Personal and Ubiquitous Computing (2021) 25:691–704
https://doi.org/10.1007/s00779-020-01394-3

ORIGINAL ARTICLE

Auditory augmented process monitoring for cyber physical production systems

Michael Iber[1] · Patrik Lechner[1] · Christian Jandl[1] · Manuel Mader[1] · Michael Reichmann[1]

Received: 7 December 2019 / Accepted: 9 March 2020 / Published online: 19 March 2020

Abstract

We describe two proof-of-concept approaches on the sonification of estimated operation states and conditions focusing on two scenarios: a laboratory setup of a manipulated 3D printer and an industrial setup focusing on the operations of a punching machine. The results of these studies form the basis for the development of an "intelligent" noise protection headphone as part of Cyber Physical Production Systems which provides auditorily augmented information to machine operators and enables radio communication between them. Further application areas are implementations in control rooms (equipped with multi-channel loudspeaker systems) and utilization for training purposes. As a first proof-of-concept, the data stream of error probability estimations regarding partly manipulated 3D printing processes were mapped to three sonification models, providing evidence about momentary operation states. The neural network applied indicates a high accuracy (> 93%) of the error estimation distinguishing between normal and manipulated operation states. None of the manipulated states could be identified by listening. An auditory augmentation, or sonification of these error estimations, provides a considerable benefit to process monitoring. For a second proof-of-concept, setup operations of a punching machine were recorded. Since all operations were apparently flawlessly executed, and there were no errors to be reported, we focused on the identification of operation phases. Each phase of a punching process could be algorithmically distinguished at an estimated probability rate of > 94%. In the auditory display, these phases were represented by different instrumentations of a musical piece in order to allow users to differentiate between operations auditorily.

Keywords Auditory augmentation · Process monitoring · Auditory display · Cyber physical production systems · Error prediction estimation

1 Introduction

A side effect of the transition from traditional production processes to smart manufacturing and Industry 4.0 is a steady increase in complexity not only regarding the variety and diversity of products to be manufactured but also in terms of operating and maintaining production plants in general. The decreasing number of employees (e.g., caused by a higher degree of automated processes) goes along with an increased complexity of operations which need to be controlled as well as an escalating amount of data generated by these processes. As a further consequence, the degree of professional knowledge and expertise that enables operators to withdraw information from the collected data grows rapidly. Cyber Physical Production Systems (CPPS) support the mediation of *implicit* and *explicit knowledge* and can be adapted to the employees' individual level of expertise [1]. In the context of manufacturing, the term *explicit knowledge* comprises any kind of information that can be stored and made accessible to

This publication is an extended version of our contribution to the proceedings of the 2019 Audio Mostly Conference *Auditory Augmented Reality for Cyber Physical Production Systems* (https://doi.org/10.1145/3356590.3356600).

✉ Michael Iber
michael.iber@fhstp.ac.at

Patrik Lechner
patrik.lechner@fhstp.ac.at

Christian Jandl
christian.jandl@fhstp.ac.at

Manuel Mader
manuel.mader@fhstp.ac.at

Michael Reichmann
dm171550@fhstp.ac.at

[1] Institute of Creative Media Technologies, St. Pölten University of Applied Sciences, St. Pölten, Austria

operators, such as instruction sheets describing the handling of certain tools to be used for specific processes. *Intrinsic knowledge*, also known as *tacit* or *working knowledge*, on the other hand, refers to information based on personal experience that cannot be articulated easily, such as riding a bicycle or drilling a hole in a wall. Future human-centered intelligent factories aim to utilize information and knowledge derived from the production for tomorrow's production staff to optimally reinforce their skills in terms of creativity and innovation generation. As a result, operators will also have a higher job satisfaction resulting in increased productivity. Stocker et al. suggest application fields for information and communication technology (ICT) solutions based on four potential implementations [2]:

1. the "personalized augmented operator," which means the support of operators through augmented reality content
2. "worker-centric knowledge sharing," which establishes a culture in which knowledge sharing is the dominant strategy to provide actionable and decision-relevant information at the right time
3. "self-learning manufacturing workplaces," which corresponds to the approach of self-learning workplaces as a fixed component of Smart Factories, where operators are supported by intelligent data linking and analysis regarding Big Data, and finally
4. "in situ mobile learning in production," which describes mobile, personalized, and situation-adaptive learning systems for lifelong learning in enterprises and the generation-spreading transfer of know-how

To simplify the development of human-centered tools for manufacturing, [2] defined a framework containing 16 building blocks to ensure a successful implementation of the abovementioned application fields. These building blocks include categories such as hardware devices (for instance head mounted displays or wearables), communication infrastructure, data analysis tools, and worker environments including sensors, knowledge management systems, or enterprise resource planning tools. CPPS introduced by [3–5] are based on the proposed framework, focusing on visualization tools for information display. Thereby, beneficial aspects of advanced devices that emphasize other senses, such as (hands-free) auditory or haptic displays, to support operators in production plants have been rather neglected. Up to now, auditory alarms have usually still been restricted to intermittent, at times even uninformative, warning sounds, leaving out the potential of monitoring approaches based on continuous sonification, which have shown to improve situation awareness [6–8].

