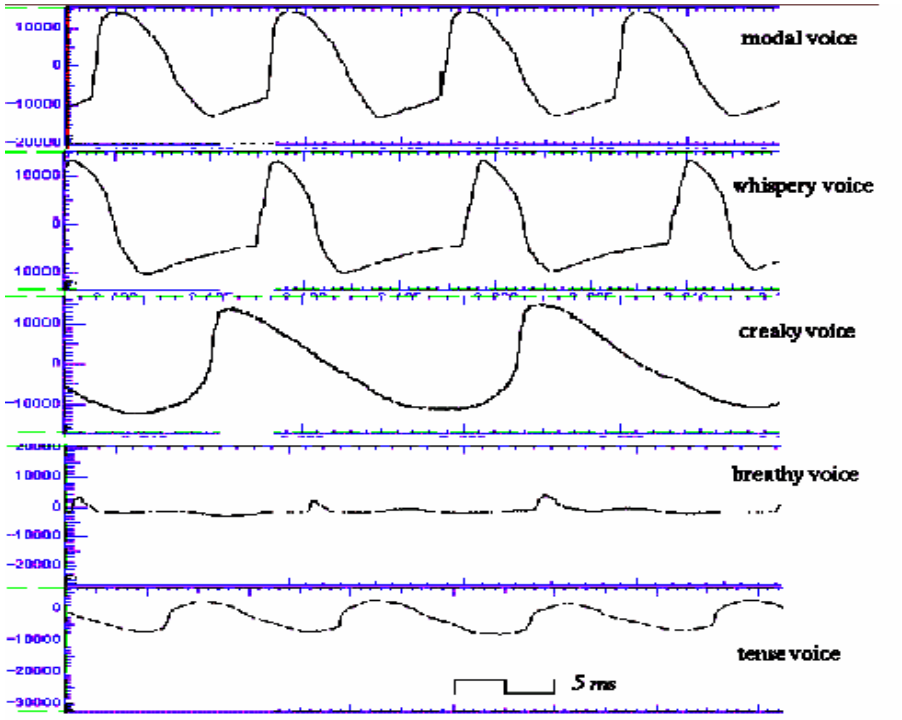

Fig. 7.1 : Impedance glottograms of normal, whispery, creaky, breathy and tense voice [9]

For a whispery voice, the increase and decrease in impedance are much faster. Moreover, the duty cycle is lower than in normal voice.

For a creaky voice, the shape of the waveform is triangular with smoothened edges. The impedance rises fast but falls gradually. So, defining the signal duty cycle is very difficult as no knee can be identified.

Breathy voice has a lower peak amplitude due to poor contact between the vocal folds or incomplete closure. The maximum contact phase is extremely short.

A strongly increased muscular tension in the larynx results in tense voice. Notable features are an almost rounded waveform, low peak-to-peak amplitude, gradual increase as well as decrease in impedance, a comparatively long maximum contact phase and a duty ratio comparable to that of normal voice.

7.2 Spectral analysis

Conventional FFT-based spectral analysis techniques as well as modern spectral estimation procedures usually employ windows like Hamming, Hanning, Kaiser etc. Inclusion of glottal open regions and differing amounts of excitation periods inside the arbitrarily placed analysis window cause errors. This can be solved to a large extent if the analysis window is restricted to a maximum of one pitch period in length, and that too, extending over only the glottal closed region. This information is available from the time aligned, appropriately scaled period of the IGG signal, which results in proper frame positioning for speech analysis for vocal tract-related parameters [2].

7.3 Formant tracking

Linear Predictive (LP) analysis helps in following the trajectories of the first three or four formants. The LP parameters are found using a fixed frame, pitch-asynchronous autocorrelation or covariance analysis. Its performance often degrades when the fundamental frequency of phonation approaches the location of the first formant. Also, past use of a closed-phase pitch-synchronous analysis suffered from the difficulty of isolating the closed phase region in successive periods of speech. By exploiting the glottal open status information available in the time-aligned IGG signal, proper frame positioning can be easily done for speech analysis [2]. A two-channel system using pitch-synchronous and pitch-synchronous closed phase analyses performed better than other existing formant tracking techniques.

