MARCH 18 2021

Explainable machine learning determines effects on the sound absorption coefficient measured in the impedance tube[a)]

Special Collection: Machine Learning in Acoustics

Merten Stender; Christian Adams; Mathies Wedler; Antje Grebel; Nobert Hoffmann

J. Acoust. Soc. Am. 149, 1932–1945 (2021)
https://doi.org/10.1121/10.0003755

JASA ARTICLE

Explainable machine learning determines effects on the sound absorption coefficient measured in the impedance tube[a)]

Merten Stender,[1,b)] Christian Adams,[2,c)] Mathies Wedler,[1] Antje Grebel,[2] and Nobert Hoffmann[1,d)]
[1]*Dynamics Group, Mechanical Engineering, Hamburg University of Technology, Am Schwarzenberg-Campus, Hamburg 21073, Germany*
[2]*Mechanical Engineering Department, System Reliability, Adaptive Structures, and Machine Acoustics, Technical University of Darmstadt, Otto-Berndt-Straße, Darmstadt 64287, Germany*

ABSTRACT:
Measurements of acoustic properties of sound absorbing materials in impedance tubes show poor reproducibility, which was demonstrated in round robin tests. The impedance tube measurements are standardized but lack precise definitions of the actual measurement setup, specimen preparation, and other factors that introduce uncertainty in practice. In this paper, machine learning models identify those factors that mostly affect the sound absorption coefficient from a large data set of more than 3000 absorption spectra measured in one impedance tube. The specimens are manufactured from one polyurethane foam, and different cutting technologies, different operators, different specimen diameters, different specimen thicknesses, and two different approaches to mount the specimens in the impedance tube are considered. Explainable machine learning techniques allow the identification and quantification of the most influential factors and, furthermore, the frequency ranges that are the most affected by the choice of these setup factors. The results indicate that besides the specimen thickness, also the operator affects the absorption coefficient by a directional and non-random relationship. Hence, it needs to be controlled carefully. The method proves to be a promising pathway for knowledge discovery from acoustic measurement data using explainability approaches for machine learning models. © *2021 Acoustical Society of America.*
https://doi.org/10.1121/10.0003755
(Received 16 November 2020; revised 19 February 2021; accepted 22 February 2021; published online 18 March 2021)
[Editor: Peter Gerstoft] Pages: 1932–1945

I. INTRODUCTION

The acoustic properties of porous materials, such as the sound absorption coefficient, are important to room acoustics design or noise control treatments as well as to determine material parameters of porous material models. The sound absorption coefficient describes the ratio of the absorbed sound power and the incident sound power. It is commonly measured by the transfer function method according to ISO 10534-2 (1998) in an impedance tube. A known issue of impedance tube measurements is the poor reproducibility among different laboratories and material specimens (Horoshenkov *et al.*, 2007; Pompoli *et al.*, 2017). It is concluded in these papers that the standard ISO 10534–2 should be clarified, particularly on (i) how to manufacture a suitable specimen, (ii) how to correctly mount it in the tube, (iii) how to process the data, and (iv) how to calibrate the measurement setup so that reproducibility can be improved. Similarly, other material properties of porous materials such as flow resistivity (Pompoli *et al.*, 2017) and viscoelastic properties (Bonfiglio *et al.*, 2018) have been reported to lack reproducibility as well.

Several studies (Cummings, 1991; Kang and Bolton, 1995; Kino *et al.*, 2012; Kino and Ueno, 2007; Pilon *et al.*, 2003, 2004; Song and Bolton, 2003; Song *et al.*, 2001; Tsay and Yeh, 2006; Vigran *et al.*, 1997) show the effect of edge constraints on the absorption coefficient. These edge constraints can stiffen the material specimen, which can affect the acoustical properties at low frequencies (Kang and Bolton, 1995; Vigran *et al.*, 1997). Different materials react with different sensitivity on edge constraints, e.g., air gaps affect the acoustic properties particularly of materials with high flow resistivity (Cummings, 1991). The effect of air gaps was quantified in Pilon *et al.* (2004) by a ratio between the macroscopic air gaps and the microscopic porous material structure. Similarly, a criterion was developed (Pilon *et al.*, 2003) that assesses how close the measured absorption coefficient matches the theoretical one of an infinitely large material specimen. ISO 10534-2 (1998) requires that the specimen must fit into the tube without being compressed or shaped convexly. Any gaps between the tube inner wall and the specimen should be closed to avoid air gaps, e.g., using modeling clay, petrolatum, or tape (ISO 10534–2, 1998). An appropriate specimen size is proposed in Kino and Ueno (2007), where the specimen is cut slightly smaller (approximately 0.5–1 mm) in diameter than the tube inner diameter to avoid effects due to edge constraints on the material properties. Further, this reduces shearing resonances of the specimen (Kino and Ueno, 2007). Shearing

[a)]This paper is part of a special issue on Machine Learning in Acoustics.
[b)]Electronic mail: m.stender@tuhh.de, ORCID: 0000-0002-0888-8206.
[c)]ORCID: 0000-0002-7307-8744.
[d)]ORCID: 0000-0003-2074-3170.

 0001-4966/2021/149(3)/1932/14/$30.00

https://doi.org/10.1121/10.0003755

resonances of the specimen cause the transmission loss to become minimal (Song *et al.*, 2001), which is studied numerically (Inoue and Sakuma, 2017; Song *et al.*, 2001) and experimentally (Song and Bolton, 2003). In practice, these edge constraints and the air gaps need to be controlled by appropriately manufacturing and mounting the specimens into the tube. How this can be achieved in practice is yet not described by ISO 10534–2 (1998), but it could be one of the reasons for poor reproducibility of impedance tube measurements (Horoshenkov *et al.*, 2007; Pompoli *et al.*, 2017). In practice, it is common to test multiple specimens of one material (Seybert *et al.*, 2013) or to carefully manufacture and install the specimen into the tube (Stanley, 2012) to reduce measurement uncertainty due to diameter or shape deviations as well as material inhomogeneity.

Experimental factors such as the mounting of the specimen into the impedance tube, diameter, or cutting technology were shown to affect the absorption coefficient in previous measurements of the authors as well (Wenzel *et al.*, 2020). Design of experiments (Montgomery, 2009) was used to determine those factors that may introduce significant uncertainty into the absorption coefficient. Five factors have been varied to determine how they affect absorption coefficient spectra, which resulted in more than 3000 measurements (Grebel, 2020): The material thickness, which is the most obvious one, the cutting technology, the diameter, the mounting direction, and the operator that manufactures the specimens. This paper builds on this large database and sets the starting point for a data-driven approach to study the sensitivities of the aforementioned factors. Given the large data set, this work employs machine learning (ML) methods to gain deeper insight into the effects of the aforementioned factors on the absorption coefficient. The ML models are built to output the correct factor value for a given absorption coefficient as input. In the second step, model-agnostic explainability techniques are utilized to understand the decision-making process of the ML model. Such explanation approaches have proven successful for scientific knowledge discovery from high-dimensional data recently (Kellner *et al.*, 2019). By quantifying the importance of input values, i.e., the absorption spectrum $\alpha(f)$, for the correct prediction of factor values, this work studies which factor affects the absorption coefficient the most. Since the thickness will obviously affect the absorption coefficient, it is taken into account in order to verify the ML approach and the explainability techniques.

The other factors may affect the edge constraints of the specimen mounted inside of the tube or may determine the amount of air gaps, which have been shown to affect the absorption coefficient measurements (Cummings, 1991; Kang and Bolton, 1995; Kino *et al.*, 2012; Kino and Ueno, 2007; Pilon *et al.*, 2003, 2004; Song and Bolton, 2003; Song *et al.*, 2001; Tsay and Yeh, 2006; Vigran *et al.*, 1997). However, these effects are yet not modeled by state-of-the-art absorption models such as the Biot–Allard model and extended ones (Allard and Atalla, 2009). A data-driven reverse engineering approach is used to identify the aforementioned factors from absorption coefficient spectra. The analysis of a neural network identifies important features in the absorption coefficient spectra. These features are affected to a certain amount by the factors of the specimen. The results indicate the most affected frequency ranges and also the amount to which different factors affect the absorption coefficient measurements. These results may help to better understand how different manufacturing technologies or mounting approaches affect the absorption coefficient at which frequencies. This may be a starting point for better controlling the uncertainty in impedance tube measurements, which has already been suggested in previous papers (Horoshenkov *et al.*, 2007; Pompoli *et al.*, 2017).

II. SOUND ABSORPTION COEFFICIENT MEASUREMENT

The sound absorption coefficient measurements for planar [i.e., one-dimensional (1-D)] sound wave propagation is briefly described in this section as well as the measurement setup of the impedance tube used to acquire the data and the experimental factors that are varied.

A. Theoretical background

The sound absorption coefficient reads (ISO 10534–2, 1998)

$$\alpha = \frac{P_\alpha}{P_\mathrm{I}} = 1 - |r|^2, \tag{1}$$

where P_α, P_I, and r denote the absorbed sound power, the incident sound power, and the reflection coefficient, respectively. For a planar acoustic wave that incidents in normal direction on the absorbing material, the reflection coefficient can be obtained from the transfer function

$$H_{12} = \frac{p_2}{p_1} = \frac{\mathrm{e}^{\mathrm{j}kx_2} + r\mathrm{e}^{-\mathrm{j}kx_2}}{\mathrm{e}^{\mathrm{j}kx_1} + r\mathrm{e}^{-\mathrm{j}kx_1}}. \tag{2}$$

p_1 and p_2 denote the sound pressures of two microphones that are located at the positions x_1 and x_2, respectively. k denotes the wave number. Solving Eq. (2) for r yields

$$r = \frac{H_{12} - H_\mathrm{I}}{H_\mathrm{R} - H_{12}} \mathrm{e}^{2\mathrm{j}kx_1}, \tag{3}$$

where

$$H_\mathrm{R} = \mathrm{e}^{\mathrm{j}k(x_1 - x_2)} \tag{4}$$

and

$$H_\mathrm{I} = \mathrm{e}^{\mathrm{j}k(x_2 - x_1)} \tag{5}$$

are the transfer functions of the reflected wave and the incident wave, respectively (ISO 10534–2, 1998).

