ACCELERATION SENSOR MEASUREMENTS OF SUBGLOTTAL SOUND PRESSURE FOR MODAL AND BREATHY PHONATION QUALITY

Wolfgang Wokurek¹, Manfred Pützer²

¹Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Deutschland
²Institut für Phonetik, Universität des Saarlandes, Saarbrücken, Deutschland

Abstract: We present a non-invasive attempt to indirectly measure the subglottal sound pressure. This quantity opens an additional acoustical path to observe the voiced sound source. The subglottal sound pressure contours of two phonation qualities, the modal phonation quality and the breathy phonation quality, are compared. The electroglottographic signal was recorded simultaneously as a well known reference basis for physiological details of voice production.

Keywords: acceleration sensor, subglottal sound pressure, phonation quality

I. Introduction

The subglottal sound pressure is of interest in order to study fine details of speech production, because the subglottal resonances may interact with voiced sound production, diphthongs in particular [1], [2], [3]. So far, every attempt to place a pressure transducer in the subglottal cavity resulted in an invasive method. That sort of methods was ruled out in advance for our study.

Placing a microphone as transducer at the fossa jugularis (suprasternal notch) was reported to yield good agreement of the transducer signal with direct subglottal pressure signal [4]. We decided for this study not to measure at the fossa jugularis to avoid a possible resonance influence from the chest that is known by audio engineers to be in the 1 kHz region. But the fossa jugularis is a place that is less covered by the cables of the electroglottogram electrodes and is likely to be investigated in the future.

The availability of micro electro mechanical systems (MEMS) to measure the acceleration, a quantity that is proportional to the force moving the sensor mass, led us to construct an external sensor to track the subglottal sound pressure at the skin of the neck. The sensor is gently pressed at the neck of the speaker in front of the cricothyroid ligament, located near the lower end of the larynx. The acceleration signals are recorded and chances may not be too bad that the tissue passes the subglottal pressure to the sensor. Due to the mass an the compressibility of the tissue only a (low pass) filtered version of the subglottal pressure may arrive at the sensor. Moreover we do not have a true reference signal of the subglottal pressure. But we do have the electroglottographic (EGG) signal as a phonatory reference. And when the EGG indicates a decreasing tissue contact of the vocal folds, we attribute the subglottal pressure to be the cause.

This acceleration sensor method was previously applied to measure the resonance parameters of the subglottal cavity [3]. Each of our studies is used to review and possibly improve the sensor and the evaluation procedure.

II. Method

A. Sensor

A three axis acceleration sensor is pressed gently against the skin of the neck and the sensor signals are recorded. The precise position at the neck is crucial. We identified the skin over the cricothyroid ligament as a potentially very good position to access the subglottal pressure. The cricothyroid ligament is a soft elastic tissue between the cricoid and the thyroid cartilage in the lower part of the larynx. It may be localized by touching the larynx and sensing for a soft gap in the elsewhere hard larynx structure.

Figure 1 shows the sensor in a test environment. The sensor currently is a cube of 1 cm edge length containing two ADXL202E two axis micro electro mechanical acceleration sensors. At one side of the cube a plastic nose of 5 mm height made out of hot glue is attached to improve the contact to the cricothyroid ligament. This nose is important to avoid loss
of contact during speech when the larynx moves up and down. The opposite side of the cube is glued to a balloon inflated with air to a diameter of about 6 cm. The upper frequency limit of the accelerations transducer is between 3 and 4 kHz.

The suspension by a hand held balloon yields more stable results than previous attempts to attach the sensor by a glue tape or by pressing the sensor to the neck by the second finger. The elastic suspension of the sensor together with the elastic and tissue (having its own mass) is a resonatory system by its own. And reasonable measurements of the subglottal pressure can not be expected in the frequency range of this resonance. The test environment shown in Figure 1 indicates a wide maximum between 100 and 200 Hz (similar to Figures 4 and 5). By a softer and less inflated ball the peak can be shifted to the 100 Hz region, but so far not further down and we currently interpete this fact as a limit caused by the elastic tissue and the mass of the sensor.

B. Signal processing

Each axis signals is lowpass filtered by an analog active filter to achieve the maximum bandwidth of 3.5 kHz specified by the data sheet of the sensor. As in a previous study [6] a 4th order Butterworth characteristic is implemented by two active Sallen-Key lowpass filter circuits. An additional 10 dB amplifier and a line buffer prepare each channel for cable transmission and for the level required by the soundcard.

During the recording session the sensor is slowly turned relative to the movement direction of the cricothyroid ligament by unconsciously changing the position of the hand holding the balloon and by vertical movements of the larynx. Hence no single axis signal shows the subglottal pressure contour. A principle component analysis of the vector signal uncovers the direction of the strongest oscillation, and the projection of the acceleration vector signal on this direction is considered as the subglottal pressure signal. It is labeled as main pressure component, or MPC signal.