While such advanced implementations of auditory displays for process monitoring are still restricted, major advances in regard to the incorporation of acoustic information for automated condition monitoring have recently been achieved [9–13]. Research by Fraunhofer Institute for Digital Media Technology in Ilmenau resulted in a service[1] for automated quality assurance combing recording of airborne sounds with machine learning approaches [14]. Due to its comparatively simple implementation in existing infrastructure and the good recognizability of altering sound attributes by means of machine learning, more and more companies supplying automation technology offer machine learning-based methodology as a product for industrial applications, especially in energy, infrastructure, and manufacturing domains. These approaches mostly focus on automated, algorithmically classified, and evaluated condition monitoring processes. Their results are usually displayed in the visual domain exclusively without any involvement of auditory monitoring.

In highly interconnected manufacturing environments, however, advanced auditory displays (AD) would offer plenty of advantages that have already been utilized in other fields of expertise. For instance, AD allow operators to freely interact during primary activities, such as walking, running, or driving, where the visual attention of humans is focused on navigation or orientation [15]. For this reason, interactive auditory displays have been developed for gait representations without barriers that may have impact on the physical posture [16, 17]. In comparison with other user interfaces such as tablets or head mounted displays, auditory interfaces also require less physical activity. These facts make AD interesting for smart manufacturing applications, especially for monitoring and controlling operations in production lines and shop floors.

Based on these considerations about (i) the potential of auditorily enhanced CPPS, (ii) acoustic condition monitoring based on machine learning, and (iii) the importance of auditory feedback for work experience and the buildup of working knowledge [18, 19], we designed a research project that combines these three aspects. It pursues the development of a method for the sonification of processed machine emission data that equips operators with unobtrusive acoustic information about the actual state or condition of one or more machines and ongoing operations in quasi real time. In the context of process monitoring, [20] differentiates between direct, peripheral, and serendipitous-peripheral monitoring modes. Peripheral monitoring relies on "information that is not central to a person's current task, but provides the person the opportunity to learn more, to do a better job, or to keep track of less important tasks" [21]. Serendipitous-peripheral monitoring is meant as "information that is useful but not required" [20]. The implementation of AD within our research project will focus on the former, i.e., peripheral monitoring, in order to confirm the smoothness of the processes and to make operators betimes aware of non-optimum behavior and future threats by continuous sonification (cf. Sects. 2.1, 2.2, and 3.3). In the long term, our research aims to develop an

[1] https://www.idmt.fraunhofer.de/de/business_units/ima.html

intelligent noise protection headphone with an integrated assistant system that provides supportive information about machine states, operations, setup, and maintenance routines to be used at production plants. We also plan a multi-channel loudspeaker implementation for control rooms.

In this publication, we will present two proof-of-concept sonification approaches that focus on process identification and error probability based on operation state and condition classification data. In order to gain basic knowledge on how to map classification data resulting from machine learning to auditory feedback information, we equipped a 3D printer with several microphones and recorded printing processes in normal and manipulated conditions. We then analyzed the resulting audio files by machine learning algorithms and developed several sonification models to display the error probabilities of the processes. In our second study, we focused on the algorithmic identification and sonification of operation phases occurring during the processes of a punching machine.

The next section gives insight into the state of the art of the several fields this paper is related to. This includes aspects of process monitoring and ecological interface design, sonification, and auditory augmentation, as well as approaches based on machine learning and feature classification for acoustic monitoring. We will then present our proof-of-concept approach on mapping error probability rates of the mentioned 3D printing processes to continuous sonifications followed by the study on operation phase identification and sonification. After a discussion of the results, we will conclude with an outlook on future steps of our research.

2 Related work

Despite an increasing demand for automated process surveillance and condition monitoring in manufacturing environments, the human factor in terms of manual handlings of processes, judgments of situations, and the reliability of decisions to be made may by no means be neglected in order to ensure a successful and sustainable production. We will focus on three aspects of this rather broad topic which all fall within the scope of acoustic information design and the relationship between operators and their work. Firstly, we will present aspects of process monitoring and workplace design. Particularly concerning the implementation of warning sounds, a vast amount of research has been conducted in recent decades. Secondly, we will discuss approaches referring to the term auditory augmentation or auditorily augmented reality, emphasizing the extension of auditory spaces for additional information. Finally, we will conclude the section with a review on industrial setups in which machine learning approaches on acoustic emissions have been used to analyze and categorize production processes.

2.1 Auditory display for process monitoring

The design and implementations of warning sounds in critical situations has been discussed for many years. Patterson and Mayfield [22] and Edworthy et al. [23] elaborated criteria concerning the attributes of warning sounds in order to distinguish them from the production environment. Various authors (see, e.g., [24–26]) mention "alarm fatigue," "alarm flooding," and sequential "alarm showers" as aspects needing to be considered for the design of warning sounds. The avoidance of too many sounds, i.e., also of too much information to be handled properly [27], appears to be as important as the prevention of inattentional deafness, i.e., the failure of noticing warning sounds [28]. Most implementations of alarming sounds are based on intermittent, event-based auditory displays presenting one or a sequence of sounds, either in anticipation of or on the actual occurrence of critical situations [29]. However, [8] indicate the advantages of continuous sonification for process monitoring, since instead of displaying warning sounds only related to specific situations, the auditory display will permanently represent the state of the monitored system. This enables operators to anticipate upcoming problems. A major challenge for the sonification design is to create a sonic environment that is meaningful and unobtrusive at the same time in order to be well conceived by the operators. In [7], we present a comprehensive overview on auditory displays for process monitoring including their various fields of application (e.g., air traffic control, plant surveillance) as well as the design criteria to be considered.