TABLE I. Mean values of the material's acoustic and mechanical properties.

Parameter	Value
Density	22 kg/m^3
Young's modulus	148.6×10^3 N/m^2
Poisson's ratio	0.47
Loss factor	0.07
Flow resistivity	6900 Ns/m^5
Porosity	0.98
Tortuosity	1.33
Viscous length	74×10^{-6} m
Thermal length	166×10^{-6} m

Planar waves propagate in an impedance tube with circular cross section up to a frequency of (ISO 10534–2, 1998)

$$f_u = 0.58 \frac{c}{d}, \tag{6}$$

with c and d being the speed of sound in air and the tube diameter, respectively. ISO 10534–2 (1998) recommends a microphone distance of at least 5% of the wavelength that corresponds to the lowest frequency of interest. This leads to

$$f_l = 0.05 \frac{c}{x_1 - x_2}, \tag{7}$$

which is the lower frequency bound.

B. Foam material

A conventional sound absorbing polyurethane foam is chosen as test material. The acoustic and mechanical parameters are determined from measurements of five samples with 40 mm thickness. Table I summarizes the results in terms of mean values of the parameters. The mechanical parameters are determined according to ISO 18437–5:2011 (2011).

C. Measurement setup

The experimental setup of the impedance tube (SAB 102, TFS, Darmstadt, Germany) is illustrated in Fig. 1. Its measuring tube length and entire length (measuring tube and specimen mount) are 950 and 1200 mm, respectively. The tube's inner diameter is 90 mm, which leads to $f_u = 2200$ Hz according to Eq. (6), where $c = 343$ m/s is considered. The lower frequency yields $f_l = 357$ Hz for a microphone distance of $x_1 - x_2 = 48$ mm. The loudspeaker is driven by a white noise signal, which is generated by a noise generator (CAL102, TFS). The specimen is mounted inside the specimen mount. A plunger needs to be adjusted to the specimen thickness so that the specimen closes flush with the specimen mount. The microphones (1/4″ MI-17, Roga Instruments, Nentershausen, Germany) are flush mounted in the tube's wall. They are located with $x_2 = 59$ mm and $x_1 = 107$ mm distance from the specimen's surface.

The microphone signals are acquired with a RogaDAQ2 data acquisition system (Roga Instruments) and processed in MATLAB, where Eqs. (3)–(5) have been implemented as in-house code. Due to the high number of measurements, the measurement time was reduced as much as possible to perform the measurements within an acceptable time frame. An average over three measurements of 1 s each is found to result in sufficient absorption coefficient spectra. However, the raw data are more noisy than those for longer measurement times, but this can be handled by an appropriate data pre-processing before they are fed into the ML models (see Appendix A 1 for details). The transformation into the frequency domain is performed with $n_f = 1960$ frequency points.

The absorption coefficient is determined for $N = 3073$ different measurements of the polyurethane foam material. For this study, $k = 5$ factors are considered as listed in Table II. Each factor has an individual number of class members m.

Five factors are selected from the large number of factors that may affect the absorption coefficient, based on the literature and experience obtained in previous measurements by the authors. These factors are subsequently described in more detail, including their classes. The specimen thicknesses of 30, 40, 50, and 80 mm listed in Table II result from the available foam plate thicknesses. The specimens can be cut more or less true to size and shape from plates of these thicknesses with the cutting technologies listed in Table II. Nevertheless, specimens cut with different technologies are different in shape upon visual inspection. Figure 2 illustrates different specimens cut with band saw and water jet. Particularly, the cutting edges of the band saw are rougher than those from the water jet. A different edge

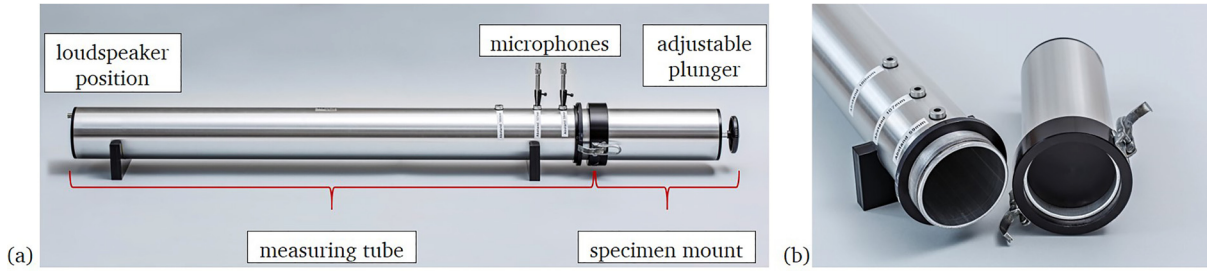

FIG. 1. (Color online) Impedance tube for sound absorption coefficient measurements with specimen mount attached (a) and disassembled (b); source: Fraunhofer LBF.

https://doi.org/10.1121/10.0003755

TABLE II. Overview on the factors and classes per factor evaluated in this study. *m*, number of measurements available for the respective factor class. Abbreviations used to classify cutting technology: S, band saw; L, low speed; S, high speed; F, foam guided manually; H, foam guided using the cutting device shown in Fig. 3(b).

Factor	Class description	*m*
Specimen thickness	30 mm	768
	40 mm	768
	50 mm	769
	80 mm	768
Specimen diameter	89 mm	1008
	90 mm	1040
	91 mm	1025
Cutting technology	Hot wire (HZD)	211
	Water jet cutting (WSS)	216
	Circular blade (MES)	66
	Band saw SLF	642
	Band saw SLH	648
	Band saw SSF	648
	Band saw SSH	642
Mounting	Plunger moved	1537
	Plunger fixed	1536
Operator	Computer (pc)	427
	Person 1 (p1)	930
	Person 2 (p2)	852
	Person 3 (p3)	864

roughness may lead to different edge constraints when the specimen is mounted in the tube. Thus, it can be assumed that different cutting technologies affect the absorption coefficient. Hot wire and water jet cutting are computer-controlled cutting technologies (see Table II). These technologies are classified into the class *computer* under the factor *operator* in Table II. A circular blade is another common cutting tool (Stanley, 2012), but it is available with nominal diameter of 90 mm only [see Fig. 3(a)] so that only specimens with 90 mm diameter are obtained for circular blade cutting. These specimens are cut by *person 1* in factor *operator* of Table II. The band saw (abbreviated S) is operated under four different conditions: The cutting speed is set to low (abbreviated L) or high (abbreviated S), and the specimens are cut by manually moving the foam plate (abbreviated H) or by using the cutting device (abbreviated F) shown in Fig. 3(b). The cutting device helps to manually guide the specimen along the band saw, which improves the accuracy of the specimens' dimensions and shape. It can, thus, again be assumed that these different manufacturing

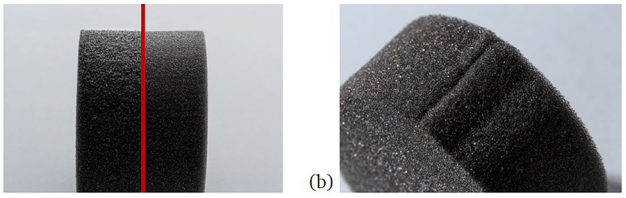

FIG. 2. (Color online) (a) Specimens cut with band saw with slow speed (left) and water jet (right); (b) example of a shape deviation when manually cutting a specimen; source: Fraunhofer LBF.

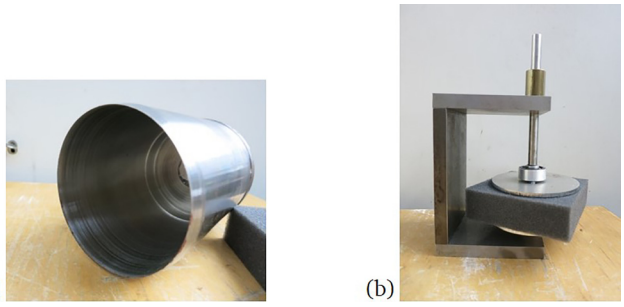

FIG. 3. (Color online) Tools for cutting foam specimens: (a) circular blade; (b) cutting device for band saw cutting.

techniques of the specimens affect the edge constraints and consequently the absorption coefficient. The band saw is operated by person 1, person 2, or person 3. Altering the operating person is considered a factor to check how interpersonal changes affect the absorption coefficient measurements. The different specimen diameters according to Table II still fit into the tube as described in ISO 10534–2 (1998), but they represent typical tolerances that have been observed during manual cutting of the specimens using the band saw without the cutting device. Finally, the *mounting* of the specimen into the impedance tube is altered. This should represent different experimenters and their possibly different approaches to mounting the specimens into the tube. Either the plunger is adjusted to the specimen thickness and the specimen is flush mounted into the specimen mount (called plunger fixed) or the plunger is moved to maximum thickness, the specimen is flush mounted to the plunger, and the specimen is slowly pushed with the plunger until it is flush mounted in the specimen mount (called plunger moved). It can again be assumed that such different ways of mounting the specimens will affect their edge constraints and, thus, the absorption coefficient.

