Erik Engelhardt, Adrian Zaman, Eric Elzenheimer, Norbert Frey, and Gerhard Schmidt*

Towards Analytically Computable Quality Classes for MCG Sensor Systems

https://doi.org/10.1515/cdbme-2022-1176

Abstract: While metrics such as SNR (signal-to-noise ratio) and ASC (application-specific capacity) can be used to describe signal quality quantitatively, they do not allow qualitative assessment. If only the SNR of a signal is given, it is not clear for which applications it is suitable. The subjective evaluation of signals by application experts, on the other hand, can provide such a qualitative assessment, but it is not reproducible. In this study, we investigate the relationship between these easily computable metrics such as SNR and ASC and the subjective evaluation of cardiologists using MCG signals as an example. To do this, we define four quality classes and generate a collection of noisy prototype signals. The cardiologists' assessments are then compared with the SNR and ASC of each signal. Due to the small amount of data, we only use a linear regression to predict quality classes based on common quantitative metrics. The achieved coefficients of determination are $R^2_{SNR} = 0.612$ and $R^2_{ASC} = 0.593$. In perspective, it is therefore plausible to analytically calculate the quality class of a MCG sensor system from the system characteristics with the support of a larger cohort of cardiologists. This will allow sensor system manufacturers to optimize their systems more efficiently for specific applications, enabling faster innovation loops at lower cost. For users of sensor systems, it will simplify the selection of the most suitable system for their application.

Keywords: mcg, ecg, sensor system characterization, sensor system performance, quality classes, biomagnetic sensors

***Corresponding author: Gerhard Schmidt,** Digital Signal Processing and System Theory, Institute of Electrical Engineering and Information Technology, Faculty of Engineering, Kiel University, Kaiserstr. 2, 24143 Kiel, Germany, gus@tf.uni-kiel.de
Adrian Zaman, Medical Department III specialised in Cardiology and Angiology, University Medical Center Schleswig-Holstein, Arnold-Heller-Str. 3 / House K3, 24105 Kiel, Germany, Adrian.Zaman@uksh.de
Erik Engelhardt, Eric Elzenheimer, Digital Signal Processing and System Theory, Institute of Electrical Engineering and Information Technology, Faculty of Engineering, Kiel University, Kaiserstr. 2, 24143 Kiel, Germany, {eren, ee}@tf.uni-kiel.de
Norbert Frey, Internal Medicine III: Heart, Vascular and Lung, University Medical Center Heidelberg, Im Neuenheimer Feld 410, 69120 Heidelberg, Germany, norbert.frey@med.uni-heidelberg.de

1 Introduction

The rapid development of biomagnetic sensors (e.g. [1–4]) has created the need for an intuitive sensor evaluation. While it is relatively easy to compare sensors using standard specifications such as sensitivity, power spectral density (PSD), and bandwidth, it is not trivial to find appropriate metrics to evaluate a sensor for specific applications such as magnetocardiography (MCG). Using quantitative metrics such as SNR (signal-to-noise ratio), or ASC (application-specific capacity) [5], it is possible to consider a sensor system within a biomedical application. However, it is still difficult to interpret the resulting values in a meaningful way. While we know that higher SNR/ASC values correspond to a better signal, it is not clear at what particular point a sensor system becomes clinically relevant. On the one hand, sensor manufacturers have difficulties assessing how to modify their product to make it more suitable for certain applications, and on the other hand, users find it challenging to choose the sensor system that is best suited for their application.

The simplest solution would be to show a signal recorded by a particular sensor system to an application expert, e.g., an experienced cardiologist in the case of MCG, who can then assess its suitability for the application. Unfortunately, this is usually not reproducible, as the assessment varies from cardiologist to cardiologist. Even the same person may assess the quality of the same signal differently at different times. Therefore, it is not practical to have an expert evaluate each iteration of a sensor system.

