IMEKO 24th TC3, 14th TC5, 6th TC16 and 5th TC22 International Conference
11 – 13 October 2022, Cavtat-Dubrovnik, Croatia

FIRST STEPS TOWARD LEVERAGING ARTIFICIAL INTELLIGENCE FOR PRECISE CHARACTERISATION OF FORCE TRANSDUCERS

D. Mirian[1], R. Kumme[2], R. Tutsch[3]

Physikalisch-Technische Bundesanstalt, Bundesallee 100, 38116 Braunschweig, Germany
[1] davood.mirian@ptb.de, [2] rolf.kumme@ptb.de
TU Braunschweig, Institut für Produktionsmesstechnik, Braunschweig, Germany, [3] r.tutsch@tu-bs.de

Abstract:
This work is dedicated to the demonstration of a dynamic force measurement system for precise characterisation of the force transducers. The rocking motion of the system as a main dominant source of uncertainty in the acceleration is investigated. We propose a novel method based on the application of an artificial neural network for evaluation of the data as an alternative to traditional approaches to get low-uncertainty calibration measurements. In the end, two special architectures of the artificial neural network, namely Long Short-Term Memory LSTM and Gated Recurrent Network GRU are introduced, and their appropriateness for our use case is discussed.

Keywords: dynamic force calibration; rocking motion; measurement uncertainty; machine learning; deep learning; recurrent neural networks

1. INTRODUCTION

A comprehensive characterisation of a force transducer requires various investigations of sensors under static, continuous, and dynamic force excitation. Despite the well-established procedure for static calibration of force transducers described in ISO 376 [1], the dynamic calibration of the sensor has been always challenging because of some reasons such as the sophisticated nature of dynamic measurement, insufficient structural equipment, and parasitic effects which yield to higher measurement uncertainty in comparison to static measurements and therefore inappropriate characterisation of the force transducers. On the other hand, the growing need for performing precise force measurement under dynamic conditions in a wide range of applications from crash-test in the automobile industry to material testing machines and production processes has triggered enormous interest during the last years.

To fulfil these needs, up to now a lot of effort has been put into developing traceable methods to characterise force transducers under dynamic conditions by some National Metrology Institutes including Physikalisch-Technische Bundesanstalt PTB [2], [3], National Institute of Standards and Technology (NIST) [4], [5] and Centro Español de Metrología (CEM) [6]. However, a high measurement uncertainty caused by tilting of calibration assembly and relatively large variation in the determined model parameters, especially damping coefficient is reported by independent groups. As an alternative approach to other works three transfer functions including Autoregressive Moving Average eXogenous (ARMAX), Box-Jenkins (BJ) and Vector Auto-Regression (VAR), are successfully utilised for modelling of the signal in the time domain by PTB [7], which yields to a stable estimation of the dynamic parameters of force transducers.

On the other hand, with the advancement in computational power of computers and more importantly development of deep learning-based algorithms in the last few years, Artificial Neural Networks ANN as powerful models have been used in a variety of problems such as time series forecasting, speech recognition, language modelling, image captioning, and many other applications. Among different deep learning-based algorithms, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) have achieved excellent performance on difficult learning tasks and demonstrated their superiority in precision and accuracy in many disciplines. We address whether similar approaches can be used to solve complex problems in the characterisation of force transducers.

In this paper, special attention is paid to the rocking (tilt) motion of the calibration system as the main dominant source of uncertainty in the acceleration and hence force. A dynamic force calibration setup is introduced based on the use of laser interferometry for performing precise acceleration measurement on both the top and the bottom of the transducer. The novel method proposed for the evaluation of measurement data in this work should pave the way to using Artificial Intelligence (AI) in dynamic force metrology.

2. CALIBRATION SETUP

A schematic demonstration of the calibration setup is provided in Figure 1. The force transducer under test is mounted on a circular metal plate that is attached to the table of an electrodynamic shaker. A calibrated load mass is fixed to the top of the transducer to generate the dynamic force. A Polytec PSV-400 scanning vibrometer is used to sequentially measure the acceleration at 16 points on the circular plate and at 24 points of the load mass (see Figure 2). The acceleration measured by the scanning vibrometer is traceable to the unit of length.