In their approach towards an integration of auditory displays into the proceeding scheme of Ecological Interface Design, [24, 30] develop a multimodal monitoring environment for the demands of anesthesia and intensive care units using intermittent and continuous sounds. The vital conditions of these environments require high accuracy in terms of the unambiguousness of the acoustic information to ensure that operators make the right decisions. At the same time, auditory displays should be reasonably pleasant to listen to in order to prevent distraction of the operators by inducing stress and to avoid an interference with the healing process of the patients. Baldwin et al. [31] provide an overview of perceptual advantages and disadvantages of multimodal displays regarding the complexity of information.

In a dual task experiment on multimodal displays, Hildebrandt et al. demonstrated that an AD based on continuous sonification outperformed event-based, intermittent sonifications in terms of accuracy and timing [8, 32]. Haas and Erp [33] supported these findings in their overview on multimodal warnings. Using continuous sonification models based on music, [34] showed that background music does not "distract users from their primary task, and […] can effectively convey information." Assumedly, it can be listened to over long periods of time, especially when the kind of music can be selected by the user.

2.2 Auditory augmentation

Mynatt et al. implemented an "audio augmented reality" in the context of office environments [35]. They developed a system using active infrared badges for tracking the position of persons and wireless headphones to deliver information through auditory cues built on the peripheral acoustic office environment. The system was behavior dependent, everyday routines such as walking through the office for instance could trigger additional auditory information, notifying a specific person about her meetings or the status of incoming emails. The authors were aware of avoiding the "alarm paradigm" (intermittent auditory displays) and integrated the auditory cues into a continuous "low-level soundtrack" (continuous sonification) as a combination of music, sound effects, and voice.

In their approach to auditory augmentation, [36] overlaid natural acoustic emissions, for instance noises that arise from typing on a computer keyboard with sound effects controlled by parameter values of an independent data domain, in the given example weather data. Serving as a basis for the generated sounds, keystrokes were recorded by vibration pickup microphones and played back in quasi real-time after data dependent sound processing had been applied. Users classified this unobtrusive additional weather report as useful information. Although the processed sounds were clearly distinguishable, nobody "mentioned the system to be bothersome." In their resume of a workshop with 19 participants from the sonification community, interaction experts, composers, sound engineers, one musicologist, and one sociologist [37] extended this approach and designed three prototype scenarios of auditory augmentation, which they defined as "the augmentation of a physical object and/or its sound *by sound* which conveys additional information." Together with Grosshauser [38], Hermann, one of the authors of the mentioned publication, equipped a drilling machine with sensors and sonified deviations from the optimum angle within an auditory interaction loop as a further example of auditory augmentation.

2.3 Machine learning approaches for process classification in production environments

To the authors' best knowledge, there is no published previous research in terms of an implementation of continuous sonification based on classification or continuous data retrieved by machine learning. There are several examples of prior work facilitating a combination of acoustic condition monitoring and machine learning. Pasha et al. present an overview of multiple supervised machine learning techniques (such as SVM, J48, and Deep Learning) used in the scope of acoustic condition monitoring [13]. Acoustic condition monitoring is, for instance, applied to the detection of air leaks in a sintering plant. The algorithm that performs best here is a Recurrent Neural Network (RNN). By directly feeding a collection of output frames of a Short Time Fourier Transform (STFT) into the network, a classification accuracy of more than 80% was achieved. Zafar et al. describe the use of RMS measurements as one of four input features to an artificial neural network in order to classify tool conditions in a wood milling process [12]. They demonstrated that the addition of airborne acoustic emission measurements increases accuracy in classification.

Both the approach of using the complete spectral magnitude data as input for a machine learning model, as well as using feature extraction (e.g., a combination of spectral moments, mel frequency cepstrum coefficients (MFCCs) and other audio features) beforehand have been explored in prior work. Liang and Wang describe the application of condition monitoring to a Desktop CNC 3D engraver machine [39]. They extracted vibrational features using a piezoelectric sensor and calculating spectral features in the range of human hearing by using a spectrum analyzer with seven frequency banks. The individual energies of these bands were divided by each other, resulting in 21 additional dimensionless indices. Combined, the energies of the frequency bands and the 21 dimensionless indices resulted in a total of 28 features as input for machine learning. All applied algorithms (random forest (RF), K-nearest neighbor (KNN), and support vector machine (SVM)) proved suitable for classification.

Grebenik et al. presented the detection of a roller element bearing failure via smartphone microphone recordings of airborne acoustic emission [40]. Three multi-dimensional features were derived from the audio recordings. The audio spectrum was divided into several bins to be analyzed separately, and three features were extracted from each bin: the number of peaks above a specified threshold, the number of peaks, and the product of the amplitudes of all the peaks in the bin. Deciding against artificial neural networks (ANN) in favor of a multi-SVM approach, a 95% accuracy could be achieved. This decision for SVMs was largely due to the quicker training times in comparison to ANNs. Yang et al. compared the performance of several classifiers, among them ANNs and SVMs [41]. They also argue in favor of SVMs because of more efficient risk minimization that "leads to a better generalization performance." They conclude that SVMs and LVQ (learning vector quantization) provide the highest accuracy "for classifying healthy and faulty conditions of small reciprocating compressors."