The two sensors are orthogonally mounted at two sides of a cube and one axis of each points into the same direction, to and from the neck, where the main oscillation is expected. A simplifying assumption of this study is that the cube is only moved parallel and not turned by the vibrating skin of the neck. In this case the signals from the common axis should be basically the same. At the end of this study we discovered, that some of the recordings show different waveforms in that direction, contradicting the simplifying assumption. Hence, an advanced kinematic model should be added in further surveys, in order to transform the sensor acceleration measurements to the acceleration of the skin of the neck.

C. Speech material

The acceleration sensor signals and the EGG signal were recorded simultaneously for a single male speaker. Sustained vowels, nasals and diphthongs were uttered with two phonation qualities: modal and breathy. These phonation qualities are produced with different tensions of the vocal folds in the larynx. Breathy phonation has shorter and less complete closure phases and longer open phases compared with modal phonation. The cavity resonance oscillations are stronger damped when the vocal folds are open, due to the larger wall surface of the total coupled cavity [5]. Furthermore the center frequencies of the coupled cavity are slightly decreased.

III. Results

The excitation signal is compared to the electroglottographic signal for vowels uttered with modal and breathy phonation quality. In Fig. 2 (modal phona-
tion quality), the beginning of the closing phase of each pitch cycle is displayed as a steep ascent of the EGG contour. The ascent ends in the contact phase. The locally maximal contact is marked by the upper peak. With a short delay, the first cycle of the MPC signal starts. The opening phase increases the damping and slightly lowers the frequency.

In Fig. 3, breathy phonation quality is characterized by shorter and less complete closure phases and longer open phases. The longer open phase lets the oscillation amplitude of the main pressure component (MPC) signal descend much more compared to modal phonation. The shorter and less complete closure phase reduces the excitation and the amplitude of the first cycle of the MPC signal.

Both Figures 2 and 3 clearly show oscillations, but in spite of the elastically suspended sensor discussed in section II./A., a short term spectrum may increase insight. The MPC contours are quasi periodic. In order to ignore the associated harmonic structure of the spectrum, the analysis window is limited to the fundamental period of each signal. A kaiser window shape is selected to have control on the contrast between frequency resolution and spectral leakage. The window parameter $\alpha = 4$ suited both phonation qualities.

The modal recording has a lower fundamental frequency and an 8 ms window is the longest that is not dominated by a harmonic structure. Fig. 4 shows the magnitude spectrum from 0 to 2 kHz. This frequency range includes the regions of the first and second subglottal resonance of [500 Hz-700 Hz] and [1300 Hz-1500 Hz] respectively. It clearly shows a dominant component slightly above 200 Hz. Very likely it corresponds to the oscillation that is visible in the time domain in Fig. 2. There are peaks in the range of the first and second subglottal resonance, but we have no means to identify them.

The speech sample from the breathy voice quality recording has a higher fundamental frequency and re-
quires a shorter analysis window of 7 ms duration. Now the strongest component is centered slightly below 200 Hz. Again this is the excited and damped oscillation with a period of about 5 ms that is visible in the MPC part of Fig. 3. The subglottal resonance may be visible in Fig. 5 but is not identified.

IV. Discussion

One intention of this study was to obtain a detailed recording of the subglottal pressure up to 2 kHz by pressing an acceleration sensor to the skin of the neck. After the interpretation of modal and brathy vowel recordings in time and short term frequency domain we see oscillations that may be caused by a resonant structure consisting of the balloon suspension, the sensor and the skin and subcutan tissue. This structure has a center frequency around 200 Hz and a bandwidth in the 100 Hz range. Attempts to lower the resonance frequency of the suspension sensor system, did not substantially decrease the center frequency of the system including skin and subcutan tissue.

Otherwise, this unintended oscillation is driven by the main ‘force’ of interest, the voice source. This cycle of glottal movements and contacts introduces boundary conditions that influence the subglottal pressure contour that drives our oscillation as well as the time variant damping of that oscillation. Quantities related to the temporal open quotient and the damping that is related to the degree of opening between the vocal folds may be extracted by advanced signal processing.

From the point of causality a conservative summary might be the following: the sequence of opening and closing of the vocal folds is visible to a certain extent in the EGG waveform and our new MPC waveform does not contradict.

Finally, since the turning movements of the sensor do not proof to be neglectible, a kinematic model of the sensor is required to transform the acceleration sensor signals to the acceleration of the sensor nose at the neck.

V. Conclusion

The present study demonstrates that a subglottal sound pressure signal (the MPC signal) reveals the phonation physiology of modal and breathy phonation quality, similar to the electroglottographic signal. During different phases (closing phase, closed phase, opening phase, open phase) of the glottal cycle, the intensity of subglottal pressure changes due to a different contact status of the vocal folds. The physiologic differences cause corresponding changes in the amplitude, frequency, and damping of the oscillations in the MPC signal. These observations encourage a further look at other phonation qualities (e.g. hoarseness quality) for a better understanding of the representation of the healthy and pathological phonatory cycle in the MPC signal.

References