Acoustic monitoring using PyzoFlex®: a novel printed sensor for smart consumer products

M Blass¹, F Krebs¹, C Amon¹, M Adler², M Zirkl², A Tschepp² and F Graf¹
¹Joanneum Research, DIGITAL, Graz, Austria
²Joanneum Research, MATERIALS, Weiz, Austria
E-mail: martin.blass@joanneum.at

Abstract. Acoustic monitoring has always been a niche area in the field of monitoring applications compared to other modalities, such as computer vision. Over the last decades, the number of applications for acoustic monitoring has been growing and ranges from predictive maintenance within the industrial sector to acoustic scene classification and security monitoring in traffic and urban scenarios. With the rise of the internet-of-things (IoT) and artificial intelligence (AI) in recent years, smart consumer products and devices have pushed forward using different sensor technologies to enhance the user experience. To this end, acoustic monitoring is still an underestimated discipline with great potential to serve as a missing link in smart sensing within environments where other modalities face difficulties. In this paper, we present PyzoFlex®, a printable sensor technology which facilitates accurate measurement of pressure and temperature changes in objects and their environment, as a sensor interface for acoustic monitoring applications. In contrast to microphones or acceleration sensors, PyzoFlex may be printed onto any curved or textured surface. To demonstrate the possibilities, we present a case study in which we equip a coffee machine with PyzoFlex to acoustically monitor the machine states in real-time using a machine learning model.

1. Introduction
Acoustic monitoring has always been a niche compared to visual monitoring applications. However, over the past years, the number of applications for acoustic monitoring has been growing, especially in areas where acoustic sensors are either advantageous over other sensor modalities, or complement them. Among others, typical applications include condition monitoring in manufacturing [1, 2, 3], structural health monitoring [4, 5] and quality control [6, 7], monitoring of road [8, 9, 10, 11] and rail traffic [12, 13], medical monitoring [14, 15], security monitoring [16, 17, 18] as well as acoustic sensing [19] and voice assistants [20]. The range of acoustic sensors used within these applications ranges from various types of microphones to vibration sensors, which capture airborne sound and structure borne sound, respectively.

In the field of acoustic monitoring and sensing, vibration sensors, such as accelerometers, use the piezo-electric effect to transform mechanical vibrations of a physical device into electrical voltage. After amplifying and converting analog signals to the digital domain, signal processing methods are used to analyze their temporal and spectral content. In addition, machine learning opens up possibilities for pattern recognition, classification and prediction by learning from recorded data.
These techniques are well-established in the industrial sector but usually rely on high-quality sensors and expensive data acquisition hardware. With the rise of the internet-of-things (IoT) and artificial intelligence (AI) as well as low-cost and low-power sensors in recent years, it has now become viable to also equip smart consumer products with sensors and artificial intelligence.

One research field with high growth rates is printable electronics. The aim is to create electronics in completely new form factors by adapting existing printing processes for functional, electronic materials, thereby opening up completely new fields of applications. Special inks with different electrical properties are printed onto large-area and flexible substrates such as plastic films, paper or textiles. The basic prerequisite for the success of printed electronics is that many materials with different electrical properties (conductive, semiconductive or insulating) can be brought into a printable form. Furthermore, functional materials with electroluminescent, sensory or energy generating properties can be realized. Since organic materials (e.g. organic polymers) can be used for printing processes, their mechanical properties allow the development of flexible electronics. In contrast to conventional microelectronics, printed electronics is characterized by simpler, more cost-effective and large-area mass production. This results in a high potential for a broader penetration in everyday life and the integration of simple electronics into everyday objects.

In this paper, we present PyzoFlex [21, 22] as a passive low-cost printed sensor technology for consumer products to enhance them with new functionalities and intelligent sensing capabilities. The novelty of this work lies in the use of PyzoFlex as an acoustic vibration sensor as the front-end of a machine learning model capable of determining multiple system states. Unlike conventional vibration sensors, PyzoFlex may be applied directly onto any curved or textured surface of a device by adhesive bonding or ink-printing over a desired area. We present a case study in which we equip a coffee machine with PyzoFlex to monitor the machine states in real-time using different machine learning models. The results indicate, that PyzoFlex is well-suited for being used as a vibration sensor for acoustic monitoring purposes and yields performances comparable to high-quality acceleration sensors.