3 Design of a sonification model based on real-time process classification

3.1 Approach

In order to generate representative data of printing operations, we equipped a 3D printer[2] with two types of microphones. Four

[2] BQ Witbox 2

miniature condenser vibration pickups[3] were mounted to the four stepper motors (responsible for the *X*-, Y-, and Z-axis movements) of the extruder as well as the conveyer of the printing material (filament). Another two pickups of the same type were fixed to the connecting part between the guidance rods of the X- and Y-axes and the filament spool holder. Additionally, two hyper cardioid instrument microphones[4] with magnetic mounts were placed on the printer frame (Fig. 1).

For printing, we adjusted a given 3D model of a hollow cube model without the top and bottom[5] to a volume of 2 cm^3. We then printed the cube three times under normal conditions. Thereafter, we induced three faults to the printer which are not unusual to occur during printing operations and may affect print quality or lead to complete failure. For the first of the three printing errors, the grease on the guidance rods of the *x*-axis was removed. To create the second error condition, the screws of the extruder fan were loosened. For the third error, we reduced the tension of the hobbed bolt of the extruder, which causes issues with the conveying of the filament.

All printing operations were recorded on eight audio channels. For reasons of reproducibility, a synchronized video of the operations was captured additionally. The recordings were manually edited, labeled, and prepared for data analysis.

An informal evaluation of the recorded material conducted among the authors revealed that the error states of the printer cannot be distinguished from the normal operation state aurally. The authors' expertise and experience in the fields of music and sound engineering, and the subsequent inability to aurally distinguish the different printing states from each other, led to the relinquishment of further aural experiments testing uninvolved subjects.

3.2 Data analysis[6]

The analysis aimed at finding an appropriate method to retrieve the recorded printing states (error condition vs. normal operation) from the audio data. This method had to fulfill three main requirements:

1. Verify that the information (acoustic cues) is contained in the data.
2. Provide insight into where or how the information is contained in the recorded data.
3. Provide preferably low-dimensional data to keep the complexity of the data sonification as low as possible.

Using the raw spectral data of all microphones would have resulted in a high-dimensional (8 × frame size) input vector for the machine learning algorithm. In order to reduce data complexity, we therefore performed feature extraction and feature selection first. We built a suitable training set by framing the audio data of all recordings using a frame size of 65,536 (= 2^{16}) samples and a hop size of 4096 samples. This rather large frame size facilitates a high frequency resolution as a basis for further processing. From the obtained spectral data, the following features were chosen for their general acceptance in audio machine learning applications: five MFCCs, root mean square (cf. [12]), spectral bandwidth, spectral centroid, and spectral roll-off. Feature calculation was based on the libROSA python package [42]. Other than [13], we did not run machine learning algorithms on the complete data generated via spectral analysis, but rather performed feature extraction and selection[7] to achieve a quicker convergence of the machine learning algorithm. That way, we also obtained means of getting insights into the data by automatic feature selection.

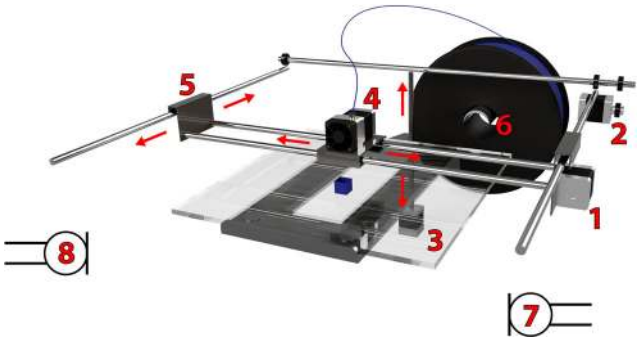

Fig. 1 Schematic representation of the microphone placement on the 3D-printer: 1–4: vibration pickups on the stepper motors (*x*-, *y*-, *z*-axes, extruder); 5: vibration pickup on the connection of the rods of the *x*-axis and the *y*-axis; 6: vibration pickup on the filament spool holder; 7 and 8: small diaphragm condenser microphones (DPAs) with hyper cardioid characteristics on the frame

As result of the analysis, we obtained a table containing 18,760 labeled observations (audio frames) with 64 audio features (8 microphones × 8 audio features). The gathered dataset was subsampled to obtain a balanced distribution of 50% error states and 50% states of normal operation. For automatic feature selection, the chi-squared test (X^2) [43] was chosen for its generality, simplicity, and effectiveness [44]. The 15 most relevant features of all recordings were selected as input to the network model (cf. Fig. 2).

The application of an SVM did not deliver satisfying results. Therefore, we utilized a neural network-based classifier (Fig. 3). The model was built using the python libraries Keras [45] and TensorFlow [46]. While [13] applied a recurrent neural network (RNN) to their audio data in a similar approach, for a start, we opted for a standard forward one. Comparing several configurations

[3] AKG C411
[4] DPA d:vote 4099
[5] https://www.thingiverse.com/thing:187707
[6] Documented code available at https://github.com/fhstp/AARIP
[7] That is, not all calculated audio features of all microphones were used but a subset that could be shown to correlate most with the printer state

