Influence of Speach Activity on Vibrometer Signals to Extract Vital Parameters of Humans

Kristian Kroschel[1], *Jürgen Metzler*[2]

[1]*Karlsruhe Institute of Technology (KIT),* [2]*Fraunhofer Institute IOSB*
kristian.kroschel@kit.edu

Abstract: For the measurement of vital parameters of the heart the elekrocardiogram or ECG is used. If a non-contact measurement is required the vibrocardiogram or VCG is more appropriate. During the measurement normally the client is quiet because she or he is realizing that the measurement is going on. This is not the case when the VCG is picked up because for comfort and to protect the eyes infrared lasers are used.

The VCG measures the speed of the surface of the skin invoked by the activity of the heart. Therefore clients are asked not to move since this would influence the result of the mesurement. It can be shown that also speaking influences the VCG significantly despite the fact that airborne sounds like the human utterance cover a different frequencey range than the heart activity. Obviously the structure-borne sounds emitted from the vocal chords influence the VCG.

In this paper this phenomenon is investigated in detail and a method is derived with which the utterance of a human can be detected based on a figure of reliability extracted from the VCG.

For the measurement of the heart beat frequency only those sections of the VCG are exploited in which the client does not speak.

1 Introduction

The gold standard to extract vital parameters of the heart activity of a client is the electrocardiogram or ECG [1]. The drawback of this method is that contact with the skin of the client at least at the arms and one leg is required. By this the client might become stressed and the measurement result is not reliable. Another reason to look for a method without contact is given in case of burnt clients or premature infants with a highly sensitive skin.

A non-contact method is based on a infrared laser Doppler vibrometer with the eye-friendly wavelength of $\lambda = 1550$ nm. From this signal the vibrocardiogram or VCG [2] is derived which may replace the ECG in the cases mentioned above. Furthermore the measurement of the heart function of a client might be executed without the knowledge of the client. Brushing the teeth in the morning in front of a mirror with an integrated device based on laser technology, the vital parameters might be measured and the result is sent to a central institution for further processing in case of abnormal findings.

In contrast to the ECG which is an electrical signal, the VCG is a signal measuring the speed of the surface of the skin by exploiting the Doppler phenomenon. Thus all mechanical effects like movement of the client, coughing etc. influence the VCG. Therefore the question arises if speaking of the client influences the VCG, too. At the first glance this will not be the case because the VCG is located in the frequency range of a few Hz whereas the utterance of a human is in the range of a few kHz except the male fundamental frequency around $f = 100$ Hz and of the female fundamental frequency around $f = 200$ Hz. But the movements of the vocal chords are not only the source of the sounds in the air but also of the structure-borne sounds of the tissue etc. Therefore the question arises whether or not the utterance of a client influences the VCG and how this influence can be detected using the VCG. Of course, the output of a microphone would tell whether the client speaks or not. But this requires an additional device which also might violate the privacy. Therefore the question will be discussed how the VCG can be used to decide whether the client speaks or is quiet.

2 System to extract the heart beat frequency from the vibrometer signal

The measuring system consists of the components shown in Figure 1. Despite the fact that all signals in the processing chaine are digital signals for which the discrete time parameter k is in common use, for all signals in the time domain the continuous parameter $t = k \cdot t_s$ is preferred to make the time dependence transparent. The sampling frequency of the raw signal is $f_s = 480$ Hz so that the distance between two samples is $t_s = 2.083$ ms.

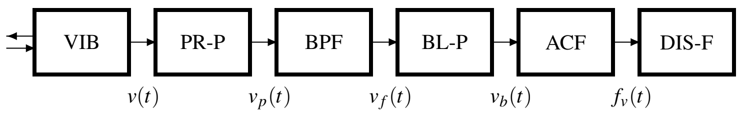

Figure 1 - Signal processing chain from the vibrometer to the display of the estimated heart beat frequency.