7.4 Voiced-Unvoiced classification

IGG helps in simple and accurate classification of speech into voiced, unvoiced, mixed and silent regions. For voiced-unvoiced classification, thresholding of IGG signal suffices because IGG is almost zero in unvoiced regions and non-zero as well as periodic in voiced regions. Hybrid speech IGG algorithms are used for unvoiced-silent or voiced-mixed segregation. For example, the algorithm used by Childers and Larar [2] first computed the statistics of a short silent region assumed at the beginning of the utterance. Then energy and zero-crossing thresholds for both speech and IGG were set accordingly. Next, an interval is checked for silence. If it is not silent, glottal activity is monitored for voiced-mixed classification. If all three classifications fail, it is categorized as unvoiced speech.

7.5 Pitch estimation

The different IGG-based algorithms used for computing the pitch contour of an utterance differ from each other in whether the pitch is computed per frame or per period, and which IGG feature is used. In the algorithm used by Childers and Larar [2], for a given frame, a pitch estimate based on IGG zero-crossings is first calculated. Using a search region based on this estimate, a second pitch value is then computed from the distance between the minima in the differentiated IGG. The former estimate suffers from the drawbacks that zero-level crossings only approximately correspond to the separation of folds during the open phase, and the exact point of zero crossing is influenced by local mean removal, tape recorder distortion, etc. This estimate is used only if the second one is zero, or when the two estimates differ widely but the first one is closer to the pitch frequency of the previous frame. In all other cases, the second estimate is more reliable.

7.6 Speech synthesis

Synthesis of high quality speech can be done using IGG aided analysis methods [4]. This is the case since naturalness and intelligibility of synthesized speech are influenced by accuracy in vocal tract modeling, voiced-unvoiced classifications, pitch detection, and the nature of the excitation used. These parameters can be accurately obtained using IGG.

7.7 Frequency and amplitude perturbation analysis

Perturbation analyses of frequency and amplitude in a voice signal have been developed to provide objective parameters for early detection of laryngeal disorders [11]. IGG is more suitable than the voice signal for this because the waveform obtained is much less complex than that of the voice signal, and is unaffected by the acoustic resonances of the vocal tract.

Frequency and amplitude perturbations of the IGG reflect different aspects of irregularity of vocal fold vibration. To assess the clinical significance of the frequency and amplitude perturbations of the IGG signal, indices were calculated and compared to the degree of hoarseness evaluated by hearing the recorded voice and seeing the spectrograms. Frequency perturbation was expressed as the mean difference in frequency between consecutive cycles measured in semitones. Amplitude perturbation was expressed as the mean difference in amplitude between consecutive cycles in dB.

In the results, frequency and amplitude perturbations were found to be significantly correlated. Males and females differed considerably in amplitude perturbation, not in frequency perturbation. Spearman's rank correlation coefficient was used to evaluate the association between amplitude perturbation, frequency perturbation, perceptual auditory rating and visual spectrographic rating. Each of the four pairs showed good correlation. Highest correlation was observed between amplitude perturbation of IGG and the auditory-perceptual rating of hoarseness. Frequency perturbation and spectrographic rating showed the lowest correlation. From this pattern, it was inferred that amplitude perturbation was a better indicator of hoarseness than frequency perturbation. Both amplitude perturbation and frequency perturbation could differentiate among extremely hoarse and moderately hoarse voices. However, only amplitude perturbation could significantly segregate between moderately hoarse and slightly hoarse voices. Also, this

showed that perturbation analyses were more closely related to auditory perception of hoarseness than to sound-spectrographic evaluation.

Among the drawbacks of this analysis was the inability of both the IGG measurements to differentiate between slightly hoarse and normal voices. This may be because the IGG samples were less than 0.5 s long, or because IGG reflects irregularity or roughness, while hoarseness consists of breathiness, turbulence or noise. Also, clear IGG signals were more difficult to obtain from women than men because of anatomical differences. The misclassification of subjects, according to Childers and Bae [12], may be due to several reasons. One of them was that there may not be a consistent perturbation pattern in the IGG. Instead, the vocal fold vibratory pattern might be stable for few cycles and then become unstable for cycles with no predictable pattern. The perceptual evaluation always assesses the entire data file of the subject, but the data analysis only examined the most stable portion of the data file, as determined by the visual inspection of the speech and IGG waveforms. Another factor was that for some subjects, especially patients undergoing therapy, the speech waveform might be stable and regular, but the IGG waveform might be irregular.