For this reason, we strive to create a model that combines the familiar and easy-to-measure metrics with the easier-to-understand but difficult-to-quantify expert opinions. For this purpose, we define four quality classes. In a survey, different cardiologists are then shown noisy MCG signals and asked to assign them to one of the quality classes. Finally, we adapt a linear model to predict quality classes from the power spectral density of signal and noise, eliminating the need to involve application experts for future sensors.

Although this approach is applicable to other biomagnetic measurements such as MEG (magnetoencephalography), we focus on MCG in this work. This is because the heart generates the strongest biomagnetic field in the human body [6] and therefore this particular application is suitable for a variety of magnetic field sensors.

2 Methods

Quality Classes

Following the, currently internal, International Telecommunication Union (ITU) definition for quality classes of telecommunication sensor systems, we define the following classes for the quality of MCG sensor systems:

1. A sensor system of quality **class one** can be used to perform all routine assessments without effort. For this, every detail of the measured signal must be clearly visible and free from noise.
2. A sensor system of quality **class two** can be used to perform most common clinical assessments, except those that rely on small details of the measured signal (e.g., the shape of the P wave or subtleties of the QRS complex and T wave).
3. A sensor system of quality **class three** can only be used for diagnosis based on very rough MCG morphology such as the periodicity and amplitude of the largest part of the measured signal (usually the R peak).
4. A sensor system of quality **class four** is not suitable for medical diagnosis, e.g., due to excessive noise in the measured signal.

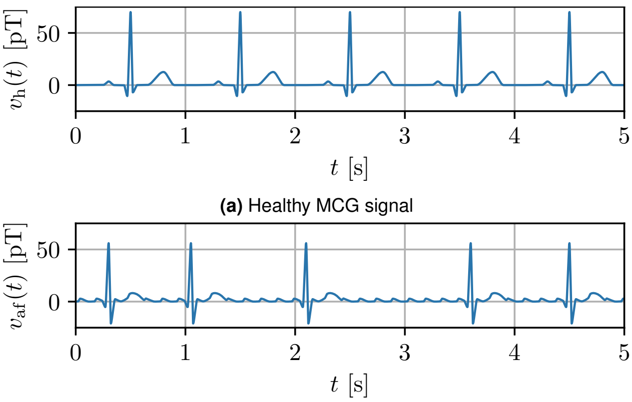
Parameters

The performance of a sensor system depends on three parameters. The desired input signal $v(t)$, the additive noise term caused by unwanted signals, and the post-processing steps applied. With an MCG sensor system, the desired signal may vary from patient to patient and depends on the sensor placement, but cannot be arbitrarily selected by the user. The primary noise sources are the earth's magnetic field, fields emanating from electrical equipment, and sensor noise itself. While the first two can be mitigated by using a magnetically shielded chamber, this is not possible with sensor noise. Finally, post-processing steps aimed at suppressing unwanted signals without affecting the desired signal can be chosen relatively freely. Limitations that may apply to post-processing steps are available processing power and real-time requirements.

This study aims to investigate the feasibility of calculating quality classes, rather than establishing a scheme that will work for all scenarios. Therefore, to reduce the complexity of the task, we only focus on the influence of the unwanted signal amplitude. The desired signals, the shape of the unwanted signal, and the post-processing steps remain fixed. To work in more realistic scenarios, further studies are needed to investigate the influence of the fixed parameters.

Prototype Signals

Since the quality classes are defined in terms of diagnostic value, we decided to use different MCG prototype signals for the survey. More specifically, we use one healthy or normal signal and four pathological signals. The healthy signal $v_\mathrm{h}(t)$ is based on a SQUID measurement performed at the Physikalisch-Technische-Bundesanstalt (PTB) in Berlin [5]. Because access to pathological MCG examples is limited, the shape of these signals is based on common ECG examples. The four pathological signals are $v_\mathrm{awi}(t)$ *anterior wall infarction*, $v_\mathrm{lbbb}(t)$ *left bundle branch block*, $v_\mathrm{af}(t)$ *atrial fibrillation*, and $v_\mathrm{lqt}(t)$ *long-QT syndrome*. The length of the signals is $T = 5\,\mathrm{s}$ and the sampling frequency $f_\mathrm{s} = 3200\,\mathrm{Hz}$. Two of the signals are shown as examples in Figure 1. See the GitHub repository[1] for the rest of the signals and more detailed information about them.