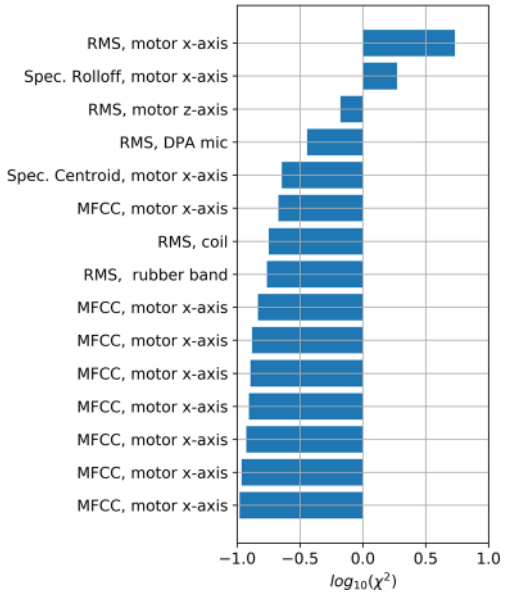

Fig. 2 Fifteen most relevant features according to X^2 of all recordings. "DPA mic" refers to a single DPA microphone (cf. no. 7 in Fig. 1)

(concerning the number of layers, neurons, and layer types), this network model showed to be sufficiently accurate for our purposes. It requires a relatively low number of features as an input, which in turn reduces the requirements for a real-time classification or the development of sonification models. However, a model that also makes use of past states (such as RNNs) is very likely to further improve the obtained accuracy and we will consider this for future developments.

The obtained data was split into a training set, a validation set and an independent test set. Using a training/validation split of 0.3 during the training process, an accuracy of > 93% on the independent test set could be achieved. This indicates that the collected data was meaningful (cf. Table 1 and Fig. 4 for a receiver operating characteristic (ROC) plot using only independent test data).

We therefore conclude that the chosen model fulfills the requirements in terms of prediction reliability. Through feature selection, we were able to identify information-rich features and the model allows the classification of system states and conditions. Thus, the hypothesis that information is contained in the data is confirmed. Furthermore, the network model generates low dimensional data streams which makes it particularly suitable for the subsequent sonification.

The results of our data analysis offered three starting points for sonification approaches:

1. Data of the identified most relevant features are directly mapped to a sonification model.
2. Data of the identified most relevant features are used as metadata to manipulate incoming audio signals of the monitored machines. Thus, relevant sonic information within these signals can be emphasized and conditioned.
3. The information on the confidence (error probability) of the model is used directly instead of thresholding this value to retrieve a classification. This provides a continuous data stream which is one-dimensional, meaningful and already normalized (Fig. 5).

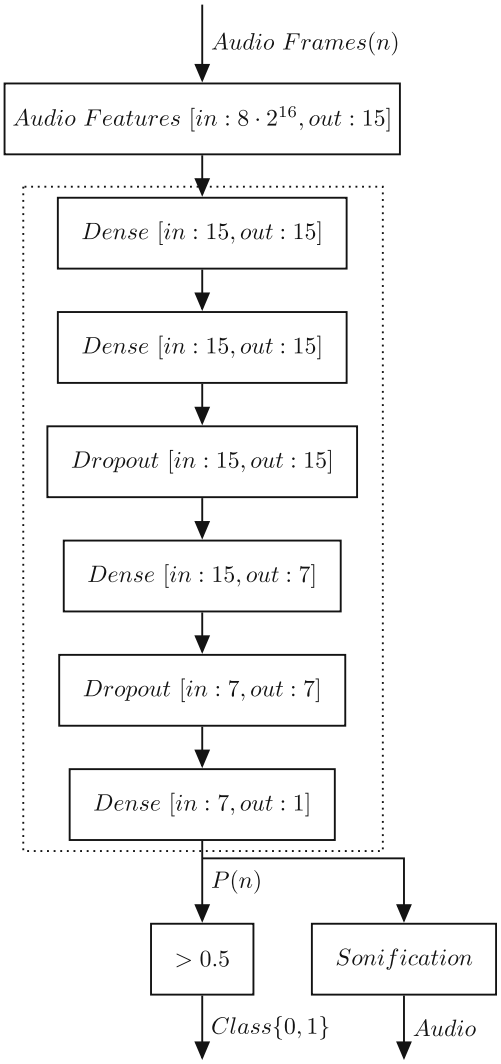

Fig. 3 Scheme of the analysis and classification processes. The *n*th audio frame of all eight microphones is fed into the feature extraction section. The 15 most relevant features are then transferred to the artificial neural network (contained in the dotted rectangle). The network calculates an error prediction $P(n)$. A threshold is used to generate the classification data for training and evaluation. Additionally, $P(n)$ is fed to the sonification section

Table 1 True/false positive/negative rates of the independent test set

True negatives	False positives	False negatives	True positives
482	34	35	505

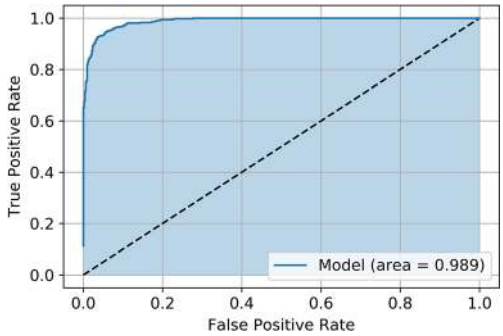

Fig. 4 ROC of independent test dataset

By reason of simplicity and efficiency, we chose the third of these starting points as data basis for our sonification model.

