Acceleration Sensor Based Estimates of Subglottal Resonances: Short vs. Long Vowels

Wolfgang Wokurek, Andreas Madsack

Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Deutschland
wokurek@ims.uni-stuttgart.de, andreas.madsack@ims.uni-stuttgart.de

Abstract

The current version ACCV4 of our acceleration sensor device is presented and used to study the influence of vowel duration to the estimates of resonances of the subglottal system of 7 female and 9 male speakers. The sensor records movements in all three spatial directions below 5 kHz. It is gently pressed to the neck of the speaker in front of the cricothyroid ligament, a soft tissue in the lower part of the larynx. Linear prediction is used the estimate three resonances between 500 Hz and 2 kHz. Statistically significant differences in the estimates taken for the first two subglottal formants are found.

Index Terms: subglottal resonances, acceleration sensor, vowel duration

1. Introduction

We consider oscillations of the air in the subglottal system as a source of information of the voiced sound source. Subglottal resonances are of interest because they may interact with voiced sound production, diphthongs in particular [1], [2], [3], [4].

In this study the spatial vibrations of the current version ACCV4 of the acceleration sensor device are recorded simultaneously with the nasal and oral sound pressure signal captured separately through a Rothenberg mask. The technical development of the sensor is accompanied by recording and analyzing well known phonetic material. Short and long vowels are extracted from read sentences.

A tiny (around 4 Hz) but statistically significant influence of the vowel quality on the estimates of the subglottal formants were documented earlier [5]. The mechanism of this influence may be a length variation of the subglottal cavity by vertical shifts of the larynx caused by different tongue articulation positions. These positions are more central (schwa like) for short vowels compared to long vowels. Therefore the expectation is, that vowel duration has a significant influence on the subglottal formant frequencies.

Using our acceleration sensor device is a non-invasive procedure. The subglottal sound pressure is sensed only indirectly through body tissue. Consequently there is only a lot of educated hope to arrive at estimates of the subglottal resonances. Direct measurement of the subglottal pressure and the comparison to the waveforms recorded of our acceleration sensor device is still pending. So far we only inspected some of our acceleration waveforms with subglottal pressure waveforms that were recorded by Donald G. Miller. This informal comparison shows enough similarity to let us think we are not too far away from the subglottal pressure [6].

2. Acceleration Sensor

Preceeding versions of the acceleration sensor device ACCV4 were presented in [5], and [7]. Two major aims of the development are reducing the mass of the moving part as well as reducing the stiffness of its suspension to shift the mechanical resonance frequency of the device below the frequency range of the subglottal resonances. The third aim is tracking the spatial movement of the body tissue with a high bandwidth.

![Acceleration sensor device](image_url)

Figure 1: Acceleration sensor device

The acceleration sensor device is shown in Figure 1. The tip T is mounted at the acceleration sensor A. Both are held by spiral springs in the suspension ring S. The suspension ring is glued to the handle H that also contains electrical connectors to conduct the analog signals to an external preamplifier by a cable.

The acceleration sensor A consists of three ADXL202E two axis microelectromechanical acceleration sensors that are glued to different planes of an aluminium cube. The tip T is a plastic screw fixed at the cube by a counter nut. The electronic components are soldered to a flexible printed circuit board (PCB). The mechanical resonance of T and A elastically suspended only by the flexible PCB is in the 10 Hz range. This suspension alone would be too weak to press the sensor to the neck. Hence two spiral springs were added. Now the resonance is closer to 100 Hz and the force of the tip T to the neck is about 0.2 N – which proved to be strong enough to keep tissue contact but is hardly noticed by the speaker. The sensor resonance seems to be sufficiently low for this study not to disturb the measurement of the first subglottal resonance in the 600 Hz range. By using less stiff springs it can be further reduced to be appropriate for the F0 range in future studies.
The arrangement of the ADXL202E chips tracks the acceleration along each spatial direction at two different points of the cube. Hence, this six signals are sufficient to compute the spatial vibrations of the body tissue. In this study only the two sensors along the axis of T are evaluated. They are added to increase the signal to noise ratio of this main direction of tissue movement since an increase of the subglottal pressure pushes the tissue and the sensor out a bit.