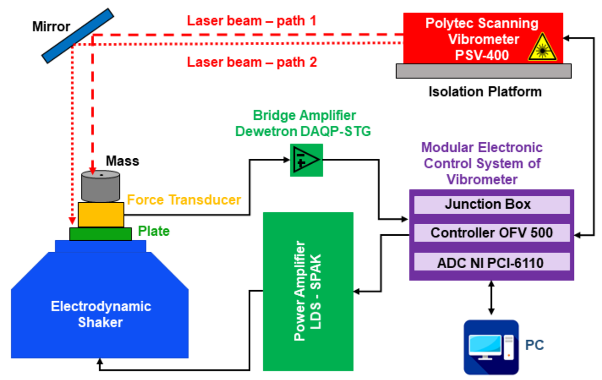

Figure 1: Schematic presentation of the measurement setup

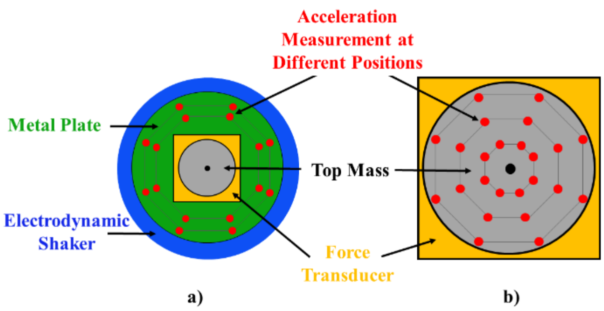

Figure 2: Setup top view – Acceleration measurement at different points a) on the plate all around the transducer b) on the top mass

The scanning vibrometer is located on an isolation platform to prevent vibration of the interferometer. The acceleration measurements at several positions all around the force transducer give a comprehensive understanding of the frequency-dependent tilting movements of the shaker while it is exited with a periodic chirp signal generated by the Polytec control system and amplified by the power amplifier LDS – SPAK. Measurement data obtained at different positions on the surface provide a dataset that can be used to feed into the artificial neural networks ANN and train them.

In the same way piston and tilt movements of the inertial load mass are measured. Because our scanning vibrometer is not able to perform measurement on two surfaces of different heights simultaneously, two independent measurement sets are carried out directly after each other under the assumption that the shaker generates the same movements in both measurement sets.

3. MEASUREMENT PROCEDURE AND DATA GENERATION

As an alternative to sinusoidal excitation, a periodic chirp signal is used to measure the frequency response of the system same as in [7]. This enables us to perform high-speed measurements with high spectral resolution rather than recording each frequency point individually or using a sweep signal. Furthermore, due to the periodic nature of the drive signal with period T, the recorded signals only contain an integer number of periods. In other words, it contains only frequency components at harmonics of $1/T$ and therefore fit on the FFT grid. Therefore, spectral leakage is prevented when the data are transformed to the frequency domain. The signal is sampled with a frequency rate of 25.6 kHz and a resolution of 39.06 µs. The acquisition time is 0.16 s for the generating 4096 sample points. The measurement settings are set in the same way in both measurement sets and therefore all collected data have the same dimension whereby a lot of data pre-processing before training the neural network is prevented. Figure 3 illustrates the recorded signal by the laser interferometer for one point on the top mass in the time domain.

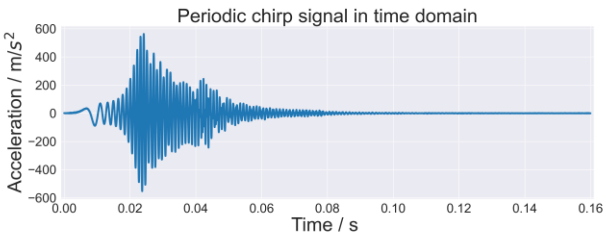

Figure 3: Acceleration measured by laser scanning vibrometer at one point on the top mass

Applying a fast Fourier transform (FFT) to the recorded signal yields the frequency response of the system shown in Figure 4.

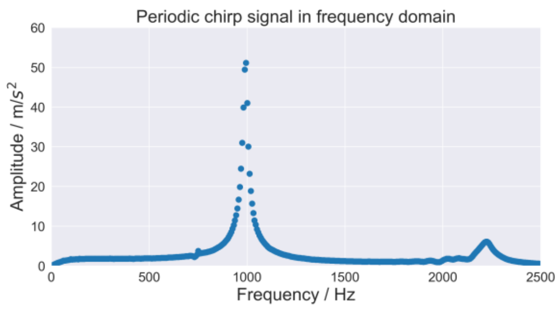

Figure 4: Fast Fourier representation of the recorded acceleration signal in the frequency domain

According to DKD 3-10 [8], the acceleration is measured by a one-point laser vibrometer and piezoelectric transducer on the top and bottom respectively. To determine the damping and the stiffness of the force transducer, the ratio of the measured accelerations on the top mass and the bottom must be created and a Lorentzian function must be fitted to the resonance curve in order to determine the resonance frequency and full width at half maximum FWHM.