3.3 Design and application of auditory display

By the application of machine learning algorithms, a highly complex input situation (eight channels of audio data) could be simplified to a one-dimensional data stream giving evidence of the error probability of the monitored operations. Thus, the challenge to deliver easily accessible and distinct information that is frequently put on an auditory display could be enormously reduced. The error probability indices of the previous condition classification comprised a value range from 0.0 to 1.0 for each analysis frame (at about 12 frames per second). These incoming values were smoothed by a moving average window of 10 frames length. To distinguish between normal operation states and error conditions, we applied a threshold at 0.7 on the weighted data stream. Error probability values below that threshold were unambiguously considered as normal state operations, values above gradually indicated increased probabilities of errors.

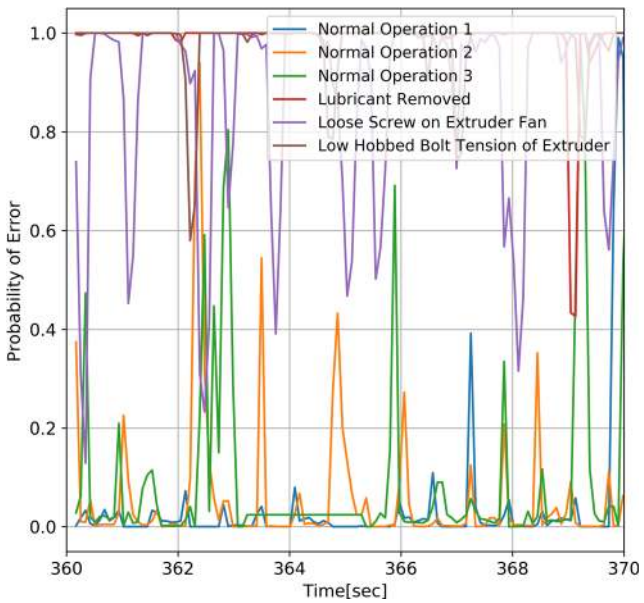

Fig. 5 Prediction of error probability $P(n)$ over time within a 10-s segment of recordings for the different operation states

As a proof-of-concept, we designed three sonification models utilizing rather diverse approaches based on the following metaphors: (i) heartbeat, (ii) soundscape, and (iii) music listening. Thereby, we considered five fundamental requirements:

1. In terms of an auditory augmented reality approach, the classification processes and the sonification processes are altogether realized in quasi real time.
2. Normal states are unobtrusively represented by continuous sonification [47] to affirm that everything is working alright.
3. Error conditions are clearly distinguishable without being considered as alarms.
4. Silence indicates a dropout of the complete system.
5. None of the represented states must acoustically hinder verbal communication (e.g., via radio).

The "heartbeat model" was chosen for its simplicity and its inherence to human activity. The characteristic double beat was generated by envelope shaped sinusoids. By default, the basic meter was set to 60 bpm and represented normal operation states reassuring a well-functioning system. As soon as the value stream of error probability indices exceeded the threshold, the meter started to fluctuate and speed up. Also, the volume of the heartbeats increased.

For their dual task experiment, Hildebrandt et al. [8] designed a soundscape based on a "forest" metaphor that included sounds such as a woodpecker pecking a tree or breaking twigs. We picked up this concept of utilizing nature sounds for the development of the "soundscape model." Based on procedural synthesis models provided by [48], we implemented a natural environment that included bird tweets and flaps, crickets, wind, thunder, and rain. All parameters, such as, e.g., wind speed, triggering of chirps and tweets, and positioning in stereo panorama, were driven by random values. Only the individual contribution of the elements (= mixing) to the scene was controlled by the error condition parameter values. Therefore, good weather conditions including sounds of birds and crickets represented normal operation states, while upcoming storm and rain sounds indicated an increase of error probability over the threshold at 0.7.

As mentioned in Sect. 2.1, Barra et al. [34] developed and evaluated a continuous sonification model that included background music which was enriched by additional musical information. Based on this rather complex concept, we designed a much simpler "music listening model" that respects the habit that many operators have, according to our observations, of listening to music (via headphones or loudspeakers) during

work.[8] Using our model, operators continue to listen to the music of their preference. However, in case of an increased error probability, a gradually narrowing bandpass filter is applied to the music playback (patina effect). Accordingly, the speed of the music starts to fluctuate in order to make operators aware of increasing error probability. The implementation of the speed fluctuation is based on the "supervp~"-external of the MuBu Library provided by IRCAM[9] [49] which allows tempo manipulations independent of frequency shifts in decent quality.

3.4 Results of 1st pilot study on error estimation sonification

Our results in detail are as follows:

1. Combining feature extraction and a custom artificial neural network (ANN), the applied model indicated a high accuracy (> 93%) concerning error probability identification distinguishing between operation states. None of these states could be identified by listening. An auditory augmentation, or rather the sonification of this classifier, provides a considerable benefit to process monitoring.
2. The data stream of error probability values was mapped into a sonification model, providing evidence about momentary operation states. Three models relying on different acoustic metaphors (heartbeat, soundscape, music listening) were implemented as a proof-of-concept. These models were designed to be unobtrusively perceived during normal conditions, clearly indicating error states without shifting into warning sound characteristics.
3. The system works in quasi real time; the application of the analysis buffer causes a delay of about 85 ms; the input and output buffers of audio interface add another 10 ms.

Due to the simplicity of the sonification models and the one-dimensional, almost Boolean information stream, error conditions are easily distinguishable from normal states in all three models. We therefore decided to forego a formal perceptional user study for now and restrict our approach to a proof-of-concept. Also, the general benefit of continuous sonifications for early identification of upcoming issues has already been evaluated by in vitro studies (see, e.g., [32, 34]). As the latter pointed out, long-term observations under real-world conditions are necessary in order to evaluate the impact, benefit, and, most importantly, willingness of operators to accept exposure to the provided acoustic information on a day-to-day 8-h basis. While we expect a good chance for an implementation of the music listening model in manufacturing environments, we doubt the potential of the two other models since they appear quite uniform and fatiguing overall. For our 2nd proof-of-concept study, we therefore focused on the musical aspect.