3. Sensor Placement

The glottis is located in the larynx and separates the supraglottal from the subglottal cavity. It lies behind the thyroid cartilage. A soft tissue – the cricothyroid ligament – connects the lower end of the thyroid cartilage to the cricoid cartilage. Previous studies indicated that an acceleration sensor gently pressed at the neck at that position is moved at least in part by the subglottal pressure [5], [6], and [7]. There may well be additional sources of sensor movement: e.g. the vocal fold vibration conducted through the thyroid, arytenoid, and cricoid cartilage, respectively. One might also expect influence of the supraglottal pressure (from the vocal tract). But we didn’t find that vibration neither in the lower part of the thyroid cartilage nor at the cricothyroid ligament. We only found oscillations from the vocal tract sound pressure (i.e. the vowel formants) at some points in the upper region of the thyroid cartilage.

The cricothyroid ligament can be found by touching the larynx with a finger and searching for a small soft gap in the elsewhere hard larynx structure. The sensor tip T is placed perpendicular to the neck and pressed gently to the soft gap until the suspension ring S touches the skin as shown in Figure 2. Now the speaker is asked to speak. The correct placement of the sensor is immediately seen in the amplitude display of the six accelerator channels. The amplitude of the two channels corresponding to the tip axis rise to high levels, the other four stay at low levels. In many cases this situation holds for several minutes. Sometimes the perpendicular position of the tip to the neck is lost and the signal amplitude distributes over more than two channels. In that case the session is paused and the sensor is arranged correctly again. In our recordings usually the sensor device was held by an assistant.

Unfortunately the electroglottogram (EGG) had to be banned from the recording equipment, since for many speakers the neck band that holds the EGG electrodes either totally or partly covered the cricothyroid ligament or touched the tip of the acceleration sensor. So the contact between the tip and the neck was disabled or the tip movement disturbed. Since the EGG is a widely accepted method of displaying aspects of the glottal cycle that are related to the amount of contact of the vocal folds we investigate technical solutions to attach the EGG electrodes to the acceleration sensor device.

4. Spatial Analysis

To get an overview on the modes of vibration in the six dimensional sensor output a (6 × 6) correlation matrix of the vector sequence is computed together with its eigendecomposition. Looking at a small number of samples of some normal voices most eigenspectra have a single dominant eigenvalue that we interpret as a single spatial mode if vibration. This strongest eigenvalue is typically more than the factor of 30 larger than the second largest eigenvalue. It seems that certain voice disorders (e.g. uncompensated unilateral vocal fold paralysis) lead to additional modes of vibration reducing this factor e.g. to 5.

The speakers of this study have no voice disorders and the amplitude is concentrated at two channels. This situation corresponds to a movement of the sensor tip in a single spatial vibration mode. Separate recordings of speakers with voice disorders indicate that asymmetries in the vocal fold vibration raise the amplitude of the second spatial vibration mode of the sensor tip. In such cases the sensor could not be arranged to give high amplitudes in only two channels. One or two more channels displayed intermediate amplitude levels [8].

5. Recording and Speech Material

Speech sounds are recorded via two electret microphones mounted in the oral and nasal section of a Rothenberg mask. For this study both sounds are added and used for labelling the short and long vowels. The lower part of the mask is visible in Figure 2. The mask was held by the speaker.

The recordings were made in a room that is reasonably anechoic above 200 Hz and has a reverberation time of about 27 milliseconds. Eight channels were recorded simultaneously, six channels of the acceleration sensor as well as the oral and the nasal sound of the Rothenberg mask. The first order 5 kHz RC-lowpass recommended by the ADXL202E data sheet was implemented by analog hardware. All channels were digitized with a sampling rate of 48 kHz. The RME soundcard offers only AC coupling, hence no static acceleration signals like the gravitation vector are available as a direction reference in the evaluation.

The sustained vowel part and the sentence part of a slightly larger Corpus was recorded and used. About 45 minutes were recorded with each speaker. The recording contained three repetitions of the following material: the short text “Nordwind und Sonne”, sustained vowels and nasals and about 120 sentences like “Im Spreewald hat sich der Chemieunfall ereignet”. The long and short vowels were taken from accented words (e.g. the [e] from “Spreewald” and [i:] from “Chemieunfall”). To minimize the learning of sentence sequences the sentences appeared in a randomized order in each of the three parts of the recording session.

The following German short vowels [i, a, ò, o] and the long vowels [iː, aː, oː, uː] were used. All analyzed vowels were labelled manually.
The speakers produced their normal (usually modal) voice. The sounds are spoken three times by 7 female and 9 male speakers. The majority of 13 speakers are native speakers of German, 4 speak German as their second language.