The vibrometer VIB emits the laser signal which is reflected by the surface of the skin of the client. The Doppler velocity $v(t)$ is extracted from the received signal and sampled. An example of such a signal is shown in Figure 2 on the left side. Obviously, it is not a stationary process because there is a significant change in amplitude at $t = 60$ s. A detailed analysis reveals that the signal below $t = 60$ s is periodic as can be seen from the section in the center of Figure 2 and looks random on the left plot in Figure 2. The reason is that at $t = 60$ sec the client starts to speak.

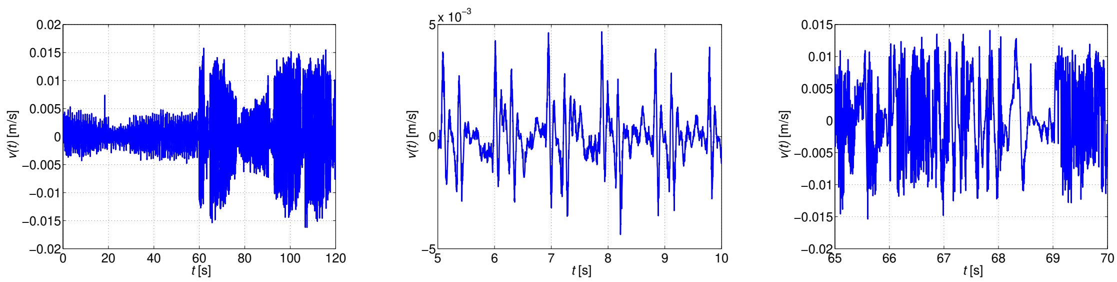

Figure 2 - Vibrometer signal. Total (left), left section (center), right section (right).

The different character of the signal in the first and second part is also visible in the frequency domain. The magnitude of the spectrum of the first half and the second half are given in Figure 3 on the left and right side, respectively. Due to the sampling frequency the bandwidth is limited to $f = 240$ Hz. The spectrum of the heart beat only is more or less limited to $f_u = 15$ Hz whereas the spectrum of the heart beat plus the utterance is much more broad-band and shows a higher intensity around $f = 100$ Hz. This is due to the fundamental frequency of the male speaker. But it is surprising that also in the range of the heart beat additional frequency component become visible which is a structure-borne sound caused by the utterance. The airborne sound covers a much higher frequency band around some kHz which is not picked up by the vibrometer signal. The spectrum of the heart beat has some dominant peaks. One of them might be associated with the heart beat frequency extracted from the data within a time interval of rouhgly $\Delta t = 60$

s. The most likely frequency would be around $f_h = 1.5$ Hz which corresponds with the lowest maximum. This has been extracted from a long time interval which is not sufficient for practical application. Instead, the frequency should be extracted from a time interval as short as possible and updated every second.

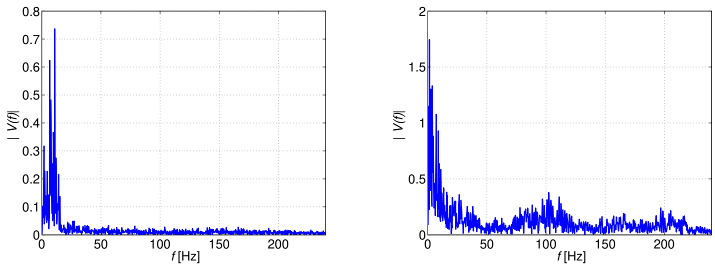

Figure 3 - Vibrometer signal spectra. Heart beat only (left), heart beat plus utterance (right).

In the pre-processing block PR-P the mean is suppressed because it does not carry useful information. Furthermore, the signal is downsampled by $M = 4$ to $f_s = 120$ Hz so that the usable bandwidth becomes $f = 60$ Hz which is sufficient to pick up not only the heart beat frequency of a relaxed client but also under physical or psychological stress plus details of the VCG structure. Because this structure like the length of the so-called p-wave is not in the focus of this investigation, the signal is filtered by a band-pass filter with the pass-band $0.582 \text{ Hz} \leq f \leq 15$ Hz which is given by block BPF in Figure 1. Frequencies below $f = 0.582$ Hz are suppressed because they originate from breathing. To avoid distortion, the filter is of the FIR type with linear phase, length $n = 512$ and a magnitude and impulse response as shown in Figure 4.