Besides more accuracy and agility, the measurement procedure introduced in this work gives a great opportunity to investigate the rocking movement of the calibration setup at each point of the top mass surface and on the plate, in comparison with DKD 3-10 [8]. Therefore, we conduct a point-by-point investigation of each surface (top mass and plate) before creating the ratio.

First, in order to find the optimum number of the resonance peak, for measurement data in the frequency domain and for just one point on the top mass, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are used for the model selection among a set of candidate models (number of peaks) which can be used in the fitting model.

As shown in Figure 5, both criteria exhibit the first minimum at three components and a further increase in the number of components (Lorentzian peaks) would result in overfitting of the model to data.

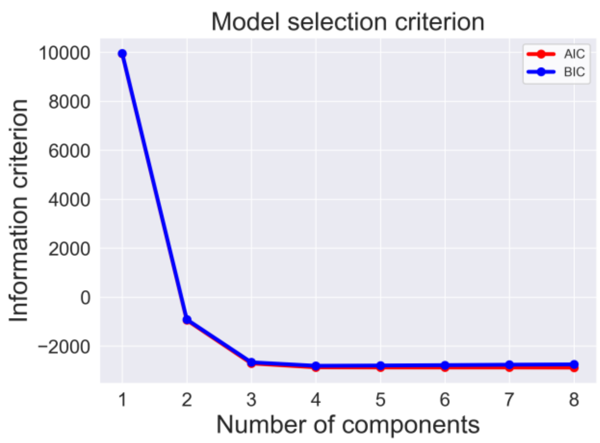

Figure 5: AIC and BIC model selection criterions

Figure 6 demonstrates the best fitting result. The fitted curve describes the measurement data as a linear sum of three Lorentzian functions with different parameters.

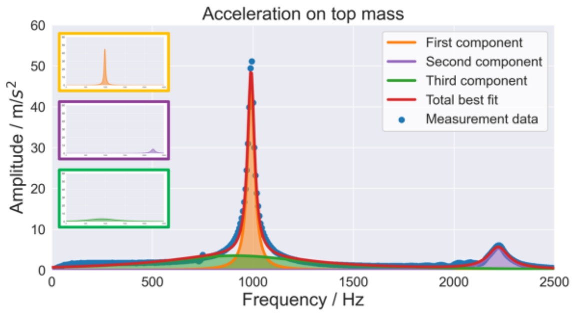

Figure 6: Best fit results as a linear mixture of three Lorentzian functions

While the first peak (shown in orange) and the second peak (shown in purple) represent the resonance of the setup and shaker respectively, the third peak (wide and small green peak) can be put down to a parasitic effect or another degree of freedom. However, in further investigation, just the first peak is characterised by its FWHM and height for all points on the top mass by performing the same method. Figure 7 provides a comparison between the FWHM and the height of the peaks. As these two parameters have different distributions, in order to make a meaningful comparison the values are normalised to their maximum.

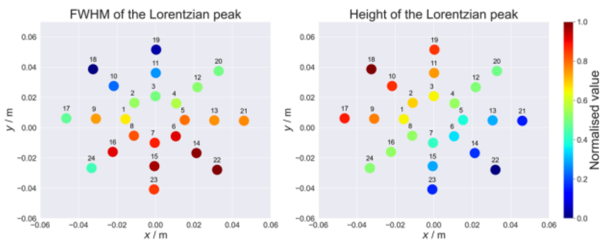

Figure 7: Comparison of the FWHM and height of Lorentzian function at different points on the top mass

Although one can easily distinguish between two areas (see Figure 8) on the top mass where the fitting model outputs in one area, wide and short peaks (marked by 1), and in the other one, narrow but tall peaks (marked by 2), averaging the acceleration over all the points on top mass seems to be a bad idea because of the asymmetry and the dissimilarity observed here.

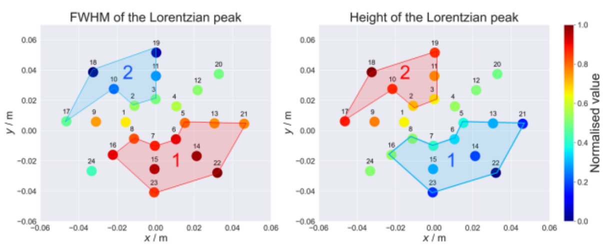

Figure 8: Two distinguished areas on the top mass

Above mentioned investigations are also important for performing comparable acceleration measurements by a piezoelectric transducer where the measurement is restricted to just one point. This means the positioning of the transducer in the middle of the surface would not necessarily yield accurate results and can be challenging.