6. Subglottal Formant Measurements

The aim of this procedure is the automatic estimation of the first and second subglottal resonances i.e. subglottal formants. The expected ranges of the subglottal resonances are [500 Hz-700 Hz] for the first and [1300 Hz-1500 Hz] for the second [1].

The six channels of the acceleration sensor device were reduced to a single channel by selecting the two that are aligned with the sensor tip either alone (channel2, channel4) or adding them (channel2+channel4). In a later study this simple approach will be compared to the eigenvector procedure described in section 4. The single acceleration signal is filtered with a 500 Hz linear phase (window design) highpass to suppress the fundamental frequency and the lower resonance modes of the sensor device. The subglottal formant parameters are estimated by linear prediction. To eliminate the large quantity of noise from the sensor above 5 kHz and to reduce computation time the sampling rate is converted to 10 kHz.

A standard correlation linear prediction is used to estimate the subglottal formant parameters: preemphasis with a zero at unity, order 10, window duration 20 ms, step size 1 ms. The small step size is used to ensure a large sample even for very short realizations of the short vowels.

The first three poles with center frequency above zero are used further. We arrived at this procedure after various heuristic attempts to assign poles to the subglottal resonances (e.g. predefined frequency intervals). The heuristics did not produce better (e.g. more normal distributed) results and are always questionable.

The poles are numbered with increasing center frequency. The first is taken as an estimate of the first subglottal formant. The second pole is assigned to a so called chest resonance. And the third pole is assigned to the second subglottal resonance.

The exclusion of the second pole as a chest resonance may be viewed as heuristic. However the following audio engineering background exists. Professional microphones intended to be attached near the speakers chest are compensated for a so-called chest resonance. The compensation is a dip of decreased transfer function at or near 1 kHz.

7. Statistics

Averaging all vowels of all speakers results in: sgF1= 586 ± 69 Hz, sgF2= 1325 ± 122 Hz, and Fchest= 929 ± 103 Hz. These sample means and standard deviations stem from channel 4 but closely resemble those of channel 2 and their sum.

The subglottal formants sgF1 and sgF2 are tested for statistically significant differences between short and long vowel. The comparisons are made for each individual separately because we are analyzing very fine phonetic detail that largely depends on the individual physiology, e.g. on the size of the subglottal cavities. All statistical computations were made with the “R” software [9].

Due to non-normal distributed samples the non-parametric Wilcoxon-Mann-Whitney-test is used in section 8. Each vowel quality is tested for each speaker separately and the tables report the number of speakers that show a significant difference in the tabulated category. For normal distributed variables a t-test is the method of choice. The Shapiro-Wilk test is applied to check for the normal distribution with a level of significance of 5%. Essentially all these tests resulted in a significant deviation from the normal distribution.

Figure 3 shows a typical histogram of the chest resonance of a single male speaker. The distribution is unimodal but there the similarity to a normal distributed variable ends.

As an example for the distribution of the subglottal formants that accompany all [a]- and [ax]-sounds of that speaker figure 4 shows the respective histograms.

8. Comparison

The Wilcoxon-Mann-Whitney-test is appropriate to statistically compare samples that differ from normal distribution.

Tables 1 and 3 give an overview on the large number of test results. They show for how many speakers the vowel duration led to a significant difference in the subglottal formants.

In table 1 the majority (about two thirds) of tests where significant at the 5% level. But how much differs sgF1 between short and long vowels? Table 2 shows the average differences of sgF1 for both groups of speakers without and with significant difference. Even the speakers with a significant sgF1 difference between short and long vowels differ only about 14 Hz at a resonance in the 700 Hz range (2%). The minority of speakers without significant vowel duration influence on sgF1 differ...
about 4 Hz which are about 0.5%.

The second subglottal formant seems to be less dependent on the vowel duration since less than half of all speakers have a significant influence. Furthermore the influence is weaker as table 4 shows. The differences are only slightly increased at a roughly double resonance frequency.

The columns of tables 1 and 3 demonstrate that there is only a minor influence whether one of the two single channels aligned with the sensor tip or their sum is the basis of the subglottal formant estimation.

### 9. Discussion

As there is a significant influence on the subglottal formants for roughly half of the speakers that influence is not very strong. A smaller number of sound realizations could likely have hidden the influence. A closer look to figure 4 demonstrates the results of tables 2 and 4 for the male speaker M07, one of the common examples. The longer vowel quality slightly decreases sgF1 and sgF2 for about 10 Hz. This is an order of magnitude similar to the influence of the vowel quality found on the second subglottal formant [5]. The lower subglottal formant for the long vowels could be caused by a raised larynx and thus, lengthened subglottal cavity. Other speakers show increased subglottal resonances, so a more detailed study will be necessary.