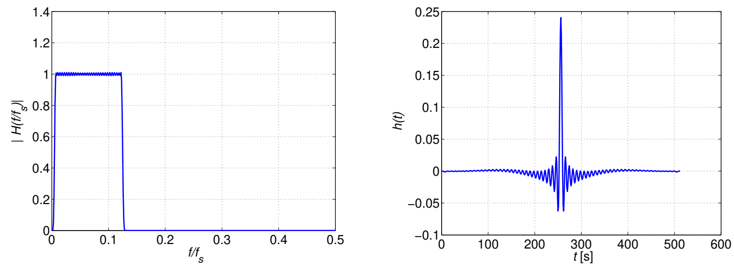

Figure 4 - Bandpass filter. Magnitude of trenfer function (left), impulse response (right).

As mentioned above, the heart beat frequency should be calculated from a data block as short as possible. Therefore the data stream is cut into data blocks of length $\Delta t = 5$ s which contain $N = 600$ samples since the sampling frequency is $f_s = 120$ Hz. Using the FFT [3] the frequency resolution is $\Delta f = 0.2$ Hz which is not sufficient for practical application. Therefore another method is used based on the unbiased estimete of the autocorrelation function [4], [5]

$$r_{VV}[\kappa, k] = \frac{1}{N - |\kappa|} \sum_{i=0}^{N-|\kappa|-1} (v_b[k+i] \cdot v_b[k+i+\kappa]), \quad 0 \leq \kappa \leq N-1 \tag{1}$$

with $N = 600$, κ the running index of the correlation function and k the time index of the data block which has to be multiplied by $t_s = 1/f_s = 1/120$ Hz $= 8.33$ ms. This calculation is executed in the block ACF. The data blocks are overlapping by $\Delta t = 4.25$ s or $N = 510$ samples so that every $t = 0.75$ s a new correlation function $r_{VV}[\kappa, k]$ is calculated. Two samples are given in Figure 5 for the first half and the second half of the data.
Since the heart beat is a periodic signal, the autocorrelation function is periodic, too, which is seen in the left part of Figure 5. In contrast the autocorrelation function on the right side

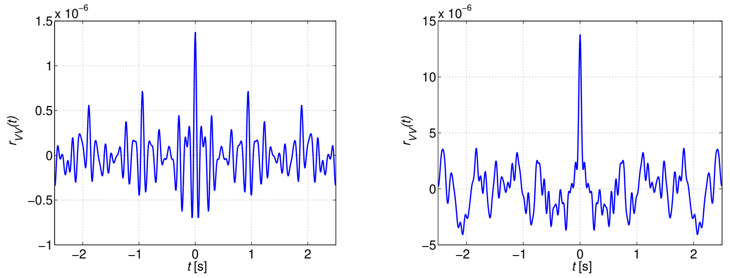

Figure 5 - Correlation function of one data block. First half of data (left), second half of data (right).

of Figure 5 has no distinct spikes except the one at $t = 0$. The reason is that speech as an information carrying process is close to a white process with a spike at $t = 0$ and a more or less constant value otherwise which corresponds with a flat spectrum. Therefore the heart beat frequency cannot be extracted from this correlation function.

In contrast, the autocorrelation function calculated from the heart beat signal is characterized by periodic spikes corresponding with the heart beat frequency. If the time instant of the maximum closest to the main maximum at $t = 0$ is $t = t_{max1}$ the heart beat frequency is

$$f_v(t) = \frac{1}{t_{max1}(t)}. \tag{2}$$

For every data block the frequency $f_v(t)$ is calculated. From the correlation function given on the left side of Figure 5 follows $t_{max1} = 0.926$ s which delivers with Equ. 2 the heart beat frequency $f_v = 1.08$ Hz or $f_v = 64.8$ 1/min. For all the available data, i.e. the section with the heart beat only and the section with the heart beat plus the utterance the result is shown in Figure 6 on the left side. In the first half the heart beat frequency is in the interval 62.65 1/min $\leq f_v \leq 67.35$ 1/min. In the second half not the heart beat rate is seen but a mixture of the heart beat rate and the stucture-borne sound component.