(a) Healthy MCG signal

(b) Atrial fibrillation MCG signal

Fig. 1: Examplary two of the five prototype signals used.

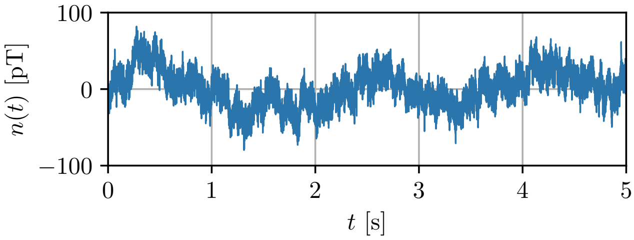
Noise

For simplicity, we use only one type of noise in this study. Namely, we use a $1/f$ or pink noise signal $\tilde{n}(t)$ since many magnetic field sensors exhibit this type of noise [7]. We apply the same noise sequence to all examples, varying only the noise power. The noise sequence used as well as its power spectral density are shown in Figure 2.

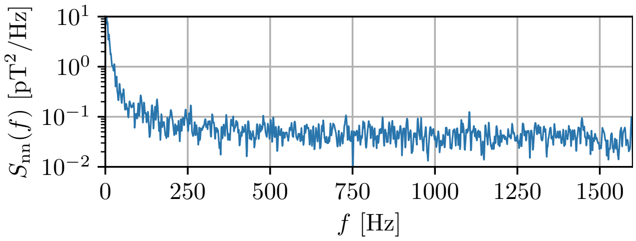
The measured sensor signals $u(t)$ are generated by adding a scaled version $n(t) = \alpha\tilde{n}(t)$, $\alpha \in \mathbb{R}^+$ of the noise term to one of the prototype signals $v(t)$:

$$u(t) = v(t) + n(t). \tag{1}$$

1 https://github.com/DenominatorIsZero/BMT-2022-MCG-Quality-Classes



(a) Time signal



(b) Power spectral density

Fig. 2: SNR 0 dB: $n(t)$ with same power as $v_\mathrm{h}(t)$.

Post-processing

The post-processing steps are based on the prevailing frequencies of the MCG signals and the usual noise sources. Since the magnetic field of the heart contains mainly frequencies between 1 Hz and 100 Hz [8], the measurement signal is first decimated to a sampling frequency of $f_{\mathrm{s},2} = 200\,\mathrm{Hz}$. Then, an FIR high-pass filter with $f_\mathrm{c} = 1\,\mathrm{Hz}$ is applied. In addition, a 50 Hz FIR bandstop filter is used. Since $n(t)$ does not contain an explicit 50 Hz component, this is not strictly necessary. In reality, however, electrical devices introduce a non-negligible amount of 50 Hz noise, so we decided to apply this common post-processing step in this study as well. Finally, to meet medical expectations, a constant is added so that the baseline of the signals is 0 pT. The resulting signal is called $s(t)$ in the following.

Metrics

The metrics computed for each signal are SNR (signal-to-noise ratio, cf. Eq. (2)), and ASC (application-specific capacity, cf. Eq. (3)) [5].

$$SNR = \frac{P_\mathrm{v}}{P_\mathrm{n}} = \frac{\int_0^\infty S_\mathrm{vv}(f)\,\mathrm{d}f}{\int_0^\infty S_\mathrm{nn}(f)\,\mathrm{d}f} \tag{2}$$

$$ASC = \int_0^\infty 10 \cdot \log_{10}\left(\frac{S_\mathrm{vv}(f) + S_\mathrm{nn}(f)}{S_\mathrm{nn}(f)}\right)\mathrm{d}f \tag{3}$$

While SNR compares the power of the noise and the signal after integration over frequency, ASC does so before integration. The result is that ASC is sensitive to spectral changes of the signal and the noise. In Elzenheimer et al. [5], we pointed out that this would allow a more relevant quantitative description of the sensor system signal.