7.8 Automatic glottal inverse filtering

The flow of air through the glottis, the glottal volume-velocity (v-v) waveform, reflects the action of the vocal folds and is thus an important indicator of laryngeal function. The estimation of glottal v-v from acoustic speech signals has applications in areas of speech analysis, synthesis and in the study of laryngeal pathology.

The process of estimating the glottal v-v by removal of vocal tract resonances from speech signals is known as glottal inverse filtering. Veeneman and BeMent developed an automated on-line method to determine the glottal v-v waveform from normal and pathological speech based on digital inverse filtering [13]. The method aimed at accurate identification of vocal tract parameters and reduction of low-frequency noise.

The voiced speech production process may be modeled by the linear vocal tract transfer function $H(z)$ which filters the glottal v-v pulse train $G(z)$ to give the oral v-v $V(z)$, i.e.

$$V(z) = G(z)H(z)$$

The glottal v-v goes to zero during the closed phase of the glottis. So during this transient period $H(z)$ may be found uninfluenced by the glottal v-v. For the closed phase identification, Wong et al. suggested analysis of a normalized linear prediction error sequence obtained by calculation of the linear prediction error on an N-point window of the speech waveform. This method was sufficient for low frequency normal speech with a well-defined closed phase. But in cases of high frequency or breathy speech when the closed phase is of shorter duration, the method often suffered from ambiguous determination of local minima in the errors.

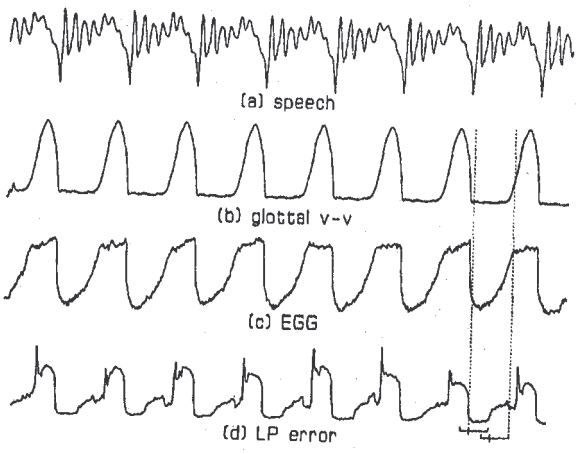

Fig.7.2(a): Example of normal voice. The vertical dotted lines denote the closed phase minimum range of error sequence. X-axis is time, Y-axis is arbitrary relative value [13]

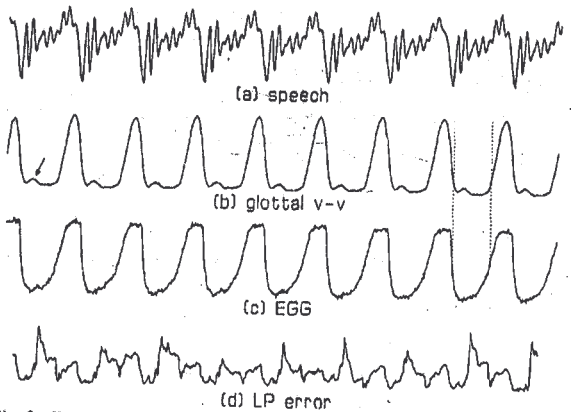

Fig. 7.2(b): Example of hoarse voice. Closed phase determination from error is ambiguous. Vertical dotted lines denote closed phase determined from 50% threshold of IGG [13]

Veneman and BeMent used an alternative method to determine the closed phase [13]. An IGG signal was sample simultaneously with the speech signal. Intervals of closed phase were determined by simple thresholding of the IGG at 50% of the peak amplitude. IGG thresholding for hoarse speech is illustrated by the dotted vertical lines in Fig.7.2 (b). Even though a well-defined region of zero glottal flow is not present, the interval shown by the vertical lines provides an analysis region of minimal glottal excitation that is not easily determined from the normalized linear prediction error sequence. Thus, by measuring glottal activity directly, the IGG provides a very robust means to determine a closed glottal phase region for the inverse filtering operation.

To summarize, the glottal v-v and IGG waveforms capture complementary features of the vocal fold motion. IGG provides information regarding the closed phase, and the latter about the open phase. The two taken together are superior to photographic means where information is lost in projecting the 3-dimensional glottis onto the horizontal plane. This method is relatively fast (3min/subject), non-invasive and automatic in the sense it requires no operator intervention once the speech and IGG signals have been obtained.