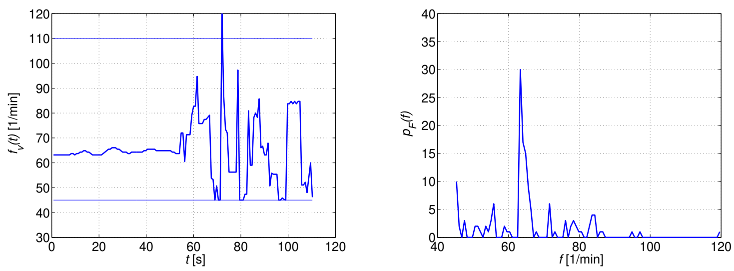

Figure 6 - Heart beat frequency $f_v(t)$ (left), histogram $p_F(f)$ (right).

For further analysis of the result given so far, the histogram $p_F(f)$ of the extracted heart beat frequency is calculated. It is depicted in Figure 6 on the right side. The main triangular section describes the variation of the heart beat frequency and corresponds with the interval mentioned above. The other values of the histogram $p_F(f)$ originate from the utterance. This opens a method to extract the heart beat frequency from measured data which contain sections with and without utterance as will be seen later.

3 Separation of sections with and without utterance

The idea behind the estimation of the heart beat frequency using the autocorrelation function is that the ideal autocorrelation function has periodical peaks of identical hight. Since only limited sections of the the VCG signal of length $\Delta t = 5$ s or $N = 600$ samples are used there is a decay of the amplitudes with increasing time. Therefore the relation of the amplitudes of the peaks with respect to the main peak at $t = 0$ is a measure of reliability whether the analysed signal is periodic or not. This figure of reliability is defined by

$$r_v(t) = \frac{r_{VV}(t_{max1}, t)}{r_{VV}(0, t)} \leq 1. \quad (3)$$

The closer $r_v(t)$ is to one the more reliable is the estimate of the heart beat frequency $f_v(t)$. The figure of reliability of the example in Figure 6 is given in Figure 7 on the left side by the solid lines.

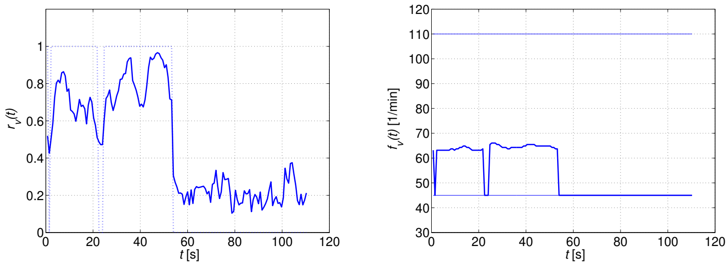

Figure 7 - Figure of reliability $r_v(t)$, sections $r_v(t) \geq 0.5$ marked (left), extracted heart beat frequency (right).

There is a significant change of reliability from the left part to the right part: whereas in the left part $r_v(t) \approx 0.7$ applies, the figure falls down to $r_v(t) \approx 0.2$ in the right part. Also in the left part where no utterance influences the VCG, there are fluctuations of $r_v(t)$ which might be caused be movements of the client.