Overall, from the variations of the factors (and classes per factor) given in Table II, 3073 valid measurements for the absorption coefficient are available. For each class, a large number of measurement results exist, which is a promising starting point for data-driven analysis techniques for revealing the nonlinear relations between factors and the measured absorption coefficient.

D. Data pre-processing

The raw measurements $\alpha(f)$ were pre-processed to facilitate a more consistent data analysis. Some measurements exhibited artifacts, like sudden drops of the absorption coefficient to vanishing values, for a range of some Hz. Besides being hard to explain physically, those outliers created long-tailed data distributions (considering the complete range of absorption coefficients α), which constitute substantial challenges for ML techniques. Therefore, these outliers were filtered out by rolling-median smoothing of the measurement curves. Appendix A 1 elaborates on the data pre-processing steps in more detail. Figure 4 displays all 3073 absorption coefficient measurements after being

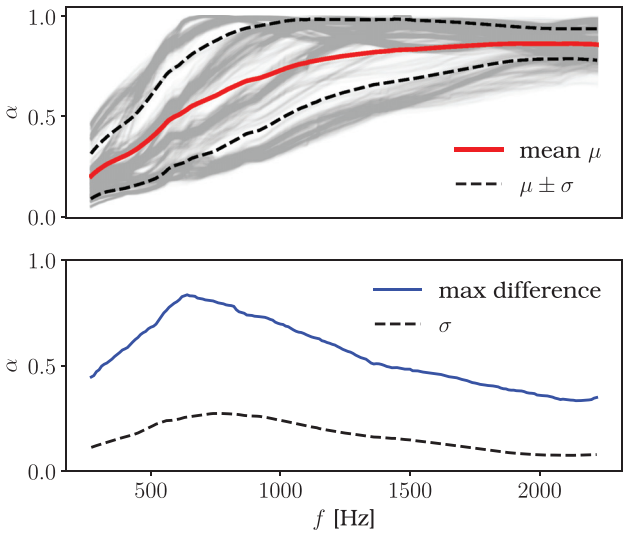

FIG. 4. (Color online) All 3073 measurements after data cleaning, their mean μ and standard deviation σ (top panel) and measures of variance (bottom panel).

cleaned, along with the frequency-wise mean value and standard deviation. There is a substantial dispersion in the absorption spectra due to the factors varied in the experimental measurements. Since we included the material thickness as a factor in the measurements, we expect that the thickness explains most of the dispersion, particularly between 500 and 1000 Hz, where the absorption coefficient spectra differ most. This can be seen from the maximum of the maximum difference, i.e., the maximum of the blue curve at the bottom of Fig. 4, which occurs between 500 and 1000 Hz. These frequencies seem to be particularly sensitive to changes in the absorption coefficient. In practice, the absorption behavior in that frequency range, e.g., of a room acoustic treatment or a noise abatement measure, can be designed toward the requirements by choosing an appropriate thickness. Nevertheless, the other factors altered during the measurements may cause significant dispersion of the absorption coefficient spectra as well. Factors such as the operator, which can be seen as some sort of unwanted variation, should hardly affect the results for a valid and standardized measurement technique.

Furthermore, a downsampling approach is implemented to reduce the frequency resolution of the absorption coefficient measurements. The underlying reasoning is related to the course of dimensionality, especially with regard to explaining a ML model that takes the input features $\mathbf{X}$ to predict outputs $\mathbf{y}$: The computation of feature importance values becomes increasingly complex as the number of input features grows (Fulcher and Jones, 2017; Guyon and Elisseeff, 2003). Given a fixed "amount of importance" to distribute among all features, large numbers of features will blur the feature importance: All features will carry *some* degree of importance, and none will stand out significantly. Therefore, dimensionality reduction of the raw input data is performed before building ML models. Fewer input features are achieved by downsampling the absorption curves to a coarser frequency grid using cubic spline interpolation. Particularly, downsampling from $n_\mathrm{f} = 1960$ (corresponding to 1 Hz frequency spacing) in the raw measurements to $n_\mathrm{f} = 980$ (2 Hz) and $n_\mathrm{f} = 196$ (10 Hz) is performed. Even though frequency resolution of the importance scoring is lost in this process, relevant feature values (i.e., frequency regimes) will be emphasized, which can help to interpret the results. Our studies show that the effect of the downsampling rate on the model performance and the explanations is mostly negligible.

The factors and their class values are considered as output variables $\mathbf{y}$ for the ML task. As these values are categorical data, they are "one-hot encoded" into m-dimensional binary vectors per factor. The output of the ML models will then indicate the predicted probability per class (summing up to unity per factor). Considering the *cutting technology* factor as an example, the one-hot encoded ground truth output vector would read

$$\mathbf{y} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}^\top \tag{8}$$

for a sample that was produced by the circular blade (MES) technology (following the vertical order of the class members given in the second column of Table II).

III. METHODS

To find out which relation between the different factors and the measured absorption coefficients exists, ML models are set up to predict the factors from a given absorption measurement curve. Subsequently, the decision-making process of the model is investigated by explainability approaches. In this way, the most relevant frequency ranges of the absorption curves are identified, which allowed the model to predict the factors of the experiment. High prediction accuracy indicates a deterministic relation between those frequency ranges and the factor. By implication, this means that the identified frequency regimes are the ones that are the most affected by one factor, which is the central objective to study in this paper: *How is the absorption coefficient measurement affected by individual factors?*

A. Formulation of the ML task

Univariate ML classification models $\mathcal{M}$ are considered in this work. These models predict a single factor from the absorption coefficient $\alpha(f)$, which is provided as input $\mathbf{X}$. Such a setup

$$\mathcal{M} : \mathbf{X} \mapsto \mathbf{y} \tag{9}$$

is typically referred to as a *single-label multi-class classification task*. The dimension of the output vector $\mathbf{y}$ is $(1 \times m_k)$, where m_k denotes the number of one-hot encoded classes for the current factor k. For example, if a model is trained to predict the cutting technology from the absorption coefficient spectra, the model will map the inputs $\mathbf{X} \in \mathbb{R}^{1\times n_\mathrm{f}}$ to the outputs $\mathbf{y} \in \mathbb{R}^{1\times 7}$, as there are $m_k = 7$ class members of the

https://doi.org/10.1121/10.0003755

factor *cutting technology*; see Table II. The univariate modeling will give first indications for factors that crucially affect the absorption coefficients. If high classification scores can be obtained for a specific factor, it can be concluded that this factor leaves a strong signature in the data, irrespective of other factor variations.

The ML models are set up with the framework TENSORFLOW (version 2.3.0) in the programming language PYTHON (version 3.7). The code used to generate the results presented in this paper has been made publicly available (Stender *et al.*, 2020).

B. Model setup and selection

A five-layer perceptron (MLP) is chosen as the *baseline* model $\mathcal{M}_1$. A deep MLP is implemented as a second model $\mathcal{M}_2$ with an adaptive number of hidden layers (as a function of the input parameter size n_f); see Appendix A 2 for more details on the models. Fivefold cross-validation (CV) (Stone, 1974) with 2458 training samples and 615 test samples is used to compute the generalization properties and compare the different model performances in terms of accuracy. Table III reports the resulting test set accuracy values by their CV mean and standard deviations. The model performance is discussed in Sec. IV.

C. Model explanations through SHAP values

From the large and quickly growing toolset of explainability approaches (Lipton, 2018; Ribeiro *et al.*, 2016) for ML models, Shapley additive explanations (SHAP) (Lundberg and Lee, 2017) by Scott Lundberg and coworkers represents an additive feature attribution method that is based on cooperative game theory. Relying on the concept of the (Nobel Prize winning) Shapley values, marginal contributions are computed to quantify the importance of each input feature value for the decision-making process, i.e., the feature's contribution to the model output. Following the desirable properties of Shapley values, e.g., local accuracy, missingness, and consistency, SHAP represents a *unique* way to assign additive feature importance measures to the input values. Using Shapley sampling, the effect of removing a variable from the input vector is estimated by sampling from the so-called *background* data set. In this work, the *DeepExplainer* (Lundberg and Lee, 2017) of SHAP is used with the complete data set serving as background data set.

To illustrate the interpretation of SHAP values, an example case is constructed. Specifically, we want to predict the specimen *thickness* (in this example 30 mm) from a single absorption coefficient measurement $\alpha(f)$. For the example displayed in Fig. 5, the model makes the correct prediction: Given $\alpha(f)$ as input, $\mathcal{M}_1$ predicts a specimen thickness of 30 mm as output. As there are $m = 4$ classes (30, 40, 50, and 80 mm) for the factor *thickness*, which are equally represented in the data set, the expected value $\mathbf{E}(X)$ is 0.25 for each class. The key idea of SHAP is to explain the difference between the expected prediction vector $[0.25, 0.25, 0.25, 0.25]^\top$ (corresponding to random guessing) and the actual model prediction $\hat{\mathbf{y}} = [1, 0, 0, 0]^\top$. Feature importance SHAP values $\phi_{i,j}$ are computed for each input quantity $i = 1, \ldots, n_\mathrm{f}$ and each target dimension $j = 1, \ldots, m$. A negative SHAP value indicates that the corresponding input feature contributed to a model prediction that is lower than the expected value, and a positive SHAP value indicates that the corresponding feature influenced the model to output a prediction that is larger than the expected value.

TABLE III. Performance of the ML models $\mathcal{M}_1$ and $\mathcal{M}_2$ for the univariate classification task. The mean fivefold CV test set accuracy score and the standard deviation are reported for both models, each dimensionality reduction rate, and each factor.