7.9 Applications to Teaching

The ability of the electroglottogram to instantaneously display the systematically changing pitch patterns has been used as the basis of a visual aid for teaching intonation and rhythm. These are crucial attributes of speech for conveying grammatical information and emotional attitudes to foreign students. According to Fourcin and Abberton, the visual display of the electroglottogram can be used in the speech training of the deaf in different ways [1]. The subject can learn to control the overall pitch of his voice by seeing his fundamental speech frequency contours (Fx) as well as the laryngograph output (Lx). Using the visual information from the Fx traces, the deaf speaker can quickly learn to produce simple intonation patterns that greatly improve the naturalness of his speech. The deaf learner can utilize the additional information about rhythm and tempo that the display provides. Since the laryngograph responds only when the vocal cords are vibrating, information about voiced or voiceless sounds can be presented.

8. HARDWARE OF THE IMPEDANCE GLOTTOGRAPH

8.1 Some commercially available glottographs

One of the first Impedance Glottographs was the "Laryngograph" or "Electrolaryngograph" introduced by the Laryngograph company. It featured circular electrodes having concentric rings that focused the sensitivity of the instrument within the neck, to reduce the noise level. Proper electrode position was determined by moving the electrodes on the neck during a prolonged vowel until a maximum IGG signal level was obtained. However, this laryngograph did not give useable waveforms for many subjects, especially women with considerable fatty tissue covering the larynx [14].

Another IGG introduced subsequently by F-J Electronics appeared to present similar problems, namely, excessive noise for some subjects and difficulty in monitoring the correct placement of the electrodes on the neck.

During the 1990-s, Glottal Enterprises developed a new IGG that employed a dual channel configuration which allowed the user to continuously monitor the location of the larynx with respect to the electrodes and provided the user an unambiguous front-panel indication of the proper electrode location. Since it was also designed with a high signal-to-noise ratio, this IGG provided a usable waveform for almost all subjects, with little or no visible random noise component [14].

8.2 Impedance glottograph developed at IIT Bombay

A block diagram of the impedance glottograph developed at IIT Bombay is shown in Fig.8.1. A sinusoidal oscillator produced a frequency of approx. 400 KHz. Wien bridge oscillator was selected due to circuit simplicity, frequency selection and amplitude stability.

This excitation is given as input to the impedance sensor, the voltage-to-current converter module. An attenuator buffer and potentiometer set the current level to be injected into the electrodes, followed by the current source. Bypass capacitors prevent the

passage of dc current through the electrodes, while resistor in the feedback path limits the dc gain of the circuit.

The electrodes are connected across the feedback path of the V/I converter circuit. Hence, a constant current flows through the two electrodes. 2 or 4 electrode arrangement can be used. In 2-electrode arrangement, both the electrodes are used for current injection and voltage sensing. There are different 4-electrode configurations possible. In one, both current injection and voltage sensing are done by center electrodes and the ring electrodes are kept floating. Another configuration is similar to the first except that the ring electrodes are grounded to reduce the common mode pick-up at the cost of some superficial current from center to ring electrodes, reducing sensitivity. In the third configuration, center electrodes are kept unchanged, while the ring electrodes are actively driven by buffers, which are in turn driven by respective center electrodes. This can reduce the common mode noise pick-up, provide better directivity to the excitation current and increase the sensitivity, but causing some additional current flow through the larynx between the two ring electrodes [4].