The paper is organized as follows: Section 2 describes PyzoFlex as printed sensor technology together with its mechanical and acoustic properties as well as possible applications. In Section 3, we first outline the experimental setup for the case study of acoustic coffee machine monitoring. We then describe the machine learning models used together with the choice of audio features, classification algorithms and model parameters. The remainder of Section 3 presents results of the classification performance evaluation of PyzoFlex compared to other sensors for two different models. Section 4 concludes this work and gives a future perspective.

2. PyzoFlex

A novel and promising sensor technology from the field of printed organic electronics is the PyzoFlex® [21, 22] technology, which has a very extensive and complex portfolio ranging from a variety of sensor applications over energy harvesting to actuation, due to the pyro- and piezoelectric properties of the sensor material. By using screen-printing, sensors can be manufactured very cost-effectively on versatile surfaces such as foils, paper or glass.

2.1. Sensor Technology and Production Process

The standard substrate of the sensor foils is a transparent plastic substrate made of PET, but other materials such as glass, TPU or paper are also conceivable. In a first step, polymer based electrodes (PEDOT:PSS) are printed. Then the functional areas are coated with the functional or active material P(VDF:TrFE). In the third step, a second electrode layer is printed on top of the functional ink. The base and cover electrodes are directly on top of each other forming a plate capacitor. Mechanical excitation of the ferroelectric layer generates an electric current due to charge separation between the two electrodes. Finally, silver conductor tracks
and an encapsulation layer are printed to finalize the sensor patches. A great advantage of this manufacturing process is that the printing process does not require any elaborate requirements (clean room environment, vacuum, evaporation, high temperatures), thus enabling a cost-effective, industrial production. The piezo- and pyroelectric properties of the sensor are activated by a poling process. This process aligns the dipoles in the material, which are randomly orientated after the printing process, vertically to the sensor electrodes.

### 2.2. Material Properties

Organic ferroelectric polymers from the PVDF class (polyvinylidene fluoride) are used as the material basis for such sensors since they exhibit strong piezo- and pyroelectric activity as well as high chemical robustness and UV resistance. The use of organic materials represents a key factor in the flexibility of the technology, since such piezo- and pyroelectric properties are otherwise only known from crystalline or ceramic materials, which are rigid and brittle due to their crystalline structure. The PyzoFlex sensor foils are based on the ferroelectric copolymer P(VDF-TrFE), which shows a piezoelectric and pyroelectric behavior upon so-called electrical poling [23]. As a result of this poling step, the ferroelectric layer possesses a remnant polarization $P_r$ with an out-of-plane orientation, i.e. normal to the electrodes’ surfaces. The value of the remnant polarization is obtained directly from the poling hysteresis and serves as a benchmark value for the piezo- and pyroelectric response of the sensor layer. A change in the out-of-plane strain component directly leads to a change in the polarization, which can be measured as a charge response $Q$. This relation between charge and strain is described by the piezoelectric constant $e_{33}$ as follows [24, 25]:

$$e_{33} = \frac{-Q}{A} \cdot \frac{1}{s_{33}} = -a \cdot P_r$$  \hspace{1cm} (1)

where $A$ is the interface area between the electrode and the ferroelectric material, $s_{33}$ is the average out-of-plane strain component present in the piezoelectric layer (the out-of-plane or thickness direction is indexed with 1), and $a$ is a material constant close to unity.

The capacitive layer structure of the sensor elements - the active layer of P(VDF-TrFE) is located as dielectric between the conductive base and cover electrodes - leads to a typical impedance curve as shown in Figure 2. The inset within this figure shows the equivalent circuit of the printed sensor, where $R_P$ is the film resistance and $C_P$ is the internal film capacitance (Figure 1). The series resistances $R_S$ take into account the resistances of the supply lines, which, in the case of printed materials, can vary between a few ohms and kilohms, depending on the material used and the length of the supply lines. The induced charge $Q$ is linearly proportional to the applied force as described earlier. The capacitance $C_P$ is proportional to the surface area of film and is inversely proportional to the film thickness. In low frequency applications, the internal film resistance $R_P$ is very high and can be ignored. The open circuit output voltage can be found from the film capacitance; i.e., $V_{\text{OpenCircuit}} = Q/C$.