	Thickness	Diameter	Cutting	Mounting	Operator
Class: $\mathbb{E}(X)$	30 mm: 0.25 40 mm: 0.25 50 mm: 0.25 80 mm: 0.25	89 mm: 0.33 90 mm: 0.34 91 mm: 0.33	HZD: 0.07 MES: 0.02 SLF: 0.21 SLH: 0.21 SSF: 0.21 SSH: 0.21 WSS: 0.07	pl. moved: 0.5 pl. fixed: 0.5	pc: 0.14 p1: 0.30 p2: 0.28 p3: 0.28
No dimensionality reduction, $n_\mathrm{f} = 1960$					
$\mathcal{M}_1$ MLP	1.00 ± 0.00	0.32 ± 0.01	0.32 ± 0.04	0.49 ± 0.01	0.85 ± 0.07
$\mathcal{M}_2$ deep MLP	1.00 ± 0.00	0.32 ± 0.01	0.20 ± 0.01	0.49 ± 0.01	0.38 ± 0.12
Moderate dimensionality reduction, $n_\mathrm{f} = 980$					
$\mathcal{M}_1$ MLP	1.00 ± 0.00	0.32 ± 0.01	0.39 ± 0.06	0.49 ± 0.01	0.83 ± 0.07
$\mathcal{M}_2$ deep MLP	1.00 ± 0.00	0.32 ± 0.01	0.20 ± 0.01	0.49 ± 0.01	0.28 ± 0.10
Strong dimensionality reduction, $n_\mathrm{f} = 196$					
$\mathcal{M}_1$ MLP	1.00 ± 0.00	0.46 ± 0.07	0.36 ± 0.06	0.49 ± 0.01	0.82 ± 0.05
$\mathcal{M}_2$ deep MLP	1.00 ± 0.00	0.32 ± 0.01	0.25 ± 0.03	0.49 ± 0.01	0.30 ± 0.03

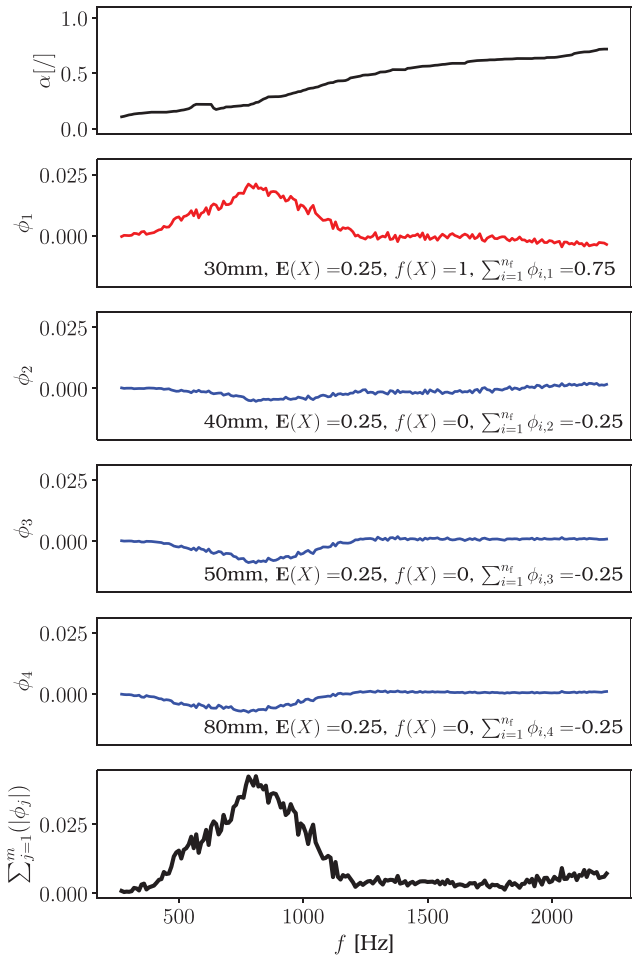

FIG. 5. (Color online) Univariate modeling case for the *thickness* factor (30 mm in this example) using $\mathcal{M}_1$ and $n_f = 196$ for SHAP analysis. Absorption coefficient used as input (top panel) and resulting SHAP values ϕ for each class m (middle panels) and the summation of all SHAP values (bottom panel).

Now consider the SHAP analysis displayed in Fig. 5. The SHAP importance values for each $\alpha(f_i), i = 1, \ldots, n_f$ for predicting an output of 40 mm are denoted as ϕ_2. Most of these values are vanishing, especially for higher frequencies. Non-vanishing negative SHAP values are found for $600 \leq f \leq 1300$ Hz, hence indicating that the corresponding input values $\alpha(f)$ are found to contribute to a model output that is below the expected value of 0.25. Hence, the absorption coefficients in this frequency range are relevant for letting the model decide that this sample does *not* belong to the class of 40 mm specimen thickness. As SHAP is an additive feature attribution method, the sum of ϕ_2 along f equals the difference between expected value (0.25) and actual model prediction (0.0), i.e., $\sum_{i=1}^{n_f} \phi_{2,i} = -0.25$. Very similar observations can be made for ϕ_3 (corresponding to 50 mm thickness prediction) and ϕ_4 (corresponding to 80 mm thickness prediction). ϕ_1 denotes the feature importance values per frequency point for making the correct prediction, i.e., in this case predicting a sample thickness of 30 mm. These SHAP values indicate that the aforementioned frequency range $600 \leq f \leq 1300$ Hz carries the most (positive) feature importance, hence pushing the model prediction from the expected value 0.25 to the prediction of 1.0. Again, the sum of ϕ_1 along the frequency dimension equals the difference between $\mathbf{E}(X)$ and $\hat{\mathbf{y}}(1)$. To aggregate the $m \times n_f$ SHAP values, their absolute values can be summed up frequency-wise, as shown in the bottom panel of Fig. 5. This plot indicates which frequency ranges of the absorption coefficients are the most important ones for both (i) predicting the 30 mm class and (ii) *not* predicting the other classes. It can be observed that for this single measurement, the absorption values above $f = 1400$ Hz do not play any role in the model's decision-making process and, therefore, that these measurement values do not carry any decisive information for inferring the specimen thickness from the absorption coefficient measurement curve, which is in line with experience and knowledge in acoustics. Very similar behavior of the SHAP values was obtained for the original frequency sampling $n_f = 1960$. Following the core objective of this work, one may thus conclude that the absorption coefficient spectra in $600\,\text{Hz} \leq f \leq 1300\,\text{Hz}$ are heavily (and deterministically) affected by the specimen's thickness. As this statement is based on one sample, i.e., on *local explanations*, this hypothesis is validated in Sec. IV by computing SHAP values for all 3073 measurements, thereby combining all individual SHAP values into *global explanations*.

IV. RESULTS

The results of the investigation are discussed regarding the model performance and the explanation of the model's decision-making process. Among the factors varied in the measurements, the thickness is expected to be the most influential factor. Although this might be an obvious result for acousticians, it is an important one to verify the proposed methodical approach to knowledge discovery from data using ML and explainability approaches. The less the remaining factors will affect the absorption coefficient spectra, the more robust the impedance tube measurement technique is against variations in these factors. Further, the explainability approaches reveal at which frequencies the factors affect the absorption coefficient most.

A. Model performance

The results of the model selection study presented in Table III indicate that both factors *thickness* and *operator* leave a dominant signature in the absorption coefficient spectra. High classification accuracy values (see Appendix A 3 for definition) are observed, indicating that the inputs carry highly distinctive characteristics for separately predicting both factors from the absorption measurements. Particularly, perfect classification scores are obtained for the thickness factor, such that this value was predicted by every model evaluated in this study without a single error. Considering the overall number of 3073 samples evaluated during the CV, this is a very promising result, which was expected from the acoustical point of view. Owing to the perfect scores across all models, the relations between the absorption coefficient and the thickness factor values are

https://doi.org/10.1121/10.0003755

necessarily deterministic. The *operator* seems to have some effect, mostly deterministic as indicated by high accuracy values in the CV study, on the absorption coefficient. Such effects can be seen as an unwanted variation, which is hard to control in experimental laboratories. This might be one of the reasons for a poor reproducibility of absorption coefficient measurements (Bonfiglio *et al.*, 2018; Horoshenkov *et al.*, 2007).

It is surprising that both models cannot predict the remaining factors *diameter*, *cutting*, and *mounting* at better-than-random accuracy values. These factors were expected to affect the edge constraints and consequently the absorption coefficients. But for the diameter, the cutting, and the mounting, the model predictions are not better than the expected values, meaning that the models do not perform better-than-random guessing. This observation indicates that either these factors do not significantly affect the measurements or that they at least do not affect the results in a univariate fashion. However, there may be multivariate effects, i.e., importance in combination with other factors. In such a situation, the classification accuracy would increase once multivariate inputs are taken into account. Following the preliminary SHAP analysis discussed in Sec. III C, the data set was truncated to a frequency range $f \leq 1400$ Hz to find out whether the results would change if the data were restricted to the most relevant frequency range. It is found that the qualitative model performance does not change, neither for the thickness and operator factor nor for the diameter, cutting, and mounting factor; see Appendix B. As a third observation, the degree of input dimensionality reduction does not affect the results significantly, especially for the thickness and the operator factors, which confirms that the models are robust and that they have successfully learned the qualitative character of the absorption coefficient measurements. Furthermore, the simpler baseline model $\mathcal{M}_1$ outperforms the more complex model $\mathcal{M}_2$ in all cases, even though a saturation in the training process was observed. Hence, longer training would not significantly increase the model performance; see model loss curves displayed in Fig. 11 and Table IV in Appendix A.