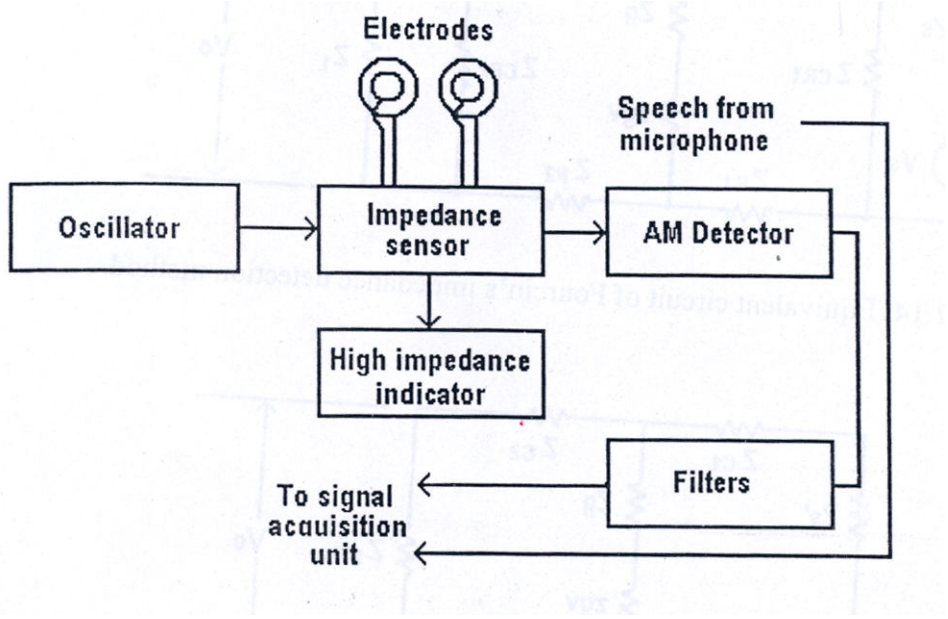

Fig. 8.1: Block diagram of the impedance glottogram developed by Patil, IITB, in 2000 [4]

A high impedance indicator module consisting of two comparators and two LED-s, one for each half-cycle, are used to verify proper contact at the skin-electrode interface. Both LED-s ON implies improper connection, both OFF implies proper connection while one ON signifies saturation of the corresponding half-cycle.

When the vocal folds vibrate, impedance between the vocal folds changes. Since the current is constant, the voltage across two electrodes changes in accordance with the impedance between the vocal folds and we get an amplitude-modulated voltage waveform. An instrumentation amplifier removes the common mode noise pick-up of this signal.

The signal is next given to the demodulator circuit, composed of a precision rectifier and a low pass filter that removes harmonics of the high frequency carrier.

The demodulated signal at the output of the low pass filter is then applied to the filter module which consists of a notch filter to remove 50 Hz noise pick-up, passive first order high pass filter to remove dc offset and fifth order elliptic low pass filter. The voltage output represents the impedance variation, and is known as the impedance glottogram.

The equivalent circuit for the impedance detection circuit used at IITB is as shown in Fig.8.2. The impedance between the electrodes is modeled as a fixed impedance Z_g in series with a time varying impedance Z_{gv}. Z_s is the source impedance. When the vocal folds vibrate, Z_{gv} varies and the output gets amplitude modulated. Thus,

$$V_o = \{(Z_g + Z_{gv}) \,||\, Z_s\} I_s$$
$$\cong (Z_s \,||\, Z_g)\{1 + Z_{gv} / Z_g\} I_s$$

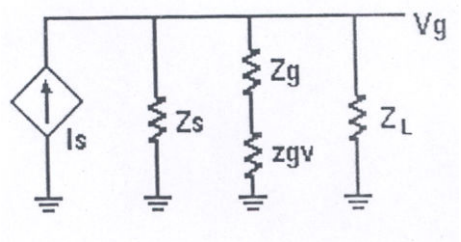

Fig. 8.2: Equivalent circuit based on current excitation, as developed at IITB [4]

The present set-up uses two Nickel-Hydride batteries 9V/12mAh. To make the instrument compact, a single battery-base power pack needs to be used. Short time stability of the oscillator needs to be improved to increase the SNR in the demodulator output. The current drain from each supply has to be greatly reduced from the present 80 mA. Study of electrode configuration needs to be undertaken to select the most appropriate one [4].

9. CONCLUSION

Glottographic techniques have been established as valuable tools for the assessment of vocal fold vibration, helping immensely in the diagnosis and documentation of laryngeal disorders. In comparison to other methods of glottography, impedance glottography allows for a better representation of the closed and closing phases of vocal fold movement, especially of the vertical contact area. Photoglottography seems particularly advantageous as far as the description of the open phase is concerned. The IGG is superior to all other methods in that it is not uncomfortable to speakers as it is completely non-invasive (it exerts no influence at all on the articulation and production of sounds).