For these measurements a sensor with an active area of 3.14 cm$^2$ (circular, with 2 cm diameter) was used. To cover the measurement range from 5 mHz to 5 MHz, the measurements were combined with two independent systems, a impedance spectrometer POT/GAL 30 V 2A from Novocontrol and a Hioki 3532-50 LCR-meter.

### 2.3. Acoustic Properties

Techniques for the acoustic characterization of thin films piezoelectric sensors typically cover sensitivity, frequency response, dynamic range, resonance frequency and the piezoelectric coefficient [26]. The difference of PyzoFlex with respect to conventional acoustic sensors, e.g. piezoelectric MEMS microphones, is that it may be used to capture airborne as well as structure...
Figure 1. Characterization of a printed PyzoFlex sensor in terms of capacity over frequency.

Figure 2. Characterization of a printed PyzoFlex sensor in terms impedance over frequency.
borne sound. For this reason, we investigate the acoustic properties of PyzoFlex in two different scenarios:

(a) mounted on a solid surface (structure borne sound),
(b) fixed into a frame as membrane (airborne sound).

Since we are mainly interested in using PyzoFlex as sensor input for a machine learning system, we only evaluate frequency responses and omit other acoustic metrics since this would go beyond the scope of this work. Frequency responses are evaluated for two printed PyzoFlex sensor samples with different types and thicknesses of material, each patch with a dimension of 9 x 2 cm:

(i) **PYZO-1**: PET substrate with a thickness of 50 µm,
(ii) **PYZO-2**: Paper strip with a thickness of 125 µm,

The measurements for both scenarios are carried out in an acoustic laboratory (5.5 x 6.3 x 2.4 m), which fulfills the *ITU-R BS 1116-1* recommendation suited for objective listening tests and acoustic measurements. As signal conditioning and data acquisition hardware we use a *GRABAU ICP-A10* adapter and a *Focusrite Scarlett 4i4 (3rd Generation)* USB audio interface.

### 2.3.1. Vibration Analysis

For scenario (a) we mount the PyzoFlex samples onto a wooden floor by applying adhesive tape to the sensor edges. In order to compare PyzoFlex to a conventional accelerometer, we attach a *PCB Model 356B18* high-sensitivity ceramic shear accelerometer (ACC) as a high-quality vibration sensor reference to the floor (Figure 3). We then create a persistent excitation of 10 Hz impulses using an electronic hammer (*Stratenschulte Messtechnik MIDI-Hammerwerk*) located 1 cm from each sensor.

![Figure 3. Measurement setup for acquiring vibration signals and performing frequency response analysis: PyzoFlex sensor samples (PYZO-1, PYZO-2) are attached to a wooden floor. A reference accelerometer (ACC) is used to compare vibration signal responses.](image)

We compute the frequency response as the average power spectral density (PSD) from a recording of 9 seconds with a sampling rate of 48 kHz, a frame size of 0.1 seconds, 8192 FFT points and a Hann window. Results are smoothed using a Gaussian window over 250 FFT bins and visualized in the frequency range from 20 Hz to 20 kHz in Figure 4. **PYZO-2** shows a similar frequency response compared to the reference ACC, whereas **PYZO-1** tends to be more sensitive above 300 Hz. All sensor signals were recorded with a gain of 56 dB.
2.3.2. Aeroacoustic Analysis

In scenario (b) we fit PyzoFlex samples within a frame of acoustic foam material to simulate a setup where the sensor is used as a membrane for capturing airborne sound (Figure 5). As a reference sensor we place a PCB Model 130F20 measurement microphone (MIC) above the foam frame and operate it at a gain of 30 dB. Gain settings for PyzoFlex sensors are equal to Section 2.3.1. As signal source we use a linear frequency sweep of 20 Hz to 20 kHz with a duration of 3 s. The signal is played from a Neumann K&H O300 loudspeaker at a distance of 0.5 m at a sound pressure level of 90 dB(A) and repeated 3 times resulting in an input signal of 9 s for measuring the acoustic frequency response. Both sensor and loudspeaker were positioned at a height of 1 m in the room center.