The task is now to introduce a threshold t_r to separate $r_v(t)$ of high reliability from low reliability. A very simple solution is given by $t_r = 0.5$. A more advanced, data dependent threshold is given by

$$t_r = \frac{r_{v_{max}}(t) + r_{v_{min}}(t)}{2} \quad (4)$$

which is $t_r = 0.5351$ for the given data. An alternative would be the mean $\bar{r}_v(t) = 0.4692$. All these letter thresholds depend on the data which have to be observed in a given interval. In the case discussed here the interval is roughly $\Delta t = 110$ s long. Since all thresholds t_r are close together the fixed, data independent threshold $t_r = 0.5$ is the most attractive one. The reliability $r_v(t)$ is shown in Figure 7 together with the sections calculated with the fixed threshold $t_r = 0.5$ and marked by the dotted lines. Furthermore, the estimated heart beat frequency is given on the right side for all those data of the VCG signal with a reliability $r_v(t) \geq 0.5$.

In the section with utterance the reliability $r_v(t) < 0.5$ holds and thus no heart beat frequency is calculated and set to the minimum $f_v(t) = 45$ 1/min. This is not only true in the section with utterance but also in sections without utterance if the figure of reliability falls below $t_r = 0.5$. This might be caused by movements of the client or unknown influences like coughing. The calculation of the correlation function and the figure of reliabilty together with the extraction of the heart beat frequency is executed in block ACF in Figure 1. The result is displayed in block DIS-F.

The heart beat is never constant, even in a relaxed state of the client. This can be read from the histogram $p_F(f)$ in Figure 6 with a center part from roughly $f = 60$ 1/min to $f = 70.2$ 1/min of triangular shape. Of course these values depend on the state of the client. They will go up in case of physical or psychological stress and also the spread is not constant. Nevertheless it might be possible to reduce the influence of the utterance by adapting the band-pass filter to this central region. In the actual case the cut-off frequencies of the band-pass filter would be

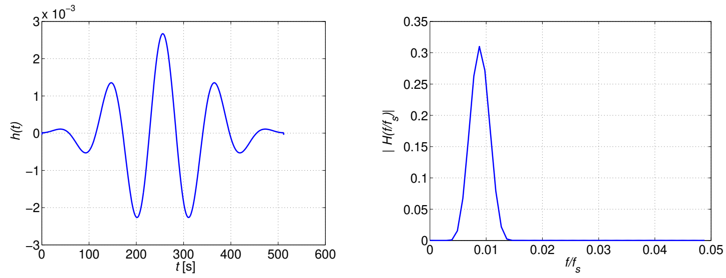

Figure 8 - Narrow-band band-pass filter. Impulse respone (left), magnitude of the transfer function (right).

1 Hz $\leq f \leq$ 1.17 Hz. The impulse response and the magnitude of the transfer function of the filter with this pass-band are given in Figure 8.
The correlation functions in the section with heart beat only and with heart beat plus utterance are given in Figure 9.

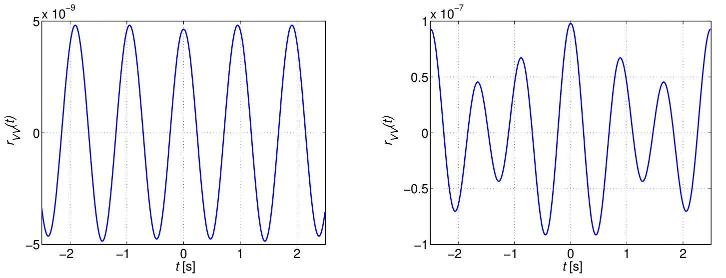

Figure 9 - Correlation functions of heart beat only (left), heart beat plus utterance (right).

The influence of the narrow-band filtering is clearly visible: the maxima of the correlation function calculated in the section with heart beat only are almost of the same hight and the peaks are much broader so that the calculation of the location of the maxima becomes less precise compared to the case of broader band width. This influences the calculation of the heart beat frequency based on the correlation function. The result is given in Figure 10 together with the figure of reliability $r_v(t)$.