Survey

For the survey, the five prototype signals are superimposed with noise at seven different power amplitudes, resulting in 35 combinations. The prototype signal $v(t)$ and the post-processed signal $s(t)$ are shown to cardiologists for each of the combinations. They are then asked to assign $s(t)$ to one of the quality classes based on their description above. The noise power are chosen to give the following SNRs for u_h: $SNR_\mathrm{h} \in \{-5, 0, 5, 10, 15, 20, 30\}\,\mathrm{dB}$, by varying α. The order in which the combinations are shown to all participants is random.

3 Results

The 35 signals were independently evaluated by $N = 4$ experienced cardiologists at the University Medical Center Schleswig-Holstein, Kiel.

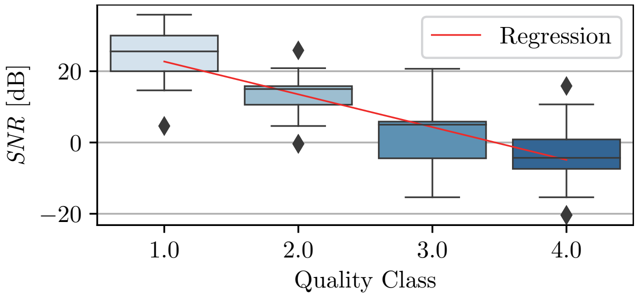
Of the 140 evaluated signals, 33 were classified as class one, 30 as class two, 38 as class three and 39 as class four. Figures 3a and 3b show the relationship of SNR and ASC to the assigned quality classes respectively. Across all signals, the average assigned class is $\overline{\mathrm{QC}} = 2.6$ with an average interobserver variability per signal of $\overline{\sigma^2}_\mathrm{QC} = 0.28$.

The signals assigned to better quality classes have, on average, a higher SNR and ASC. It can be seen that the classes can be clearly separated based on the medians. However, the ranges from the first to the third quartile overlap for classes three and four. Looking at the range from minimum to maximum SNR/ASC in the respective classes, it can be seen that there is no clear separation of the classes based solely on these characteristics.

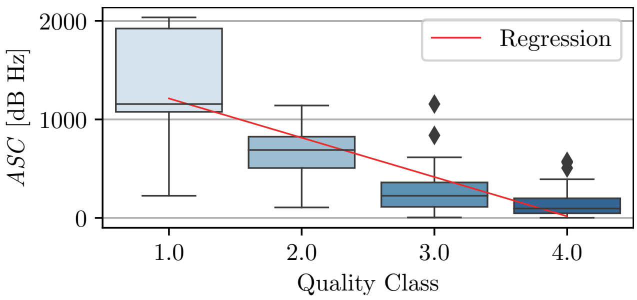
Using a least square regression, the quality class (QC) can be calculated as a linear function of the SNR (cf. Eq. (4)) or of the ASC (cf. Eq. (5)). The achieved coefficients of determination are $R^2_{SNR} = 0.612$ and $R^2_{ASC} = 0.593$. It remains uncertain if a linear function is the best fit for the relationship between SNR/ASC and quality classes.

$$\mathrm{QC}(SNR) = \begin{cases} 1, & \text{if } SNR \geq 22.45\,\mathrm{dB}, \\ 4, & \text{if } SNR \leq -4.81\,\mathrm{dB}, \\ 3.47 - 0.11 \cdot \frac{SNR}{\mathrm{dB}}, & \text{else.} \end{cases} \tag{4}$$

$$\mathrm{QC}(ASC) = \begin{cases} 1, & \text{if } ASC \geq 1216.8\,\mathrm{dB\,Hz}, \\ 4, & \text{if } ASC \leq 16.2\,\mathrm{dB\,Hz}, \\ 4.042 - 0.0025 \cdot \frac{ASC}{\mathrm{dB\,Hz}}, & \text{else.} \end{cases} \tag{5}$$



(a) SNR over assgined quality class with linear regression in red.