In the explanation step, consequently only the $\mathcal{M}_1$ models for the thickness and operator factor are studied. Explaining the other models, i.e., those with low accuracy and those for the other factors, would be meaningless: One would explain weakly performing (or potentially wrong) models. Following our reasoning for the dimensionality reduction and the results in Table III, the explainability approaches are applied to the models trained on the data with the strongest dimensionality reduction $n_\mathrm{f} = 196$.

B. Explaining model decisions

Following the previously presented *local* SHAP analysis for a single measurement (see Fig. 5), *global* explanations are presented in this section through aggregating all local explanations for all measurements. Final instances of the $\mathcal{M}_1$ model are trained on 0.75–0.25 train-test set splits (best-practice ratio in ML) of the data, and model predictions are made for all samples. Particularly, only those measurements are considered for explanation for which the model obtained a correct prediction at a minimal confidence of 90%, thus neglecting wrong or less confident predictions. As $\mathcal{M}_1$ obtains perfect predictions for the thickness factor, all 3073 measurements can be considered for this factor. For each sample, SHAP values $\phi_{i,j}$ are obtained, and the frequency-wise sum of their absolute values

$$\psi = \psi(f_i) = \sum_{j=1}^{m} |\phi_{i,j}|, \quad i = 1, \ldots, n_\mathrm{f} \tag{10}$$

is computed as depicted in the bottom panel of Fig. 5. The resulting vector ψ indicates frequency regimes of importance for the model. For aggregation of all measurements, first-moment frequency-wise statistics are computed across all ψ vectors belonging to each class. Figure 6 depicts the median values of all ψ values per class along the frequency range. The resulting picture shows a clear trend: For all classes, the ψ values increase from $f = 300$ Hz to $f = 850$ Hz and then decrease again. For the frequency range above 1200 Hz, only small values can be observed. Significant values, such as $\psi > 0.002$, can be found in the frequency range $500 \leq f \leq 1000$ Hz, meaning that the absorption coefficients in this regime are crucially affected by the thickness factor. Particularly, the amplitude of ψ measures the amount to which the measurement is affected.

Even though this compact visualization of a large number ($3073 \times 4 \times 196$) of SHAP values shows a similar behavior per class, a deeper analysis of the per-class distribution of SHAP values is shown in Fig. 7. Here, the per-class mean value and standard deviation per frequency value are depicted to display the variation between samples and thus pieces of information that get lost when showing only the median of all ψ values as global explanations. It turns out that for the first two classes (30 and 40 mm), the variation of the SHAP values is the largest in 800 Hz $\leq f \leq 1200$ Hz, while the two remaining classes (50 and 80 mm) exhibit two distinct frequency regions of large

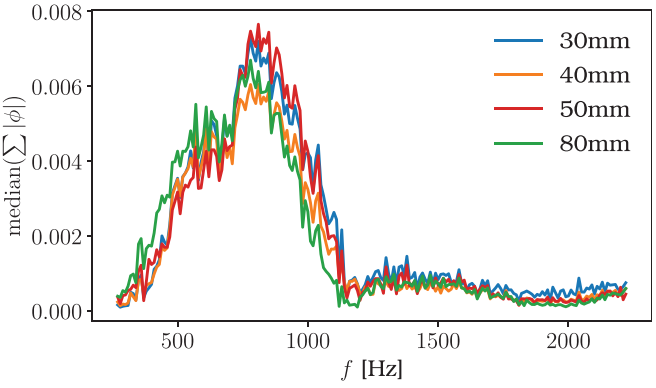

FIG. 6. (Color online) Median sum of absolute SHAP values per sample categorized by the correctly predicted class label for the factor *thickness*. Large values indicate strong relevance for the prediction process of the ML model.

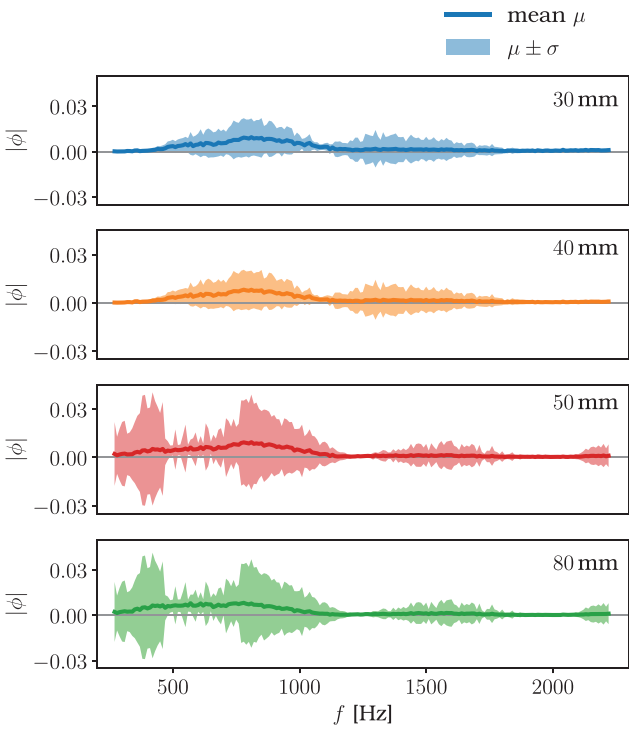

FIG. 7. (Color online) Mean values and standard deviations of the per-class sum of SHAP values for the factor *thickness*.

variations in the SHAP values. Hence, there are significant variations in the local explanations per sample underlying the global explanations shown in Fig. 7. The larger the variations, the more the per-sample SHAP values differ. Therefore, one can conclude from the rather small variations in the frequency range 500 Hz $\leq f \leq$ 800 Hz that this region is consistently highly decisive (non-zero mean) for all samples (low variance). From the viewpoint of the acoustical properties, it is well-known that the maximum absorption shifts to lower frequencies for an increasing thickness. Further, the absorption coefficient can also decrease at frequencies above the maximum. These effects are particularly prominent in the above-mentioned frequency range that is most affected by the thickness variations in our absorption coefficient measurements. In the high frequency regimes, variance is low, but so is the mean. This behavior indicates that those frequency ranges are consistently *not* affected by the change of the specimen thickness, which is also a well-known result from the acoustics perspective. For the *operator* factor, the picture is less clear (see Fig. 8). The operator factor is *not* intended to have any physical relation to the absorption coefficient spectra. It rather introduces some unwanted variation into the absorption coefficient spectra. As a consequence, the relevant frequency regime is not as pronounced as for the thickness factor. Larger values of ψ can be observed up to $f = 1500$ Hz, thus indicating the importance of this factor for the measured absorption coefficient in the lower half of the frequency range. However, there are several peaks observable, pointing at variable feature importance along the frequency axis. For example, the class *pc* has a distinguished peak at 800 Hz $\leq f \leq$ 1000 Hz,

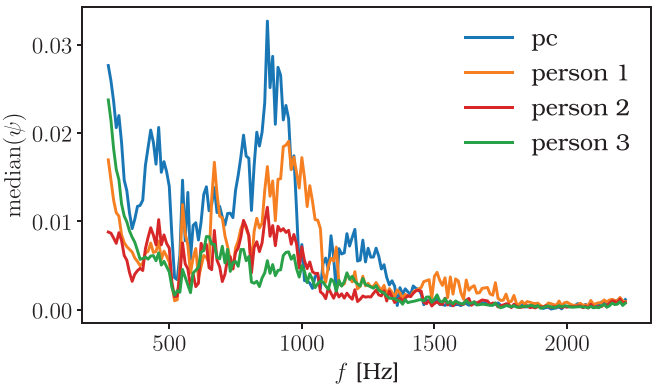

FIG. 8. (Color online) Median sum of absolute SHAP values per sample categorized by the correctly predicted class label for the factor *operator*. Large values indicate strong relevance for the prediction process of the ML model.

while the relevance for *person 3* is relatively equally distributed. Also, the variance per class, depicted in Fig. 9, shows a behavior that differs in person 1 and person 2 compared to the other two classes. For the former two classes, there are large variances in the SHAP values observable, pointing at strongly varying importance scores for the corresponding measurements. The frequency regime below 1500 Hz is crucially affected by person 1 and person 2 for some specimens, but not for others. It is concluded that the effect is either highly variable, possibly also as a function of other factors, or less deterministic. However, in the latter case, we would have observed much higher scattering in the training

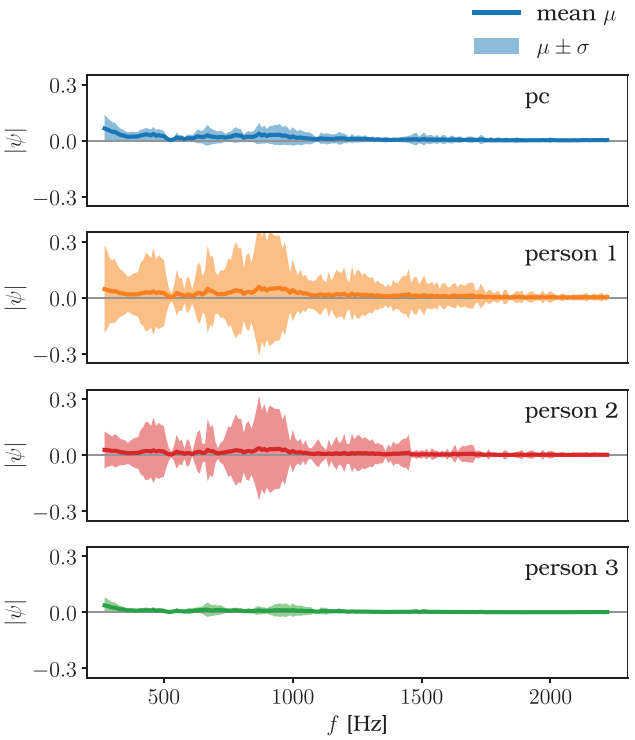

FIG. 9. (Color online) Mean values and standard deviations of the per-class sum of SHAP values for the factor *operator*.

https://doi.org/10.1121/10.0003755

process of the model. In comparison, the effect of the computer and person 3 on the absorption coefficient is much more consistent.