3.5 Design and application of 2nd proof-of-concept study: process classification

As a next step of our research, we designed a second proof-of-concept study in situ at the shop floor of a metal working company. The fluctuating acoustic environment of a real-world production scenario implicates additional challenges for airborne sound analysis and process categorization. In addition to noises caused by nearby machines, passing by forklift trucks or human activities, area-wide music playback all over the shop floor was also a source of acoustic emission that needed to be taken into consideration.

Similar to our proceeding in the first study, we equipped a semi-automatic CNC punching machine[10] with 10 small diaphragm condensers and contact microphones[11] at strategic positions which are, for instance, situated near the punching head, the work plate, the valve, the clutch, the compressor, and the transformer box. The aims of the study were as follows:

1. to test/adapt our previously established feature extraction and machine learning routines against/to environmental influences
2. to classify different operation phases during processes[12] with an accuracy in similar height to the one achieved in the first proof-of-concept study
3. to develop a sonification model that clearly displays and distinguishes operation phases and integrates them into the work environment

3.6 Process phases during operations

The processing of a single workpiece, i.e., a metal sheet, at the punching machine can be subdivided into five operation phases:

1. operator inserting the workpiece into the machine (manual operation)
2. punching processes (automatic operation)
3. re-arranging the workpiece (automatic operation)
4. punching processes (automatic operation)

[8] In later real-world implementations, operators will be encouraged to compile their individual playlists and listen to the music of their preference.
[9] www.ircam.fr
[10] Boschert Punching Machine Compact (https://boschert.de/en/products/machines/punching-machines/compact.html)
[11] Six AKG 411 contact microphones, 2 DPA 4099 cardioid clip microphones, 2 Sennheiser MKE600 shotgun microphones
[12] Originally, we also intended to analyze and classify the probability of process errors. However, since all (500+) operations during our two-day observation were processed flawlessly, this aspect of our research had to be postponed to a follow-up investigation.

Table 2 Representative timestamps of operation phases after manual labeling

Product type	Selected workpiece	Operation phases (starting times in seconds): workpiece …				
		Inserted	Punched	Re-arranged	Punched	Released
A	1	00:00	00:15	00:25	00:36	00:40
A	2	00:00	00:13	00:23	00:34	00:37
A	3	00:00	00:14	00:24	00:35	00:38
A	4	00:00	00:15	00:25	00:36	00:38
A	5	00:00	00:15	00:25	00:36	00:39
A	6	00:00	00:15	00:25	00:37	00:39
A	7	00:00	00:14	00:24	00:35	00:38
A	8	00:00	00:13	00:23	00:35	00:38
A	9	00:00	00:13	00:23	00:34	00:37
A	10	00:00	00:14	00:24	00:35	00:39

5. operator withdrawing workpiece (manual operation)

Our self-defined task for the process classification was to develop a method based on our first proof-of-concept study that automatically distinguishes between these phases with a comparable accuracy (i.e., > 93%).

We focused on the recording of the processing of one specific product type ("A"). The observed custom order comprised 500 workpieces, a sample size that we expected to deliver enough data for our analysis. The processes for this product type consisted of 10-mm-diameter stamps punching holes into a 0.55-mm electronically galvanized steel sheet. In order to be able to reproduce the operations recorded by the set of the 10 microphones described above, we filmed the scenario with a video camera that was time-synchronized to the audio recordings. Manually labeled operation phases show a maximum difference of 3 s within each of the 5 operation phases (cf. Table 2) indicating that even the processes involving manual activities ran on a stable basis.

Combining the manual operations, i.e., the "inserting" and "withdrawing" of a workpiece to an overall "handling" phase and considering the two "punching" phases as a single category, we obtain a characteristic temporal pattern of operation phases as displayed in Fig. 6.

Equivalent to our previous proceeding, we performed feature extraction on all audio recordings which were framed to a buffer of 2^{15} samples[13] using 12 MFCCs (from a mel spectrum with 128 mel bins), spectral centroid, spectral roll-off, and spectral bandwidth. Feature selection was performed using X^2 (Fig. 7).

The 30 most relevant features of a dataset of 1206 frames were selected and fed into seven network models[14] for training and testing using a train/test split of 0.5. For each input frame, the network estimates the probability $P(n)$ for each of the three classifiers representing the "handling" [0], "punching" [1], and "re-arranging" [2] phases of the operations. The classifier with the highest probability ranking determines the allocation of the analyzed frame. While most of the tested networks exhibit rather high confusion rates between phases [0, 2]—the confusion matrix of the support vector machine (Table 3) with an overall accuracy of about 80% provides a representative example—the random forest network (Table 4) performed best with an accuracy of more than 96%.