Figure 5. Measurement setup for capturing airborne sound and performing frequency response analysis: PyzoFlex sensor samples (PYZO-1, PYZO-2) are fitted into an acoustic foam frame for acting as a membrane and compared to a reference measurement microphone (MIC).
For measuring frequency responses we choose the same signal processing setup as for the vibration analysis (Section 2.3.1) apart from frame size and FFT size, which are set to 3 seconds and $2^{19}$, respectively. The frequency responses of the aeroacoustic measurement for PyzoFlex samples and the reference microphone are shown in Figure 6. In contrast to MIC, it becomes obvious that the used sensor samples PYZO-1 and PYZO-2 do not provide a linear frequency response. This may be attributed to material properties or to a mounting setup which induces specific vibration modes on the sensor surface.

![Figure 6](image_url). Frequency responses of a linear frequency sweep signal for two different PyzoFlex sensor samples (PYZO-1, PYZO-2) and reference measurement microphone (MIC).

3. Evaluation

As a practical example of how to use PyzoFlex for monitoring applications in consumer products, we equipped a conventional coffee machine with a patch of PyzoFlex and trained machine learning models to discriminate between different machine states during the coffee making process. The goal of this study is to investigate whether PyzoFlex is suitable for being used as a vibration sensor within acoustic monitoring applications.

3.1. Experimental setup

3.1.1. Hardware

The device under test is a Krups EA8108 fully automatic coffee machine, which we equipped with two sensors on the right side of its corpus: a rectangular patch of PyzoFlex with dimensions of 9 x 2 cm printed onto 175 µm of PET substrate (PYZO) and a PCB Model 356B18 high-sensitivity triaxial accelerometer (ACC) as vibration sensor reference as used in acoustic measurements of Section 2.3. Both sensors are driven at a gain of 56 dB and placed next to each other in order to minimize spatially related differences in the vibration signals. In addition, we placed a PCB Model 130F20 measurement microphone (MIC) at a distance of 2 m to capture the airborne sound signature of the coffee machine as a reference to structure borne sound. For recording the sensor signals, we used a Focusrite Scarlett 18i8 (3rd Generation) USB audio interface and specified a sampling frequency of 48 kHz and a resolution of 32 bit for recordings in the .wav file format. The full recording setup is shown in Figure 7.
Figure 7. Recording setup for acquiring audio data of the coffee machine monitoring system.

3.1.2. Data

For each recording the true reference of coffee machine states was manually annotated by labeling the respective time segments. Annotations describe three distinct states in the coffee-making process:

(i) move: the sound of internal movements of mechanical components (e.g. water pump)
(ii) grind: the sound of grinding coffee beans
(iii) brew: the sound of brewing coffee

In total, 68 minutes of audio data was recorded while the coffee machine was running. During this time, occasional background noise, speech as well as haptic interactions with the coffee machine were present and manually labeled as noise. After the annotation process, the acquired audio data was split into a training and a test dataset on file recording level, i.e. keeping the chronological sequence of coffee-making states. The size of the training and test set is 38 and 30 minutes respectively. Since the duration of states within one coffee-making cycle do not differ to a great extent, the distribution of states is similar in the training and test dataset. Dataset distributions are visualized in Figure 8.

Figure 8. Dataset distribution of coffee machine audio data: percentage of coffee machine states and noise (left) and absolute duration of states (right) for training and test data, resp.
3.2. Acoustic Monitoring System

The acoustic monitoring system in this paper infers the current state of the coffee machine from the audio signal (structure borne or airborne) using a classifier. For each sensor type, we compare two classifier models, which were created by different machine learning design principles: manual feature design (Model 1) vs. automatic feature learning (Model 2).

3.2.1. Model 1

The first model consists of a manually designed feature extraction stage and a Random Forest (RF) classifier. We create three independent RF models (one per coffee machine state) and evaluate them in a multi-class classifier setup.

