DOI 10.5162/SMSI2020/P3.11

Plastic Material Classification using Neural Network based Audio Signal Analysis

Sascha Grollmisch[1,2], David Johnson[1], Tobias Krüger[1], Judith Liebetrau[1]
[1] Fraunhofer IDMT, Ilmenau, Germany
[2] TU Ilmenau, Ilmenau, Germany
jsn@idmt.fraunhofer.de

Summary:

Analyzing the acoustic response of products being struck is a potential method to detect material deviations or faults for automated quality control. To evaluate this, we implement a material detection system by equipping an air hockey table with two microphones and plastic pucks 3D printed using different materials. Using this setup, a dataset of the acoustic response of impacts on plastic materials was developed and published. A convolutional neural network trained on this data, achieved high classification accuracy even under noisy conditions demonstrating the potential of this approach.

Keywords: acoustic quality control, industrial sound analysis, neural network, material classification

Background, Motivation and Objective

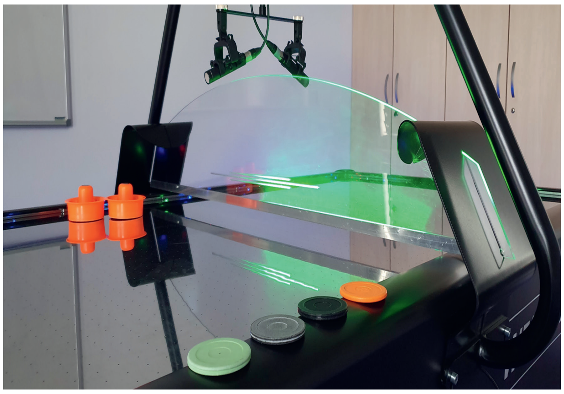

Fig. 1. Air hockey table with two microphones and pucks composed of different materials.

Analyzing acoustic signals in the human hearing range is often problematic due to the complex noise scenarios in factories. Industrial Sound Analysis (ISA) using low-cost microphones and neural networks was shown to be a potential solution for use-cases such as classifying metal surfaces or the operational state of electric engines [1]. Similarly, faults inside metal or plastic materials may only be identified through acoustic analysis because the faults are often not visible. Furthermore, applying structure-borne sensors to each product is often too costly. Analyzing the acoustic response of products such as metal screws was demonstrated to be a feasible solution for acoustic quality control [2]. Products are hit in a non-destructive way with the acoustic response being recorded. The recordings are then classified correctly under the noisy conditions of real-world production lines.

In this work, we evaluate if different plastic materials are distinguishable under noisy conditions by analyzing the acoustic response of the material on impact. Our proposed baseline system combines audio signal processing with a convolutional neural network (CNN) classifier. Additionally, we publish a novel dataset, *IDMT-ISA-PUCKS* (IIPD), alongside this paper for reproducibility and fostering further research [3].

Dataset

An air hockey table was used as a demonstration system for the simulation of difficult acoustic analysis conditions consistent with industrial manufacturing settings. Data collection was performed by equipping the air hockey table with two sE8 microphones, each recording one side of the table, as seen in Fig. 1, while a game is played. In addition to the standard pucks, several new pucks were 3D printed using one of three different materials. For each of the four materials, five three minute games were played, each with a different puck of the specific material. Further, each game was played with different sets of players. To mimic a real industrial setting, this scheme was repeated twice with industrial sounds at different loudness levels played in the background leading to a total dataset duration of 225 minutes.

Applied method

To detect the puck material during a game of air hockey, we propose a Convolutional Neural Network (CNN) classifier with five classes; four classes for each type of puck material and one class for the case when there is no puck in play.

We propose using a CNN over a fully connected neural network [1], because multiple hit events can occur in one analysis window. Since the CNN is invariant to time positions by applying pooling operations, generalization should improve. The proposed architecture should perform in near real-time and be robust to a variety of background noise.