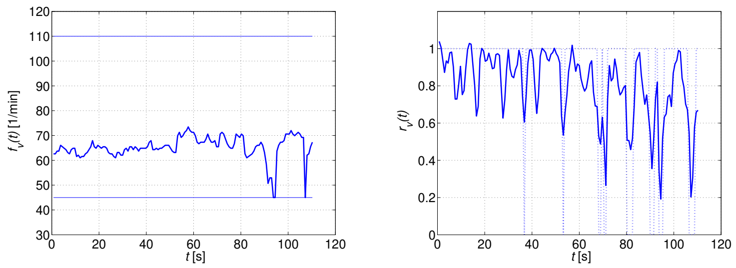

Figure 10 - Heart beat frequency (left), figure of reliability with sections for heart beat extraction (right).

As expected, the influence of the utterance on the calculation of the heart beat rate is dramatically reduced. The price for this advantage is the condition that the frequency range of the heart beat rate has to be known. The figure of reliability is significantly higher even if it is less in the region with utterance. The average value is $\bar{r}_v = 0.801$ in contrast to $\bar{r}_v = 0.469$ for the broad-band filter. Therefore it is not appropriate to choose the threshold $t_r = 0.5$ but $t_r = 0.615$ following from Equ. 4. The extracted heart beat frequency based on this treshold is shown in Figure 11 together with the resulting histogram.
In the sections where for the figure of reliability $r_v(t) < t_r = 0.615$ holds the heart beat frequency is set to $f_v(t) = 45$ 1/min. This is quite a bit arbitrary. The mean value would be more appropriate which might be extracted from past values averaged over the block length with

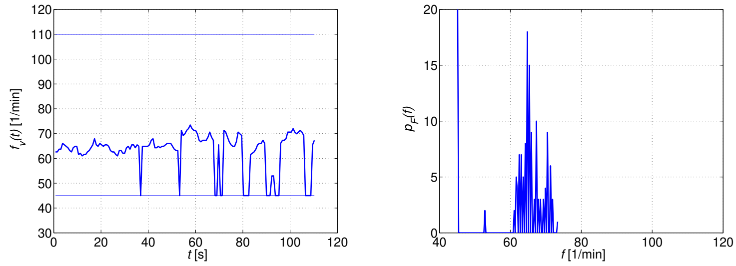

Figure 11 - Extracted heart beat frequency $f_v(t)$(left), resulting histogram $p_F(f)$ (right).

$N = 600$ samples. But this approach would generate jumps at the time instant where the figure of reliability $r_v(t)$ passes the threshold t_r.

4 An example

To test the procedure to extract sections of the VCG signal with heart beat activity only the sentence *The quick brown fox jumps over the lazy dog* with the duration of roughly $\Delta t = 5$ s, a similar sentence in German with length $\Delta t = 15$ s, the vowels with length $\Delta t = 5$ s and the numerals from zero to nine with length $\Delta t = 10$ s and a repetion uttered by a male speaker have been used for a test. The preprocessed and filtered signal together with speech pauses is shown in Figure 12 together with the extracted heart beat frequency.

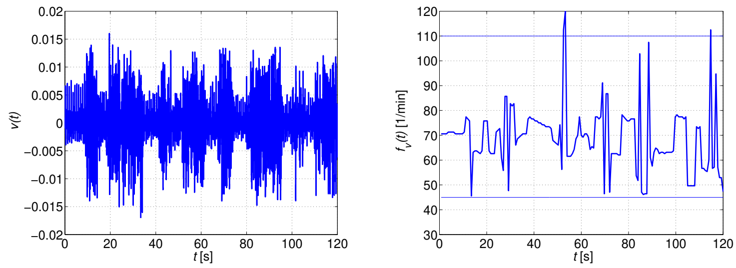

Figure 12 - A selection of speech signals (left) and the extracted heart beat frequency $f_v(t)$ (right).

The heart beat is much less powerful than the utterance. For an example using a section of length $\Delta t = 5$ s with heart beat alone and heart beat plus utterance, respectively, a distance of 10,25 dB has been calculated. It is obvious that the shown heart beat frequency cannot be used since it is not reliable enough. Therefore the figure of reliability has been calculated and the threshold $t_r = 0.5$ was used to separate reliable and unreliable sections from each other. Furthermore, the heart beat frequency has been extracted from the reliable sections. The result is shown in Figure 13.