Overall, the factor *operator* tends to affect the absorption coefficient in a much more blurred way than the thickness does. This is an important—and to the authors' best knowledge—novel result in uncertainty analyses of impedance tube measurements. It highlights the necessity to improve the procedures and, thus, to control such uncertainty.

Note that interactions are neglected in the univariate models. If a certain factor affects the absorption coefficient only via interactions with, or in combination with, specific realizations of other factors, the dependencies are multivariate. Hence, *multivariate classification models* would be required for resolving these factor dependencies. Future work will follow up on these multivariate modeling and explanation tasks.

Two further analyses are performed to confirm the findings presented before. First, the frequency range of the input features was reduced to $f < 1400$ Hz, which is found to be the most relevant frequency range in the previously discussed studies. The model performance scores, reported in Table V, do not change significantly: The thickness and operator factors can be predicted at high accuracy values only. The subsequent SHAP analysis, shown in Fig. 12, confirms the findings from the SHAP analysis run on the full frequency range. However, the results are slightly different for a thickness of 80 mm and for the factor *operator*. This highlights that (i) interactions need to be taken into account by a multivariate modeling to determine yet hidden patterns in the data and (ii) particularly the factor *operator* tends to be highly variable. In a second analysis, the complete data set was subdivided into four subsets. Each subset corresponds to a fixed sample thickness value (30, 40, 50, or 80 mm). Models are trained for each subset to determine if previously hidden relationships can be detected for those factors that have not shown high performance scores upon univariate classification yet. The results, reported in Table VI, confirm the results from the full data set: The operator factor can still be predicted at high accuracy values, while the diameter, cutting, and mounting factors cannot be predicted. For the 40 mm sample thickness subset, also the operator performance drops to the expected values. Hence, no relationship between the operator and the absorption coefficient was identified for that sample thickness, which again highlights the high variability of that factor. Overall, these two additional analyses support the previous findings.

After all, it is important to point out that only a specific model out of many good models was explained. Hence, choosing a different model may result in slight deviations in the results. However, our results are consistent for different dimensionality reduction rates, i.e., different models, which is a promising result with regard to the generalization properties of our models and, thus, the results obtained through explainability approaches.

V. CONCLUSION

In this paper, large amounts (3073) of absorption coefficient measurements of one polyurethane foam are analyzed, where five factors, i.e., thickness, diameter, cutting, mounting, and operator, are varied upon impedance tube measurements according to ISO 10534–2 (1998). ML and explainability approaches are developed and validated in several data analyses to determine correlations with the mentioned factors in the absorption coefficient spectra.

It is found that particularly the *specimen thickness* and *operator* factors are strongly linked to the absorption coefficient spectra. While this effect is well-known for the former, the latter finding is especially interesting due to the fact that the data-driven model indicates a non-random effect of the operator. Our findings are in line with reproducibility experiments (Bonfiglio *et al.*, 2018; Horoshenkov *et al.*, 2007), which show that the operator introduces some kind of variation into the absorption coefficient spectra. The variation due to the operator factor can hardly be linked to any physical effect yet. However, the statistics of the model explanations (SHAP values) indicate that different operators leave different signatures in the absorption coefficient spectra. To control this uncertainty in practice, this signature should be controlled, e.g., by using one identical operator for all specimens, whose absorption coefficients are subject to a direct comparison. This is a first step to control the uncertainty due to the operator, but it will not increase the robustness of the measurement procedure itself. To better understand the operators' effects on the absorption coefficient, a multivariate modeling is required. The remaining factors diameter, cutting, and mounting are not affecting the absorption coefficient in a deterministic manner. Hence, they show no correlation that significantly differs from a random guess. However, the univariate modeling approach chosen in this paper neglects interactions. To deepen the insight into the effects of these factors, a multivariate modeling will be considered in a future analysis. Furthermore, other porous materials such as fibrous materials may show different effects on the factors investigated in this paper. An analysis of absorption coefficients of other materials is also part of future research using the method proposed in this paper as well as multivariate methods.

The purely data-driven analysis maps very well with well-known acoustical properties of sound absorbing foam, i.e., extracting the effect of thickness on absorption coefficient spectra. This is a promising finding from the ML modeling viewpoint, as it confirms the proposed methodology to be consistent, robust, and explanatory. Careful model selection needs to be undertaken using CV and checking for non-overfitting models, such that explanation techniques can then reveal parts of the model decision-making process in a second step, which finally can help to discover knowledge from the data. Hence, the proposed method represents a new pathway for finding novel (and maybe more hidden) patterns, such as for the operator, also in other data sets. For

example, determining such correlations in the impedances or the transfer functions that can be obtained from the microphone signals is also of interest.

ACKNOWLEDGMENTS

The authors thank the anonymous person 1, person 2, and person 3, who cut so many foam specimens, which allowed us to perform this measurement campaign. Further, we thank Mr. Sören Wenzel for measuring more than 3073 specimens with the impedance tube. M.S. was supported by the German Research Foundation (DFG) within the Priority Program SPP1879 *calm, smooth, smart* under Grant No. 314996260.

APPENDIX A

1. Data pre-processing

The raw measurement data are the sound pressure time signals of the two microphones. These data are transformed into the frequency domain to calculate the absorption coefficient using Eqs. (3)–(5). Note that $k = \omega/c = 2\pi f/c$. The absorption coefficient spectra are somewhat noisy due to the short measurement time of only 1 s, since not all random disturbances are averaged out; see top panel of Fig. 10 for an example. Nevertheless, the main characteristics of the absorption coefficients are captured so that the noise can be filtered out before feeding the data into the ML models. Particularly, negative absorption coefficients were observed in some of the measurements. These artifacts are replaced by $\alpha = 0$ values. Next, sudden drops in the absorption coefficient spectra are being inpainted by (i) rolling-median (window length 25) and (ii) rolling-maximum (window length 50) processing. The resulting smoothed data follow the original data closely but do not contain the sudden drops, which cannot be explained physically. An example for the data cleaning is illustrated in Fig. 10 along with the complete set of all cleaned data sets.

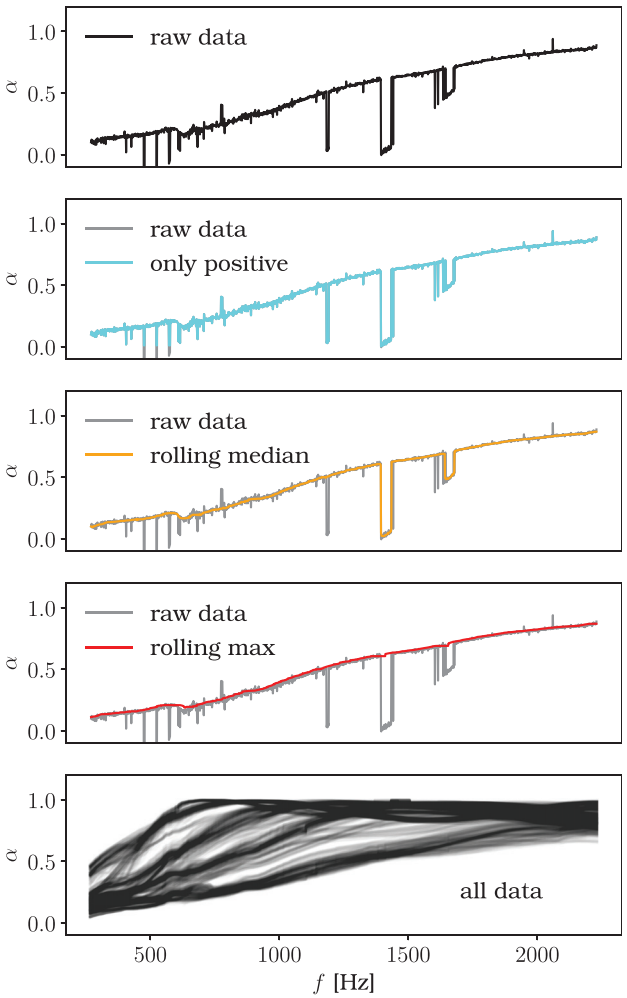

FIG. 10. (Color online) Example for the data cleaning employed prior to the modeling phase and the complete set of cleaned data (bottom panel).

2. ML models

a. Configuration of $\mathcal{M}_1$

The number of input neurons equals the number of frequency sampling points n_f, and the number of neurons in the output layer equals the number of classes m_k for factor k.

TABLE IV. Performance of the ML models $\mathcal{M}_1$ and $\mathcal{M}_2$ for the univariate classification task. The mean fivefold CV test set loss value and the standard deviation are reported for both models, each dimensionality reduction rate, and each factor.