Features

A discrete set of audio features is chosen to numerically describe the sound characteristics of the respective sound classes. These features can contain generic low-level descriptors [27, 28], as well as custom features to accurately represent the physical properties of the underlying mechanism. This kind of feature engineering marks the traditional approach of machine learning systems and can be a time-consuming and complex procedure depending on the application. However, a set of well developed and selected features may result in a robust system, which performs well in cases where few training data is available and where it is possible to properly model the underlying physical properties of a sound class. In the following, we describe our manual feature-based approach for classifying coffee machine states, which employs

- an appropriate subset of acoustic low-level descriptors,
- template matching using the similarity of averaged power spectra,
- spectral peak tracking and derived custom high-level features.

Prior to feature extraction, we compute the Short-Time Fourier Transform (STFT) of the vibration signals using a Hann window with a frame size of 2048 samples for the models **move** and **grind** and 16384 samples for the model **brew**. The hop size is 1024 samples, which yields a feature frame rate of 46.875 fps. The choice of using a greater frame size for the **brew** features is based on the requirement of a better frequency resolution in lower frequencies for this sound class (see right plot in Figure 9). FFT sizes used match the chosen frame sizes.

![Figure 9](image)

**Figure 9.** Spectral templates as normalized power spectra for sound signatures of coffee machine states: **move** (left), **grind** (middle), **brew** (right).

For all three RF models we use the same set of audio features, however with slightly different parametrization depending on the dominant frequency range of each sound class. First, we include the following low-level descriptors: zero-crossing rate, spectral centroid and flatness, and 10 Mel-Frequency Cepstral Coefficients (MFCC). Next, we perform a spectral template matching by computing a similarity score of power spectra to an averaged spectrum template which was
precalculated for each sound class (Figure 9). As similarity score we choose the Kullback-Leibler (KL) divergence and handle power spectra as probability distributions by normalizing w.r.t. the summed power.

Finally, we apply a spectral peak tracking over time similar as in [29]: After peak picking in the power spectrum of a current frame, a fixed number of peaks are selected based on their prominence and tracked using a low-complexity association strategy [30]. From the resulting peak tracks at each time frame instance we derive custom high-level peak track features. The complete description of the 38 manually selected features is shown in Table 1.

| Feature | Size | Description |
|---------|------|-------------|
| sCent  | 1    | Spectral Centroid: frequency indicating center of spectrum mass |
| sFlat  | 1    | Spectral Flatness: flatness of the spectral envelope |
| ZCR    | 1    | Zero-Crossing Rate: zero-crossings per time frame |
| MFCC   | 10   | Mel-Frequency Cepstral Coefficients of 100 Mel-bands |
| KLDIV  | 1    | Kullback-Leibler Divergence: similarity of spectrum and spectral template |
| pSize  | 1    | Peak Track Size: number of tracked spectral peaks |
| pImpc  | 1    | Peak Track Impact: ratio of peak track energy to total spectrum energy |
| pFluc  | 1    | Peak Track Fluctuation: weighted std. deviation of peak freq. over time |
| pStab  | 1    | Peak Track Stability: mean presence of peak tracks within 0.5 s |
| pCent  | 1    | Peak Track Centroid: frequency indicating center of peak track mass |
| pSprd  | 1    | Peak Track Spread: weighted std. deviation of peak track frequencies |
| pWght  | 12   | Peak Track Weight: weights computed from peak stability and energy |
| pRtio  | 3    | Peak Track Ratio: relative difference of peak indices of adjacent peaks |
| pDiff  | 3    | Peak Track Difference: abs. difference of peak indices of adjacent peaks |

Classification
For the feature engineering approach discussed in section 3.2.1 we choose a frame-based Random Forest (RF) classifier from the machine learning toolbox perClass [31] together with a temporal smoothing of confidence outputs. The RF was parametrized with 30 trees, a maximum number of 10000 nodes and 20% randomly selected features at each tree stage to split a node. Confidence smoothing as post-processing was implemented using a first-order IIR filter, also known as exponential smoothing, with a forgetting time constant of 0.1 seconds.