In the second set of evaluations, the same approach is used for acceleration on the plate. As shown in Figure 9 finding any clear pattern that describes the acceleration at different positions on the plate is not simple. Besides the rocking movement of the shaker, this sophisticated behaviour can arise from fixing components of the metal plate (screws) and different material lattice vibrations at each point.

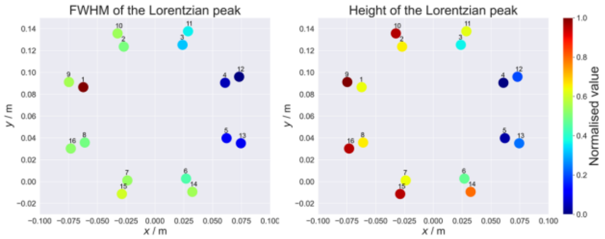

Figure 9: Comparison of the FWHM and height of Lorentzian function recorded at different points on the metal plate

The main idea that motivates this work is the utilisation of the ability of artificial neural networks to detect anomalies in signals and the arithmetically filtering of them after measurements. The developing calibration procedure would serve as a new sample-free method, applicable in all calibration facilities without the need to use an air bearing system and any special mechanical adapter and hence with a simple calibration setup.

4. ARTIFICIAL NEURAL NETWORKS

ANNs have been around for quite a while but in recent years we are witnessing another wave of interest and they have been brought into some metrology fields. They frequently outperform other Machine Learning ML techniques when it comes to complex problems due to their deep and nonlinear character. The basic idea of using ML is to infer knowledge from the data without explicitly programming. In the simplest form, ANNs have three layers, input, hidden, and output layer. In practice, they contain multiple non-linear hidden layers which enables them to learn complicated relationship between their inputs and outputs. Figure 10 illustrates a simple fully connected neural network with three hidden layers. The network can be arbitrarily deep by adding more layers.

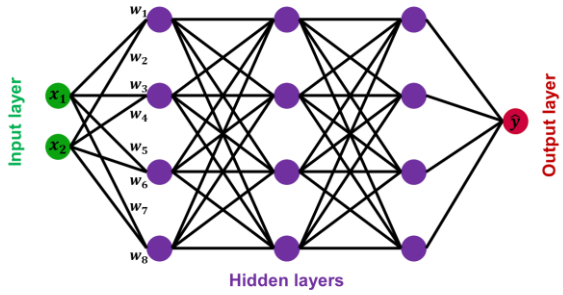

Figure 10: A simple fully connected neural network

The output of the network is calculated from equation (1).

$$\hat{y}_{W,b}(X) = \sigma(WX + b) \quad (1)$$

where X is the matrix of input features, W matrix of weights, b bias vector, and σ is a nonlinear activation function (e.g., sigmoid or tangent hyperbolic (tanh)).

The training of a network consists of three major steps namely forward pass, loss calculation, and back-propagation. In the first step, input data are passed to input nodes, these nodes are connected to the first hidden layer via connections. Each connection is associated with a weight. These weights are initialised by the network in the first step and optimised during the training. The output of each hidden layer is passed as input to the next layer until the information reach the output layer and the network makes a prediction $\hat{y}$. In the second step, the prediction is compared to the ground truth y using a loss function (e.g. Mean Squared Error MSE). The loss function outputs an error value that evaluates the network performance. Finally, starting from the output layer toward hidden layers the gradient for each node is calculated using that error, and the network finds the best weight values after several repetitions.

The choice of an appropriate network architecture which at least theoretically can use all relevant information to develop a model is the first challenge that one is faced with when it comes to utilising the ANN. The following sections provide a glimpse of neural network layers selected for this work while, to keep the discussion as simple as possible, it does not dig into the sophisticated underlying mathematics.

4.1. Recurrent Neural Networks RNN

The acceleration values in the time domain measured by the scanning vibrometer over the time $\boldsymbol{T}$ can be considered as sequence data, as given in equation (2).

$$x = (x_1, x_2, x_3, \dots, x_{T-1}, x_T) \quad (2)$$

Given time sequences as training data, the aim is to learn the rules to predict the output data given the input data. Among different possible neural network layers, Recurrent Neural Networks RNN are selected which are good at modelling the temporal sequence of the data due to their chain-like structure, shown in Figure 11.