### 10. Conclusions

The current acceleration sensor device ACCV4 seems suitable as a transducer for non-invasive measurements of subglottal pressure waveforms in the frequency range of the first and second subglottal resonances. Compared to previous versions the mass of the moving part is lower and the suspension stiffness is reduced leading to a mechanical resonance frequency in the lower region of the fundamental frequency range that can be still lowered by less stiff springs. Due to the suspension ring and the handle the sensor may be placed on the neck sufficiently stable and it follows vertical larynx movements in the course of speaking.

The subglottal formant measurements on the filtered output signals of the accelerometer device conform with the

---

**Table 1:** Wilcoxon-Mann-Whitney test for the first subglottal resonance frequencies sgF1

<table>
<thead>
<tr>
<th>Vowels</th>
<th>Channel2</th>
<th>Ch.2+Ch.4</th>
<th>Channel4</th>
</tr>
</thead>
<tbody>
<tr>
<td>i / ɪ</td>
<td>11 / 17</td>
<td>14 / 17</td>
<td>14 / 17</td>
</tr>
<tr>
<td>a / ɑ</td>
<td>12 / 17</td>
<td>13 / 17</td>
<td>12 / 17</td>
</tr>
<tr>
<td>ɔ / ɔ:</td>
<td>15 / 17</td>
<td>13 / 17</td>
<td>12 / 17</td>
</tr>
<tr>
<td>u / u:</td>
<td>12 / 17</td>
<td>12 / 17</td>
<td>12 / 17</td>
</tr>
</tbody>
</table>

**Table 2:** Average differences in Hz between the first subglottal resonance frequencies sgF1 of short and long vowels

<table>
<thead>
<tr>
<th>Vowels</th>
<th>Channel2</th>
<th>Ch.2+Ch.4</th>
<th>Channel4</th>
</tr>
</thead>
<tbody>
<tr>
<td>i / ɪ</td>
<td>1.70 / 17.5</td>
<td>3.71 / 13.3</td>
<td>3.42 / 16.7</td>
</tr>
<tr>
<td>a / ɑ</td>
<td>2.02 / 13.9</td>
<td>3.75 / 13.5</td>
<td>3.98 / 14.8</td>
</tr>
<tr>
<td>ɔ / ɔ:</td>
<td>3.85 / 13.4</td>
<td>3.20 / 10.8</td>
<td>2.98 / 10.4</td>
</tr>
<tr>
<td>u / u:</td>
<td>4.30 / 13.9</td>
<td>6.93 / 12.5</td>
<td>9.05 / 16.3</td>
</tr>
</tbody>
</table>

**Table 3:** Wilcoxon-Mann-Whitney test for the second subglottal resonance frequencies sgF2

<table>
<thead>
<tr>
<th>Vowels</th>
<th>Channel2</th>
<th>Ch.2+Ch.4</th>
<th>Channel4</th>
</tr>
</thead>
<tbody>
<tr>
<td>i / ɪ</td>
<td>3.26 / 13.6</td>
<td>5.28 / 16.2</td>
<td>4.53 / 15.9</td>
</tr>
<tr>
<td>a / ɑ</td>
<td>3.49 / 11.4</td>
<td>4.74 / 16.5</td>
<td>2.93 / 16.8</td>
</tr>
<tr>
<td>ɔ / ɔ:</td>
<td>5.59 / 13.4</td>
<td>4.53 / 17.8</td>
<td>5.05 / 19.9</td>
</tr>
<tr>
<td>u / u:</td>
<td>5.43 / 18.3</td>
<td>8.37 / 29.9</td>
<td>8.21 / 23.4</td>
</tr>
</tbody>
</table>

**Table 4:** Average differences in Hz between the second subglottal resonance frequencies sgF2 of short and long vowels

known frequency ranges. Vowels duration exerts a statistical significant influence on the subglottal formant estimates.

### 11. Acknowledgements

The authors thank Dr. med. Dr. phil. Manfred Pützer (Universität Saarbrücken) for his kind cooperation in identifying the appropriate sensor position. Many thanks to Vitalia Voronik, Olga Podushko, Tamara Zulkashivili, Matthias Lenz, and Andreas Müller for the involving task of recording and labelling.

### 12. References