3.2.2. Model 2
The second model classifies the coffee machine states from a spectrogram input using a recurrent neural network (RNN). We train one model with four outputs corresponding to the four possible sound classes move, grind, brew, and noise.

Features
Although neural network classifiers are able to learn from raw audio input [32], most current audio classifiers use variants of spectrograms as input [33, 34, 35, 36]. In our model we use mel frequency filtered spectrograms, extracted with an FFT size of 4096 and a hop size of 1024 samples to match the feature frame rate of Model 1. We reduce the FFT bins to 80 frequency bands between 50 and 3000 Hz with a mel filterbank. For amplitude compression we add a constant of one to the mel spectrogram magnitudes (to ensure positive values) and apply the decadic logarithm.
Classification
Neural networks have been employed successfully in a huge amount of audio and image related tasks. Current models consist of recurrent layers [33], convolutional layers [35] or combinations of both types [34, 36]. In this paper, we target a real-time application with a model that could equally be executed on a microcontroller with limited computational resources. Therefore, we use a lightweight model which only consists of two layers with each 30 GRU [37] cells. Fully recurrent models have the advantage that only the features of the current time frame have to be kept in memory and therefore have a low memory demand. The model used in this paper consists of only 15,784 trainable parameters. The last layer of the model is a softmax activation over the four classes output. The model is trained with categorical cross-entropy loss and Adam optimizer [38] until the model performance stops improving on a hold out validation dataset for more than 20 epochs. The validation set was randomly sampled from the training set and contains seven minutes of audio data.

3.3. Results and Discussion
We evaluate the two models of our acoustic coffee monitoring system in terms of confusion matrices and macro-average classification scores over all sound classes [39]. This evaluation is carried out for each of the three sensors: PYZO, ACC and MIC. Furthermore, we examine the real-time performance in terms of CPU load on a standard notebook PC.

3.3.1. Machine State Classification
A detailed summary of the overall classification performance is given in Table 2 in terms of macro-average precision, recall and F1-scores. The obtained results show that both PYZO and ACC achieve similar performances, whereas MIC suffers from classification errors, mainly attributed to background noise (false negatives). In addition, we can see that both classification models obtain similar classification scores, except for the MIC signal, where Model 1 achieves a slightly better performance.

A more detailed view is given in Figures 10 and 11 which present classification results as confusion matrices of Model 1 (RF) and Model 2 (RNN). Interestingly, Model 2 always outperforms Model 1 in detecting grind but always performs worse than Model 1 in detecting brew. This might be partly due to the fact that Model 1 has a better frequency resolution in the brew features. The difference between Model 1 and Model 2 becomes particularly clear with the MIC sensor: Model 2 has difficulties to distinguish move and brew from noise, while Model 1 struggles with discriminating grind and noise.

| Sensor | RF Precision | Recall | F1-Score | RNN Precision | Recall | F1-Score |
|--------|--------------|--------|----------|---------------|--------|----------|
| PYZO   | 0.9386       | 0.9673 | 0.9542   | 0.9324        | 0.9688 | 0.9508   |
| ACC    | 0.9386       | 0.9726 | 0.9550   | 0.9361        | 0.9742 | 0.9544   |
| MIC    | 0.8953       | 0.8687 | 0.8811   | 0.8809        | 0.8178 | 0.8464   |

3.3.2. Real-Time Performance
For the real-time implementation of the acoustic coffee machine monitoring system we used MATLAB together with Simulink [40, 41] to generate an executable in C++. At runtime the most computationally expensive part of the RF classifier is the spectral peak tracking, whereas for the RNN computing activations of GRU cells requires most of the processing time. In order to
reduce the computational load of the tested systems, we may reduce general model complexity, e.g. by using less neurons or hidden layers in the RNN, or by improving code efficiency. Using the feature and model parameters as discussed in Sections 3.2.1 and 3.2.2, we arrive at an average CPU load of 0.5% and 0.35% on a Windows 10 PC with an Intel Core i7 6820HQ CPU clocked at 2.7 GHz for the RF and the RNN model, respectively. A screenshot of the system’s user interface showing classifier confidences as interactive bars is depicted in Figure 12.