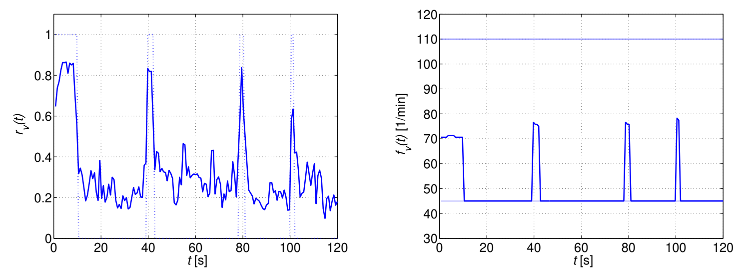

Figure 13 - Figure of reliability with marked reliable sections (left), extracted heart beat frequency $f_v(t)$ (right).

In fact, all speech pauses have been found. Obviously the speaker was so fast that no pauses within his utterance have been found. But the short pauses between the sentences and the vowels and numerals have been used to extract the heart beat frequency.

5 Summary and outlook

In all relevant publications on the application of vibrograms to extract vital parameter of the heart the clients are asked not to speak. This is acceptable since the measurements are executed in cooperation with the client. To avoid the influence of the clients attitute on this measurement she or he should not realize that a measurement goes on. But then it cannot be excluded that the client speaks and the utterance influences the measurement significantly. This can be overcome by excluding the sections with utterance. To detect time sections with utterance a figure of reliability is used. This is calculated within a data block of about five seconds duration. In case of significant differences of power between sections with heart beat alone and heart beat plus utterance the fixed threshold $t_r = 0.5$ can be used.

For atrial fibrillation similar heart beat rates are extracted [5] as with utterance. For discrimination between both cases the existence of the fundamental frequency of the speaker can be used.

Another method to reduce the influence of utterance is to narrow the bandwidth of the used band-pass filter. Here a threshold depending on the figure of reliability has to be used. The drawback of this approach is that an adaptation of the band-pass is required when the heart beat changes due to physical or psychological stress. Further investigations are required to optimize this approach.

All the measurements have been taken at the neck [6] of the client. Further measurements at the forhead, the wrist and the throat are available. At all locations the heart beat can be measured and at all of them the utterance influeneces the measurement. The best results [7] are found at the neck, the second-best at the throat followed by the forehead and wrist.

The project has been supported by a grant from the Germen Ministry of Education and Research under the contract 13N13725 and by Polytec from Waldbronn by supplying a vibrometer PDV100.

References

[1] JELIFFE, R.W.: *Fundamentals of Electrocardiography*. Heidelberg u.a.: Springer 1990

[2] TABATABAI, H., OLIVIER, D.E., ROHRBOUGH, J.W., PAPADOPULOS, C.: *Novel Applications of Laser Doppler Vibration Measurement to Medical Imaging*. Sens. Imagery 14, 2013, pp. 13-28

[3] KAMMEYER, K.-D., KROSCHEL, K.: *Digitale Signalverarbeitung*. Springer-Vieweg, Wiesbaden 2018

[4] KROSCHEL, K., RIGOLL, G., SCHULLER, B.: *Statistische Informationstechnik*. 5. Aufl., Springer, Heidelberg u.a., 2011

[5] KROSCHEL, K., LUIK, A.: *Laser-based remote measurement of vital parameters of the heart*. Proc of SPIE vol. 10680, Strasbourg 2018, pp. 1068000-2 - 1068000-8

[6] KROSCHEL, K., METZLER, J.: *Lokalisation des optimalen Messorts zur berührungslosen Bestimmung von Puls- und Atemfrequenzen*. 28. Konferenz Elektronische Sprachsignalverarbeitung, Saarbrücken 2017, pp. 300-387

[7] KROSCHEL, K., METZLER, J.: *Dependence of the Vibrocardiogram on the Measuring Point and the Influence of Speech Activity*. Internal report, Tricoder project, 2018