(b) ASC over assgined quality class with linear regression in red.

Fig. 3: Box plot showing the SNR (a) and ASC (b) of all signals grouped by their assigned Quality Class.

4 Conclusion

Due to the small number of participants, the results of this study are not sufficient to definitively clarify how subjective quality classes can be calculated using metrics such as SNR and ASC. However, the results suggest a clear interrelationship between those easily calculatable metrics and the subjectively assigned quality classes.

The possibility to determine the quality class of a signal analytically holds a lot of potential for sensor system manufacturers and users at the same time. Several steps are still necessary to achieve this goal. It is essential that the community agrees on a uniform class definition. The definitions used in this study are to be considered only as suggestions and certainly need detailed discussion on a broad level.

In addition, it has become evident that it is challenging to obtain a sufficient amount of data to develop a valid model. Although many parameters were not varied in this study, the cardiologists still had to evaluate 35 signals, which proved to be time-consuming. We intend to continue the analysis of MCG quality classes in future studies with more data and a closer look at all relevant parameters.

Author Statement

Research Funding: This work was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) through the Collaborative Research Center CRC 1261 Magnetoelectric Sensors: From Composite Materials to Biomagnetic Diagnostics. Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical Approval: The research involving human subjects complies with all relevant national regulations and institutional guidelines and was conducted in accordance with the principles of the Declaration of Helsinki. It was approved by the Ethics Committee of the Medical Faculty of Kiel University under Eudamed number CIV-20-04-032332.

References

[1] E. Elzenheimer, H. Laufs, W. Schulte-Mattler, and G. Schmidt, “Magnetic Measurement of Electrically Evoked Muscle Responses With Optically Pumped Magnetometers,” *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 28, no. 3, pp. 756–765, 2020.

[2] Y. Shirai, K. Hirao, T. Shibuya, S. Okawa, Y. Hasegawa, Y. Adachi, K. Sekihara, and S. Kawabata, “Magnetocardiography Using a Magnetoresistive Sensor Array,” *International heart journal*, vol. 60, no. 1, pp. 50–54, 2019.

[3] P. Hayes, M. Jovičević Klug, S. Toxværd, P. Durdaut, V. Schell, A. Teplyuk, D. Burdin, A. Winkler, R. Weser, Y. Fetisov, M. Höft, R. Knöchel, J. McCord, and E. Quandt, “Converse Magnetoelectric Composite Resonator for Sensing Small Magnetic Fields,” *Scientific Reports*, vol. 9, no. 1, p. 16355, Dec. 2019. [Online]. Available: http://www.nature.com/articles/s41598-019-52657-w

[4] P. Hayes, “Converse Magnetoelectric Resonators for Biomagnetic Field Sensing,” PhD Thesis, Kiel University, 2020. [Online]. Available: https://macau.uni-kiel.de/receive/macau_mods_00001263

[5] E. Elzenheimer, C. Bald, E. Engelhardt, J. Hoffmann, P. Hayes, J. Arbustini, A. Bahr, E. Quandt, M. Höft, and G. Schmidt, “Quantitative Evaluation for Magnetoelectric Sensor Systems in Biomagnetic Diagnostics,” *Sensors*, vol. 22, no. 3, p. 1018, Jan. 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/3/1018

[6] J. Reermann, P. Durdaut, S. Salzer, T. Demming, A. Piorra, E. Quandt, N. Frey, M. Höft, and G. Schmidt, “Evaluation of magnetoelectric sensor systems for cardiological applications,” *Measurement*, vol. 116, pp. 230–238, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S026322411730619X

[7] L. M. Ward and P. E. Greenwood, “1/f noise,” *Scholarpedia*, vol. 2, no. 12, p. 1537, 2007.

[8] A. M. Scher and A. C. Young, “Frequency Analysis of the Electrocardiogram,” *Circulation Research*, vol. 8, no. 2, pp. 344–346, Mar. 1960. [Online]. Available: https://www.ahajournals.org/doi/10.1161/01.RES.8.2.344