**Figure 10.** Predictions of Model 1 (RF) visualized as confusion matrices for each sensor: PYZO (left), ACC (middle), MIC (right). Prediction scores are displayed as rates, i.e. normalized by the number of true data samples.

**Figure 11.** Predictions of Model 2 (RNN) visualized as confusion matrices for each sensor: PYZO (left), ACC (middle), MIC (right). Prediction scores are displayed as rates as in Figure 10.

**Figure 12.** Screenshot of the coffee machine monitoring user interface at runtime.
4. Conclusion
In this work we evaluated PyzoFlex as a sensor frontend in an acoustic monitoring system. First, we recorded acoustic frequency responses of two printed PyzoFlex sensor samples for structure-borne as well as airborne sound sources and compared them to the frequency responses of high-quality measurement sensors. In the structure borne sound scenario PyzoFlex showed similar characteristics in terms of frequency content and sensitivity with respect to the reference accelerometer sensor. For the aeroacoustic measurement, in which we simulated PyzoFlex as membrane of a sound transducer, it became apparent that the tested sensor samples provided non-linear frequency responses and featured lower sensitivity compared to the reference measurement microphone.

Next, we tested PyzoFlex as a sensor interface within a practical acoustic monitoring application. In this proof-of-concept study we equipped a coffee machine with a PyzoFlex and a reference vibration sensor, recorded audio data while the machine was running and trained two different machine learning models to discriminate between three machine states. The evaluation shows that, although the audio quality recorded with Pyzoflex is lower than the quality of the acceleration sensor, the PyzoFlex model yields similar classification performance compared to the high-quality accelerometer reference. This observation has been verified using two different, lightweight classification models including a Random Forest classifier and a Recurrent Neural Network which operate in realtime.

Our results indicate that PyzoFlex may be used as a low-cost, passive and versatile alternative to conventional vibration sensors in monitoring applications or IoT-based interfaces. The demonstrated example of acoustic coffee machine monitoring stands representative for various use cases in which consumer products may be equipped with PyzoFlex to provide them with intelligence or sensing capabilities. Future work includes the characterization of acoustic performance parameters for PyzoFlex sensors printed onto different materials to investigate the usage in high-precision vibration measurement setups and improvements for capturing airborne sound.