	Thickness	Diameter	Cutting	Mounting	Operator
No dimensionality reduction, $n_f = 1960$					
$\mathcal{M}_1$ MLP	0.0062 ± 0.0096	1.0991 ± 0.0003	1.5070 ± 0.1042	0.6932 ± 0.0001	0.4091 ± 0.1767
$\mathcal{M}_2$ deep MLP	0.0160 ± 0.0106	1.0992 ± 0.0003	1.7663 ± 0.0197	0.6933 ± 0.0002	1.3508 ± 0.0105
Moderate dimensionality reduction $n_f = 980$					
$\mathcal{M}_1$ MLP	0.0002 ± 0.0003	1.0992 ± 0.0003	1.3770 ± 0.0993	0.6933 ± 0.0001	0.4076 ± 0.1500
$\mathcal{M}_2$ deep MLP	0.0382 ± 0.0394	1.0992 ± 0.0003	1.7665 ± 0.0199	0.6933 ± 0.0002	1.3509 ± 0.0098
Strong dimensionality reduction $n_f = 196$					
$\mathcal{M}_1$ MLP	0.0005±0.0006	0.9918±0.0573	1.3717 ± 0.1227	0.6933 ± 0.0001	0.4154 ± 0.1116
$\mathcal{M}_2$ deep MLP	0.0032 ± 0.0041	1.0991 ± 0.0003	1.6616 ± 0.0449	0.6933 ± 0.0001	0.7945 ± 0.0634

https://doi.org/10.1121/10.0003755

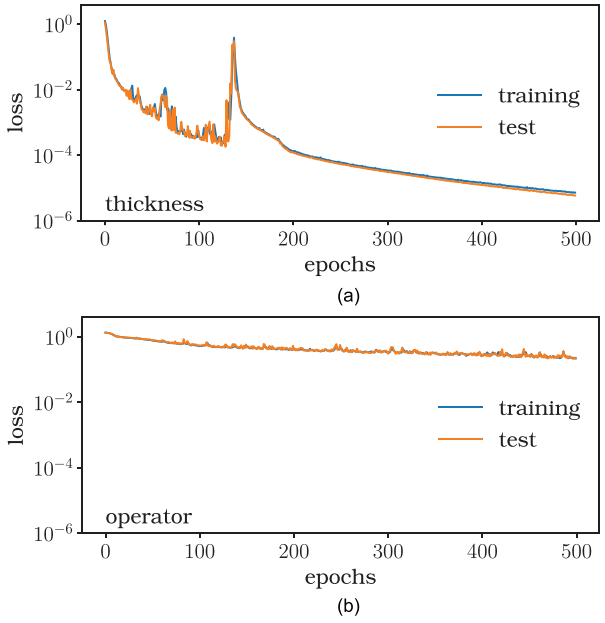

FIG. 11. (Color online) Training process of the final models for the thickness factor (a) and the operator factor (b). The categorical cross-entropy loss is depicted for the training set and the test set.

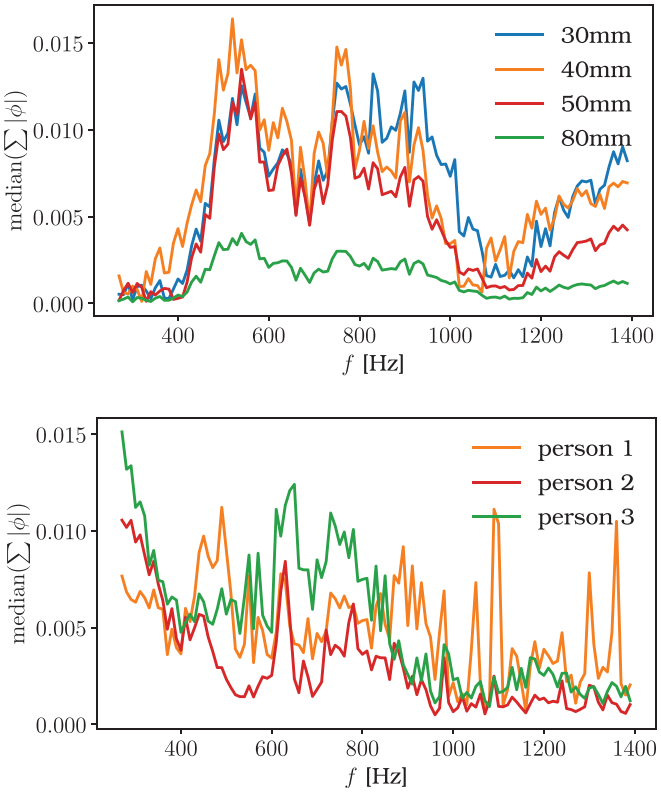

FIG. 12. (Color online) SHAP feature importance values for the reduced frequency range models.

The first hidden layer is composed of $n_f/2$ neurons, the second hidden layer carries 50 neurons, and the third layer has 25 neurons. Rectified linear unit (ReLu) activation functions are used in the hidden layers.

b. Configuration of $\mathcal{M}_2$

The number of neurons per hidden layer (again using ReLu activation) is a cascade of $(\frac{1}{2})$, such that the first hidden layer has $n_f/2$ neurons, the second layer has $n_f/4$ neurons, and so forth until a size of 30 is reached. A dropout layer (dropout rate 0.2) is introduced after each hidden layer, and a final hidden layer with 25 neurons is set before the output layer with softmax activation. Hence, the number of hidden layers depends on the input dimension n_f.

For the original sampling rate $n_f = 1960$, the two models have $\approx 5,815,800$ ($\mathcal{M}_1$) and $\approx 6,405,500$ ($\mathcal{M}_2$) trainable parameters (the actual number of trainable parameters depends on the number of classes m_k, i.e., the size of the output layer). Depending on the number of classes m_k, the loss functions are binary cross-entropy ($m_k = 2$) and categorical cross-entropy ($m_k > 2$). The softmax activation function of the output layer ensures that the output values represent a normalized probability distribution within $(0, 1)$, such that the sum of outputs equals unity, corresponding to the one-hot encoded target values. Both classifiers $\mathcal{M}_{1,2}$ are trained for 50 epochs using Adam (stochastic gradient descent with adaptive moment estimation) (Kingma and Ba, 2017) with a learning rate of 0.001 and a batch size of 64 for the CV study. The final models were trained for 500 epochs using a

TABLE V. Truncated data set for frequencies below $f < 1400$ Hz: Performance of the ML models $\mathcal{M}_1$ and $\mathcal{M}_2$ for the univariate classification task. The mean fivefold CV test set accuracy score and the standard deviation are reported for both models, each dimensionality reduction rate, and each factor.

	Thickness	Diameter	Cutting	Mounting	Operator
No dimensionality reduction, $n_f = 1960$					
$\mathcal{M}_1$ MLP	1.0±0.0	0.32±0.01	0.32±0.04	0.49 ± 0.01	0.85±0.05
$\mathcal{M}_2$ deep MLP	0.98±0.02	0.32 ± 0.01	0.20±0.01	0.49 ± 0.01	0.47±0.06
Moderate dimensionality reduction, $n_f = 980$					
$\mathcal{M}_1$ MLP	1.0±0.0	0.32 ± 0.01	0.35±0.05	0.49±0.01	0.84 ± 0.04
$\mathcal{M}_2$ deep MLP	1.0 ± 0.0	0.32 ± 0.01	0.24±0.01	0.49±0.01	0.66±0.02
Strong dimensionality reduction, $n_f = 196$					
$\mathcal{M}_1$ MLP	1.0 ± 0.0	0.54±0.02	0.33 ± 0.03	0.49 ± 0.01	0.84 ± 0.04
$\mathcal{M}_2$ deep MLP	1.0 ± 0.0	0.33 ± 0.02	0.29 ± 0.02	0.49±0.01	0.74±0.03

batch size of 256. No overfitting was observed during the training of the various models.

3. ML metrics

The classification performance of the classifiers is measured by the *accuracy* metric, which is computed as follows:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{A1}$$

where TP denotes true positives, TN denotes true negatives, FP denotes false negatives, and FN denotes false negatives.

4. Training process

Table IV lists the test loss values form the cross-validated model selection process in analogy with Table III in Sec. III B.

Figure 11 depicts the training process in terms of the training and test set loss of the final models discussed and analyzed in Sec. IV A. The loss for the thickness factor decreases rapidly to values in the range of 10^{-5} after 500 training epochs, while the decay is much slower to 10^{-1} for the operator factor. In both cases, no separation of training and test curve can be observed, i.e., indicating that no model is overfitting.

APPENDIX B

1. Reduced frequency range $f < 1400$ Hz

Table V shows all results for a truncated data set that contains the inputs $\alpha(f)$ within the frequency range $f < 1400$ Hz, as this range was found to be of major importance. It turns out that the model performance is qualitatively very similar to that of the models using the full frequency range as input, which demonstrates the robustness of the proposed ML models.

The SHAP values show a similar qualitative result for thicknesses of 30, 40, and 50 mm; cf. Figs. 6 and 12 (top panel). However, the reduced frequency range causes the SHAP value for 80 mm thickness to be less pronounced than the SHAP value for the full frequency range. The reduced frequency range seems to be less decisive for classifying the absorption coefficient into the 80 mm thickness class. This indicates that other factors and—in particular—interactions of factors, which have not been captured by the univariate model, are also having an effect on the absorption coefficients. In case of the factor *operator*, its effect is still less clear in the reduced frequency range; cf. Figs. 8 and 12 (bottom panel). For example, the SHAP values of person 1 increase at distinct frequencies above 1000 Hz, which are not shown if the full frequency range is considered. This finding confirms that the *operator* factor seems to be highly variable. No prediction for the *pc* data was achieving the 90% confidence bound, so no SHAP values were computed.

2. Thickness-separated data sets

In a second study, the complete data set was split into four subsets, each corresponding to a unique specimen thickness. Hence, the influence of the most decisive factor, the thickness, is removed. For each subset, the cross-validated model selection was performed to analyze the effect of the other factors for $n_f = 1960$ and the reduced frequency range $f < 1400$ Hz. Table VI reports the model performance for each of the data subsets corresponding to a single thickness value. For the factors diameter, cutting, and mounting, the model performance does not change by more than 0.05 compared to the full data set reported in Table V. Also, for the operator factor, the results do not change significantly, except for the data set for 40 mm sample thickness, for which no clear relation can be found by both models. The performance drops slightly for the 50 mm specimen thickness.