At present, glottographs, especially impedance glottographs, are being used at many research laboratories, but except for rudimentary applications such as the measurement of vocal period, the technique has not been accepted for general clinical use, mainly because of the problems of extremely noisy waveform for certain subjects, difficulty in electrode placement, and inadequate charting of the various waveform features of interest to the clinician [4]. As a solution, multichannel techniques can be used to produce an IGG that

can verify the fidelity of its own output waveform as an indicator of the time patterning of vocal fold contact and can yield a signal that helps the user properly position the electrodes and/or track vertical movements of the larynx during voiced speech or singing. The use of improved IGG units incorporating such techniques should make possible a higher level of confidence in the use of impedance glottography in the study of voice production, in voice analysis, and in voice training.

10. ACKNOWLEDGEMENT

I am grateful to my respected guide, Prof. P.C. Pandey, whose active help and constant guidance contributed greatly to the successful completion of this seminar report. It was really a memorable experience to work under him.

I am also thankful to the members of the SPI Lab for their co-operation extended.

REFERENCES

[1] A.J. Fourcin and E. Abberton, "First applications of a new laryngograph," *Medical and Biological Illustration*, Vol 21, pp 172-182, July 1971.

[2] D.G. Childers and J.N. Larar, "Electroglottography for laryngeal function assessment and speech analysis," *IEEE Trans. Biomedical Engineering*, Vol BME-31, No. 12, pp 807-817, Dec 1984.

[3] I.R. Titze, B.H.Story, G.C. Burnett, J.F. Holzrichter, L.C. Ng, and W.A.Lea, "Comparison between electroglottography and electromagnetic glottography", *J. Acoust. Soc. Am.* 107(1), pp-581-588, Jan 2000.

[4] Anil Luthra , " An Impedance Glottograph , " *M. Tech. Dissertation* , Supervisor : Prof. P.C. Pandey, Dept. of Electrical Engineering, IIT Bombay, July 2004.

[5] D. O. Shaughnessy, *Speech Communications Human and Machine* , 2nd Ed. , Hyderabad : Universities Press (India), 2001.

[6] J.J. Jiang , S. Tang , M. Dalal , C. Wu , and D.G. Hanson, " Integrated Analyzer and Classifier of Glottographic Signals," *IEEE Transactions on Rehabilitation Engineering*, Vol. 6, No. 2, June 1998.

[7] F. D. Minifie , C.A. Kelsey, and T. J. Hixon, " Measurement of Vocal Fold Motion using an Ultrasonic Doppler Velocity Monitor ," *The Journal of the Acoustic Society of America*, Volume 43, No. 5, 1968.

[8] J. F. Holzrichter, L.C. Ng , G. J. Burke , N. J. Champagne II , J. S. Kallman, and R.M. Sharpe "Measurements of glottal structure dynamics," *Lawrence Livermore National Laboratory, University of California report, UCRL JRNL 147775.*

[9]"Description of the EGG waveform," – A tutorial by K. Marasek, http://www.ims.uni-stuttgart.de/phonetik/EGG/pagee2.htm , downloaded on 28 Sept 02.

[10] T. Baer, A. Lofqvist and N.S. McGarr, "Laryngeal vibrations: A comparison between high speed filming and glottographic techniques," *J. Acoust. Soc. Am.*, Vol. 73, No. 4, pp 1304-1308, Apr 1983.

[11] T. Haji , S. Horiguchi, T. Baer, and W.J. Gould, "Frequency and Amplitude Perturba-

tion Analysis of Electroglottograph during Sustained Phonation," *J. Acoust. Soc. Am.* 80(1), July 1986.

[12] D.G. Childers and Keun Sung Bae , " Detection of Laryngeal Function using Speech and Electroglottographic Data," *IEEE Transactions on Biomedical Engineering*, Vol. 39, No. 1, Jan 1992.

[13] D.E. Veeneman and S.L. BeMent, "Automatic Glottal Inverse Filtering from Speech and Electroglottographic Signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-33, No. 2, Apr 1985.

[14]"Electroglottographs, "http://www.glottal.com/electroglottograph.html, site of Glottal Enterprises (founded 1979 by Martin Rothenberg, PhD), downloaded on 11 Mar 03. Reg. Office : 1201 E. Fayette Street, Syracuse, New York 13210-1953, USA. E-mail : info@glottal .com

[15] A.J. Fourcin, "Laryngographic examination of vocal fold vibrations," *Ventillatory and Phonatory Control Systems, An International Symposium*, New York : Oxford University Press, 1974.

[16] A.J. Fourcin, "Laryngographic Assessment of Phonatory Functions," *ASHA Reports,* No.11, pp 116-127, 1981.