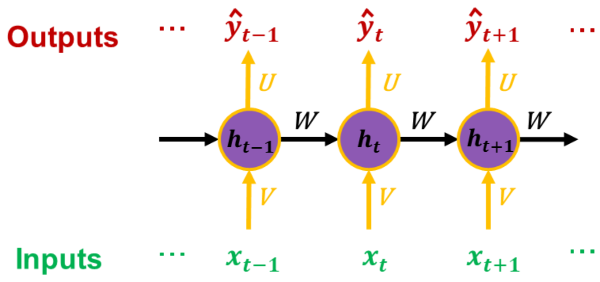

Figure 11: Unfolded Recurrent Neural Network RNN

Unlike traditional neural networks also known as feed-forward neural networks, a looping mechanism in hidden layers makes possible data transformation from one time step t to the next step $t+1$. By iterating equations (3) and (4) from $t=1$ to $t=T$, the hidden and output vector sequences are calculated.

$$h_t = \sigma(Vx_t + Wh_{t-1} + b_h) \tag{3}$$

$$\hat{y_t} = \sigma\left(Uh_t + b_y\right) \tag{4}$$

where V, W, and U are the weights matrices, and b is the bias vector.

The hidden states are a function of all previous hidden states. However, by increasing the steps processed by the RNN, the quantity of the information that can be retained from previous steps decreases dramatically. This shortcoming also known as short-term memory is caused by the "vanishing gradient" during back-propagation [9]. The gradient is a derivative of the loss function with respect to the weights and serves as a measure to tweak and adjust the weights. That means the value of the gradient defines the level of the adjustment for each node when gradients are back-propagated through layers and also through time. When the gradient becomes smaller and smaller, the weight updates become insignificant for further layers which in the worst case means no learning is done.

4.2. Long Short-Term Memory LSTM

Long short-term memory LSTM is one variant of the RNN networks which addresses short-term memory problem and aims to mitigate this issue using gating mechanisms [10]. The information is transmitted through the LSTM cells-chain via the so-called "cell state" Ct that resembles a conveyor belt running parallel to the sequence that is processing. Information can be optionally added or removed by two gates known as the input gate and the forget gate respectively. The data flow through LSTM cells is carefully controlled by the sigmoid neural layers which squash values between 0 and 1. This enables the cells to optionally let data pass through or dispose. The calculated value in each LSTM cell is based on the cell state along with the filtered and newly added data. Finally, the output gate decides which data should be sent to the next cell. On one hand, the capability of the LSTMs to learn long-term dependencies, and on the other hand the ability to delete irrelevant information, qualifies them for our modelling problem.

4.3. Gated Recurrent Unit GRU

Same as LSTM neural network, GRU implements a gating mechanism to eliminate the vanishing gradient [11]. It can be seen as an LSTM with a simplified design by which just two gates control the data flow. While the update gate can be trained to keep information from time steps long ago and determines how much of them needs to be passed to the future, the reset gate decides to remove information which are irrelevant. GRUs have shown their superiority over LSTM on some tasks in terms of speed and generalisation [12] and therefore it is chosen for this work to compare its performance.

4.4. Bidirectional RNN

Since in the signals recorded by scanning vibrometer, for a given time step the whole data also in the future are available, there is no reason not to exploit future values as well. BRNN is an advanced RNN that can be trained simultaneously in the positive and negative time directions and use all available input information [13]. This makes the model more powerful and robust against anomalies and outliers. BRNN processes forward sequence and backward sequence by two separate hidden layers which are fed to the same output layer, as shown in Figure 12.

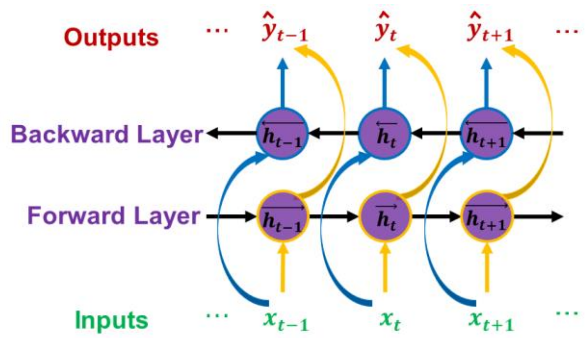

Figure 12: Bidirectional Recurrent Neural Network

Combining BRNNs with other RNNs such as LSTM and GRU enables us to access long-range data in both input directions.