TABLE VI. Thickness-separated data sets: Performance of the ML models $\mathcal{M}_1$ and $\mathcal{M}_2$ for the univariate classification task. The mean fivefold CV test set accuracy score and the standard deviation are reported for both models, each factor, and per data subset for each sample thickness value.

	Diameter	Cutting	Mounting	Operator
Data subset for sample thickness 30 mm				
$\mathcal{M}_1$ MLP	0.43±0.03	0.49±0.02	0.48 ± 0.02	0.85±0.07
$\mathcal{M}_2$ deep MLP	0.31 ± 0.04	0.32±0.03	0.48 ± 0.02	0.66±0.05
Data subset for sample thickness 40 mm				
$\mathcal{M}_1$ MLP	0.32±0.03	0.24±0.04	0.47 ± 0.01	0.24±0.02
$\mathcal{M}_2$ deep MLP	0.32 ± 0.03	0.20±0.02	0.47 ± 0.01	0.48±0.12
Data subset for sample thickness 50 mm				
$\mathcal{M}_1$ MLP	0.60±0.04	0.38±0.08	0.50 ± 0.02	0.77±0.12
$\mathcal{M}_2$ deep MLP	0.32 ± 0.01	0.20±0.01	0.49 ± 0.01	0.47±0.06
Data subset for sample thickness 80 mm				
$\mathcal{M}_1$ MLP	0.47±0.08	0.31±0.06	0.47 ± 0.01	0.81±0.05
$\mathcal{M}_2$ deep MLP	0.32 ± 0.03	0.21±0.02	0.47 ± 0.01	0.71±0.05

Allard, J., and Atalla, N. (**2009**). *Propagation of Sound in Porous Media: Modelling Sound Absorbing Materials*, 2nd ed. (John Wiley & Sons, West Sussex, UK).

Bonfiglio, P., Pompoli, F., Horoshenkov, K. V., Rahim, M. I. B. A., Jaouen, L., Rodenas, J., Bécot, F.-X., Gourdon, E., Jaeger, D., Kursch, V., Tarello, M., Roozen, N. B., Glorieux, C., Ferrian, F., Leroy, P., Vangosa, F. B., Dauchez, N., Foucart, F., Lei, L., Carillo, K., Doutres, O., Sgard, F., Panneton, R., Verdiere, K., Bertolini, C., Bär, R., Groby, J.-P., Geslain, A., Poulain, N., Rouleau, L., Guinault, A., Ahmadi, H., and Forge, C. (**2018**). "How reproducible are methods to measure the dynamic viscoelastic properties of poroelastic media?," J. Sound Vib. **428**, 26–43.

Cummings, A. (**1991**). "Impedance tube measurements on porous media: The effects of air-gaps around the sample," J. Sound Vib. **151**(1), 63–75.

Fulcher, B. D., and Jones, N. S. (**2017**). "*hctsa*: A computational framework for automated time-series phenotyping using massive feature extraction," Cell Syst. **5**(5), 527–531.e3.

Grebel, A. (**2020**). "Unsicherheiten bei der bestimmung der schallabsorption mithilfe des kundt'schen rohres" ("Uncertainty in absorption measurements in impedance tubes"), Ph.D. thesis, Technical University of Darmstadt, Darmstadt, Germany.

Guyon, I., and Elisseeff, A. (**2003**). "An introduction to variable and feature selection," J. Mach. Learn. Res. **3**, 1157–1182.

Horoshenkov, K. V., Khan, A., Bécot, F.-X., Jaouen, L., Sgard, F., Renault, A., Amirouche, N., Pompoli, F., Prodi, N., Bonfiglio, P., Pispola, G., Asdrubali, F., Hübelt, J., Atalla, N., Amédin, C. K., Lauriks, W., and Boeckx, L. (**2007**). "Reproducibility experiments on measuring acoustical properties of rigid-frame porous media (round-robin tests)," J. Acoust. Soc. Am. **122**(1), 345–353.

Inoue, N., and Sakuma, T. (**2017**). "Numerical investigation of effect of support conditions of poroelastic materials in impedance tube measurement," Acoust. Sci. Technol. **38**(4), 213–221.

ISO 10534-2 (**1998**). "Determination of sound absorption coefficient and impedance in impedance tubes—part 2: Transfer-function method," International Organization for Standardization, Geneva, Switzerland.

ISO 18437-5:2011 (**2011**). "Mechanical vibration and shock—characterization of the dynamic mechanical properties of visco-elastic materials—part 5: Poisson ratio based on comparison between measurements and finite element analysis," International Organization for Standardization, Geneva, Switzerland.

Kang, Y. J., and Bolton, J. S. (**1995**). "Finite element modeling of isotropic elastic porous materials coupled with acoustical finite elements," J. Acoust. Soc. Am. **98**(1), 635–643.

Kellner, L., Stender, M., von Bock und Polach, R. U. F., Herrnring, H., Ehlers, S., Hoffmann, N., and Høyland, K. V. (**2019**). "Establishing a common database of ice experiments and using machine learning to understand and predict ice behavior," Cold Regions Sci. Technol. **162**, 56–73.

Kingma, D. P., and Ba, J. (**2017**). "Adam: A method for stochastic optimization," arXiv:1412.6980.

Kino, N., Nakano, G., and Suzuki, Y. (**2012**). "Non-acoustical and acoustical properties of reticulated and partially reticulated polyurethane foams," Appl. Acoust. **73**(2), 95–108.

Kino, N., and Ueno, T. (**2007**). "Investigation of sample size effects in impedance tube measurements," Appl. Acoust. **68**(11), 1485–1493.

Lipton, Z. C. (**2018**). "The mythos of model interpretability," Queue **16**(3), 31–57.

Lundberg, S. M., and Lee, S.-I. (**2017**). "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30*, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., Red Hook, NY), pp. 4765–4774.

Montgomery, D. (**2009**). *Design and Analysis of Experiments*, 7th ed. (John Wiley & Sons, Hoboken, NJ).

Pilon, D., Panneton, R., and Sgard, F. (**2003**). "Behavioral criterion quantifying the edge-constrained effects on foams in the standing wave tube," J. Acoust. Soc. Am. **114**(4), 1980–1987.

Pilon, D., Panneton, R., and Sgard, F. (**2004**). "Behavioral criterion quantifying the effects of circumferential air gaps on porous materials in the standing wave tube," J. Acoust. Soc. Am. **116**(1), 344–356.

Pompoli, F., Bonfiglio, P., Horoshenkov, K. V., Khan, A., Jaouen, L., Bécot, F.-X., Sgard, F., Asdrubali, F., D'Alessandro, F., Hübelt, J., Atalla, N., Amédin, C. K., Lauriks, W., and Boeckx, L. (**2017**). "How reproducible is the acoustical characterization of porous media?," J. Acoust. Soc. Am. **141**(2), 945–955.

Ribeiro, M. T., Singh, S., and Guestrin, C. (**2016**). "Why should I trust you?," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD '16*, edited by B. Krishnapuram, M. Shah, A. Smola, C. Aggarwal, D. Shen, and R. Rastogi (ACM Press, New York), pp. 1135–1144.

Seybert, A. F., Hua, X., and Herrin, D. W. (**2013**). "Controlling uncertainty of sound absorption measurements using the impedance tube method," in Proceedings of Inter*n*oise 2013, September 15–18, Innsbruck, Austria, p. 217.

Song, B. H., and Bolton, J. S. (**2003**). "Investigation of the vibrational modes of edge-constrained fibrous samples placed in a standing wave tube," J. Acoust. Soc. Am. **113**(4), 1833–1849.

Song, B. H., Bolton, J. S., and Kang, Y. J. (**2001**). "Effect of circumferential edge constraint on the acoustical properties of glass fiber materials," J. Acoust. Soc. Am. **110**(6), 2902–2916.

Stanley, D. R. (**2012**). "Impedance tube specimen preparation and mounting issues," in Proceedings of Inter*n*oise 2012, August 19–22, New York, p. 1360.

Stender, M., Adams, C., Wedler, M., Grebel, A., and Hoffmann, N. (**2020**). "Code accompanying the paper: Explainable machine learning determines effects on the sound absorption coefficient measured in the impedance tube," https://github.com/TUHH-DYN/ExplainableML_acoustics (Last viewed March 9, 2021).

Stone, M. (**1974**). "Cross-validatory choice and assessment of statistical predictions," J. R. Stat. Soc. Series B Stat. Methodol. **36**(2), 111–133.

Tsay, H.-S., and Yeh, F.-H. (**2006**). "The influence of circumferential edge constraint on the acoustical properties of open-cell polyurethane foam samples," J. Acoust. Soc. Am. **119**(5), 2804–2814.

Vigran, T. E., Kelders, L., Lauriks, W., Leclaire, P., and Johansen, T. (**1997**). "Prediction and measurements of the influence of boundary conditions in a standing wave tube," Acta Acust. United Acust. **83**(3), 419–423.

Wenzel, S., Slomski, E. M., Adams, C., and Melz, T. (**2020**). "Bestimmung der Einflüsse von Störgrößen auf die Absorptionsgradmessung im Kundtschen Rohr mithilfe neuronaler Netze" ("Determination of the influence of disturbance variables on the absorption coefficient measurement in the Kundt's tube using neural networks"), in Proceedings of DAGA 2020—46. Jahrestagung Für Akustik, DEGA, Berlin, pp. 551–554.