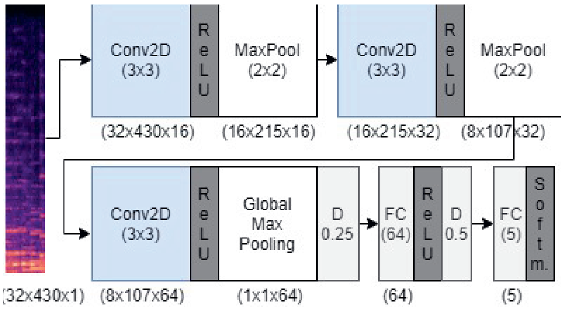

Fig. 2. CNN architecture with 3 convolutional layers, rectified linear unit (ReLU) activation, Dropout (D), a fully connected (FC) layer, and a final softmax classification layer.

To use a CNN with audio data, we prepare the data using the short-time Fourier transform (STFT) to generate a spectrogram as input to the classifier. For near real-time detection, we choose the maximum delay to be five seconds. Therefore, the analysis window is set to ten seconds with 50% overlap. For each ten second window, the raw data is downsampled to 22050 Hz and the STFT is applied with a window size of 512 samples and no overlap. We convert the spectrogram to mel-scale by applying a 32 band mel-scaled filter. This results in a 430 frame spectrogram with 32 frequency bins per ten second patch as input for the CNN. Additionally, per-channel energy normalization (PCEN) is applied to suppress background noise [5]. The CNN architecture, which was empirically designed, is shown in Fig. 2. Each CNN was trained for 1500 epochs using the Adam optimizer [5] with a learning rate of 0.001 and batch size of 256.

Experiments and Results

To validate the proposed method, we perform experiments evaluating puck material detection under different noise conditions. First, the model is trained using only recordings without background noise (nl1). This evaluates the general ability of the model to distinguish between materials. Additionally, to evaluate the robustness of the proposed method in real-world industrial scenarios with unknown background noises, the previously trained model is tested on recordings overlayed with background sounds at two different amplitudes (nl2 and nl3).

The results, evaluated using 5-fold cross validation, presented in Tab.1, show that materials are distinguishable with high accuracy when there is no background noise (nl1). Adding background sounds decreases the performance to 89%. However, when PCEN is employed the accuracy only drops to 95% in the worst case, showing the potential of this preprocessing method to reduce unwanted signal components. The PCEN parameters have been kept to default, but for future work their optimization might improve the results. Using majority voting of the predictions over a time window of one minute (instead of ten seconds) improves the accuracy to 99.7%, even with the highest background noise level. A possible reason for the performance gap on short windows is a lack of hit events in some of the 10-second time frames. This could be avoided with systems where the product is directly hit and analyzed, eliminating frames with no events.

Tab. 1: Results showing accuracy for different noise levels (nl) with and without PCEN.

	nl1	nl2	nl3	nl3 1min
w/o PCEN	99.5	90.0	89.0	91.5
PCEN	96.1	96.6	94.9	99.7

Conclusion

A dataset and classification system were created to detect the material of plastic pucks. By applying preprocessing with PCEN high accuracies can be achieved, confirming our hypothesis that plastic materials can be distinguished by their acoustic response under noisy conditions. However, material type may have a bigger influence on sound than small faults inside the products. To test the capabilities of our proposed method, faults could be printed inside pucks or other products for future research aimed at improving automated quality control systems.

References

[1] S. Grollmisch, et al., Sounding Industry: Challenges and Datasets for Industrial Sound Analysis, *EUSIPCO* (2019), A Coruna, Spain

[2] J. Liebetrau, et al., Luftschallbasierte Rissdetektion von Metallteilen, *DAGA* (2018), Munich, Germany

[3] IDMT-ISA-PUCKS dataset, Website: https://www.idmt.fraunhofer.de/datasets

[4] Y. Wang, et al., Trainable frontend for robust and far-field keyword spotting, *ICASSP* (2017), New Orleans, USA

[5] D. P. Kingma, et al.: A Method for Stochastic Optimization, *ICLR* (2015), San Diego, USA

This work has been partially supported by the German Research Foundation (BR 1333/20-1, CA 2096/1-1).