5. IMPLEMENTATION AND MEASUREMENT UNCERTAINTY

All different neural networks mentioned in this paper are available as build-in layers in the Tensorflow developed by google. A high-level API namely Keras facilitates simple implantation of the neural networks in Tensorflow. A common-sense approach (non-deep learning baseline) will be used to demonstrate the superiority of the black-box deep learning-based method. The average acceleration values for every surface (top mass and metal plate) are used as a reference. The relative deviation is calculated by taking the Root Mean Squared Error (RMSE) metric. This yields a good understanding of the anomaly distribution along with the investigation performed in this work. We will conduct an empirical investigation and evaluate each technique mentioned in this paper on the task of the advanced force transducer modelling.

The black-box nature of the deep-learning method introduces a new contribution to uncertainty calculation. This could arise from different sources such as imperfect training, systematic errors, sampling noise, and unexpected shifts in the data. However, there are some approaches to uncertainty quantification such as Bayesian neural networks, dropout-based methods, and ensemble techniques which will be examined in our further work.

6. SUMMARY

The work described in this paper investigates the rocking movement of the dynamic force measurement setup. The proposed new evaluation method based on utilising artificial neural networks can be used to reduce measurement uncertainty and thereby improve the modelling of force transducer under dynamic conditions.

7. ACKNOWLEDGEMENT

EMPIR EURAMET

The EMPIR initiative is co-funded by the European Union's Horizon 2020 research and innovation programme and the EMPIR Participating States

This work is part of the 18SIB08 project funded by the EMPIR program. The authors would like to express their gratitude to our colleague Frank Hauschild, who supports us in preparing and performing the measurements throughout this work.

8. REFERENCES

[1] ISO 376, "Metallic Materials - Calibration of Force-Proving Instruments Used for the Verification of Uniaxial Testing Machines", International Organization for Standardization, Geneva, Switzerland, 2011.

[2] M. Kobusch, S. Eichstädt, L. Klaus, T. Bruns, "Investigations for the model-based dynamic calibration of force transducers by using shock excitation", in Proc. of 22nd IMEKO TC3 Conf., Cape Town, South Africa, pp 45-51, 2014. Online [Accessed 25 November 2022] https://www.imeko.org/publications/tc22-2014/IMEKO-TC3-TC22-2014-005.pdf

[3] R. Kumme, "Investigation of a primary method for a dynamic calibration of force measuring instruments: a contribution to reduce the measuring uncertainty", Doctoral Thesis TU Braunschweig, PTB-Bericht MA-48, 1996.

[4] N. Vlajic, A. Chijioke, "Traceable dynamic calibration of force transducers by primary means", Metrologia, vol. 53, pp. S136-S148, 2016. DOI: 10.1088/0026-1394/53/4/S136

[5] N. Vlajic, A. Chijioke, "Primary sinusoidal calibration of force transducers up to 2 kilohertz", XXI IMEKO, 2015. Online [Accessed 25 November 2022] https://www.imeko.org/publications/wc-2015/IMEKO-WC-2015-TC3-046.pdf

[6] N. Medina, J. de Vicebte, "Force sensor characterization under sinusoidal excitations", Sensors, vol. 14, 18454-73, 2014. DOI: 10.3390/s141018454

[7] S. Hassan, C. Schlegel, R. Kumme, R. Tutsch, "Parametric identification of dynamic force transducers", Journal of Physics, 2018. DOI: 10.1088/1742-6596/1065/4/042043

[8] "Dynamic calibration of force transducers according to the sinusoidal method", Guideline DKD-R 3-10, Sheet 2, 2019. DOI: 10.7795/550.20190507AEN

[9] S. Hochreiter, "Untersuchungen zu dynamischen neuronalen Netzen", Diploma thesis, TU Munich, 1991.

[10] S. Hochreiter, J. Schmidhuber, "Long Short-Term Memory", Neural Computation, vol. 9, pp. 1735-1780, 1997. DOI: 10.1162/neco.1997.9.8.1735

[11] K. Cho, B. Merrienboer, D. Bahdanau, "On the Properties of Neural Machine Translation: Encoder–Decoder Approaches", 2014. DOI: 10.3115/v1/W14-4012

[12] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling", NIPS, 2014. DOI: 10.48550/arXiv.1412.3555

[13] M. Schuster, K. Paliwal, "Bidirectional Recurrent Neural Networks", IEEE Transactions on signal processing, vol. 45, no. 11, 1997. DOI: 10.1109/78.650093