References
[1] Goyal D and Pabla B 2016 Archives of Computational Methods in Engineering 23 585–594
[2] Chaudhury S B, Sengupta M and Mukherjee K 2014 International journal of scientific engineering and research 2 11
[3] Baghalian A, Senyurek V Y, Tashakori S, McDaniel D and Tansel I N 2018 Journal of Nondestructive Evaluation 37 24
[4] Behnia A, Chai H K and Shiotani T 2014 Construction and Building Materials 65 282–302
[5] Amerini F and Meo M 2011 Structural health monitoring 10 659–672
[6] Aboonajmi M, Jahangiri M and Hassan-Beygi S R 2015 Journal of food processing and preservation 39 3175–3188
[7] Shevchik S A, Masinelli G, Kenel C, Leinenbach C and Wasmer K 2019 IEEE Transactions on Industrial Informatics 15 5194–5203
[8] Pucher M, Schabus D, Schallauer P, Lypetskyy Y, Graf F, Rainer H, Stadtschnitzer M, Sternig S, Birchbauer J, Schneider W et al. 2010 Multimodal highway monitoring for robust incident detection 13th International IEEE Conference on Intelligent Transportation Systems (IEEE) pp 837–842
[9] Graf F, Rattei G and Ruhleder G 2011 Giving tunnels ears: Installation of the first acoustic monitoring system for road tunnels worldwide Internationaler Tunnelkongress
[10] Thallinger G, Krebs F, Kolls E, Vertal P, Kasanický G, Neuschnied H and Ambrosch K E 2017 Near-miss accidents-classification and automatic detection First International Conference on Intelligent Transport Systems (Springer) pp 144–152
[11] Zambon G, Angelini F, Salvi D, Zanaboni W and Smiraglia M 2015 Traffic noise monitoring in the city of Milan: Construction of a representative statistical collection of acoustic trends International Congress on Sound and Vibration, ICSV 12-16 July
[12] Ngamkhanoung C, Kaewrunrue S and Costa B J A 2018 Infrastructures 3 3
[13] He M, Feng L and Zhao D 2019 Optik 181 343–350
[14] Abbasi-Kesbi R, Valipour A and Imani K 2018 Healthcare technology letters 5 7–12
[15] Hu Y, Kim E G, Cao G, Liu S and Xu Y 2014 Annals of biomedical engineering 42 2264–2277
[16] Mydlarz C, Salamon J and Bello J P 2017 Applied Acoustics 117 207–218
[17] Goetze S, Schroder J, Gerlach S, Hollosi D, Appell J E and Wallhoff F 2012 Journal of Computing Science and Engineering 6 40–50
[18] Kwon H, Berisha V and Spanias A 2008 Real-time sensing and acoustic scene characterization for security applications 2008 3rd International Symposium on Wireless Pervasive Computing (IEEE) pp 755–758
[19] Cai C, Zheng R and Hu M 2019 arXiv preprint arXiv:1901.03450
[20] Hoy M B 2018 Medical reference services quarterly 37 81–88
[21] Rendl C, Greindl P, Haller M, Zirkl M, Stadlober B and Hartmann P 2012 Pyzoflex: printed piezoelectric pressure sensing foil Proceedings of the 25th annual ACM symposium on User interface software and technology pp 509–518
[22] Zirkl M, Scheipl G, Stadlober B, Rendl C, Greindl P, Haller M and Hartmann P 2013 Pyzoflex: a printed piezoelectric pressure sensing foil for human machine interfaces Organic Field-Effect Transistors XII; and Organic Semiconductors in Sensors and Bioelectronics VI vol 8831 (International Society for Optics and Photonics) p 883124
[23] Stadlober B, Zirkl M and Irimia-Vladu M 2019 Chemical Society Reviews 48 1787–1825
[24] Furukawa T and Seo N 1990 Japanese Journal of Applied Physics 29 675
[25] Schaffner F and Jungueck B 1994 IEEE transactions on dielectrics and electrical insulation 1 553–562
[26] Ali W R and Prasad M 2020 Sensors and Actuators A: Physical 301 111756
[27] Peeters G 2004 CUIDADO IST Project Report 54 1–25
[28] Eyben F 2015 Real-time speech and music classification by large audio feature space extraction (Springer)
[29] Gerber T, Martin N and Mailhes C 2015 Industrial Electronics, IEEE Transactions on 62 6616–6626
[30] McAulay R and Quatieri T 1986 IEEE Transactions on Acoustics, Speech, and Signal Processing 34 744–754

[31] PR Sys Design 2018 perClass URL http://perclass.com/
[32] Jaitly N and Hinton G 2011 Learning a better representation of speech soundwaves using restricted boltzmann machines 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (IEEE) pp 5884–5887
[33] Böck S and Schell M 2012 Polyphonic piano note transcription with recurrent neural networks 2012 IEEE international conference on acoustics, speech and signal processing (ICASSP) (IEEE) pp 121–124
[34] Adavanne S, Parascandolo G, Pertilä P, Heittola T and Virtanen T 2017 arXiv preprint arXiv:1706.02293
[35] Gemmeke J F, Ellis D P, Freedman D, Jansen A, Lawrence W, Moore R C, Plakal M and Ritter M 2017 Audio set: An ontology and human-labeled dataset for audio events 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (IEEE) pp 776–780
[36] Drossos K, Mimilakis S I, Gharib S, Li Y and Virtanen T 2020 arXiv preprint arXiv:2002.00476
[37] Cho K, Van Merriënboer B, Bahdanau D and Bengio Y 2014 arXiv preprint arXiv:1409.1259
[38] Kingma D and Ba J 2014 International Conference on Learning Representations
[39] Van Asch V 2013 Belgium: CLiPS 49
[40] MathWorks 2017 MATLAB version 9.3.0.713579 (R2017b) The Mathworks, Inc. Natick, Massachusetts
[41] MathWorks 2017 Simulink version 9.0 (MATLAB R2017b) The Mathworks, Inc. Natick, Massachusetts