In order to a obtain a more flexible solution for challenges of future scenarios, we continued our research by developing a custom artificial neural network (Fig. 8) based on the one we had used in our first pilot study (Fig. 3) with superior modularity, expandability, and scalability. With an accuracy rate of about 94%, this model performed slightly worse than the random forest network (about 96%). According to the confusion

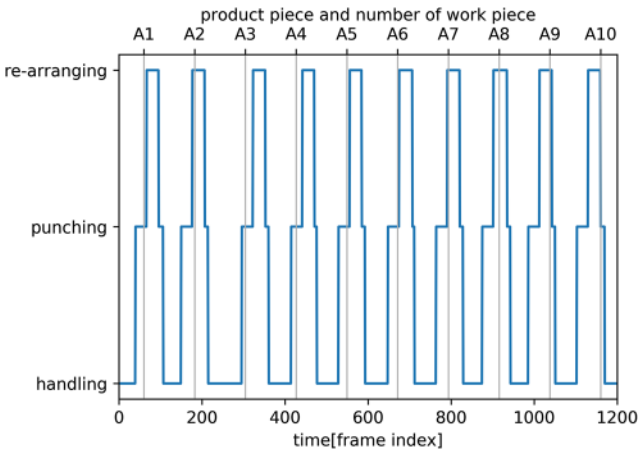

Fig. 6 Sequence of operation phases over time

[13] The use of a smaller buffer size than the one set in our first study was caused by memory restrictions.

[14] Logistic regression, decision tree classifier, K-nearest neighbor, support vector machine, random forest classifier, multi-layer perceptron classifier

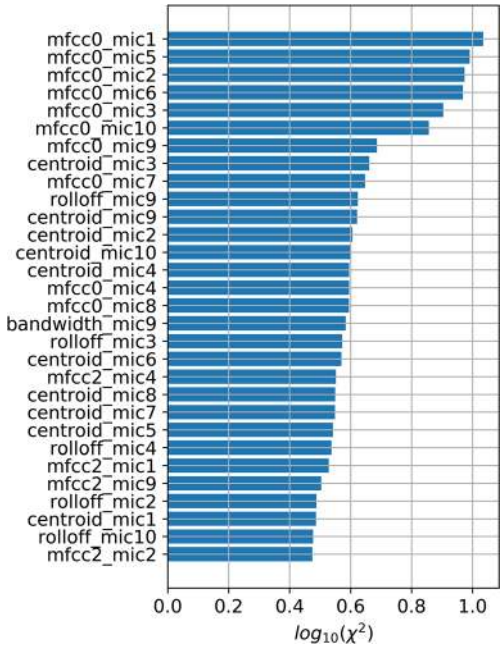

Fig. 7 The 30 most relevant features according to X^2

matrix (Table 5), however, the confusion between class 0 ("handling") and 2 ("re-arranging") is on a similar level than the one exhibited by the random forest model and also outperforms all the other tested networks models. Since also the ROC in Fig. 9 displays individual accuracies of 95% for class 2 and even better performances for classes 1 and 3, we conclude that we reached our stated target of achieving an accuracy comparable with the one we reached in our first proof-of-concept study.

The time-agnostic characteristics of the network model become evident in the noisy output of the original signal (Fig. 10). We smoothened these fluctuations by applying an infinite impulse response (IIR) filter $H(z)$ to the output of the network before allocating the analyzed frames to their most probable class via *argmax* (Fig. 8). The filter was constructed using the following difference equation, with s being a smoothing constant:

$$y(n) = \begin{cases} y(n-1) + \dfrac{x(n)-y(n-1)}{s}, for\ x(n) < y(n) \\ x(n), for\ x(n) \geq y(n) \end{cases}$$

resulting in the transfer function

Table 3 Confusion matrix for classifiers [0–2] obtained by the application of a support vector machine network

	[0]: handling	[1]: punching	[2]: re-arranging
[0]: handling	265	1	6
[1]: punching	18	153	14
[2]: re-arranging	77	4	65

Table 4 Confusion matrix for classifiers [0–2] obtained by the application of a random forest network

	[0]: handling	[1]: punching	[2]: re-arranging
[0]: handling	270	0	2
[1]: punching	1	184	0
[2]: re-arranging	12	7	127

$$H(z) = \frac{1}{s + z^{-1} - sz^{-1}}$$

for a falling signal, and

$$H(z) = 1$$

for a rising signal.

Figure 10 also shows that the accuracy of our model was essentially improved by this filtering of recent predictions. While recurrent neural networks would offer a logical next step to truly make the model aware of previous states, the presented model fulfills the given task in a satisfactory manner and can even be used to label more collected data in order to train a more general model.

3.7 Sonification model

The auditory display of error probability estimations as performed in our first proof-of-concept study suggests the implementation of sonification models that map an increasing probability of faulty operations to sonic parameters that indicate rather negative connotations. This can be realized by modeling bad weather conditions or by applying patina filters to high-end music recordings. Errors that, for instance, are caused by the deterioration of machines usually do not appear at once but develop gradually. The worsening of generated weather conditions by upcoming rain and thunderstorms or gradually applied filters according to the state of deterioration will provide useful information to experienced operators so that they are well informed about the state of machines and can decide at which point to take action.

The challenges for designing auditory displays that represent operation states are rather different, since these phases do not change gradually but immediately. The sonification should indicate the state clearly on a perceptually neutral basis without evaluating the quality of the processes. The provided information should assure operators that everything is working properly. Also, it must be kept in mind that the displayed sounds will be listened to over long periods of time. Therefore, a strategy is needed that respects the usual acoustic environment operators are accustomed to and does not essentially intrude into the auditory scene. The shop floor of the enterprise where we recorded the punching processes at was permanently flooded with music. Listening to music during work has been a

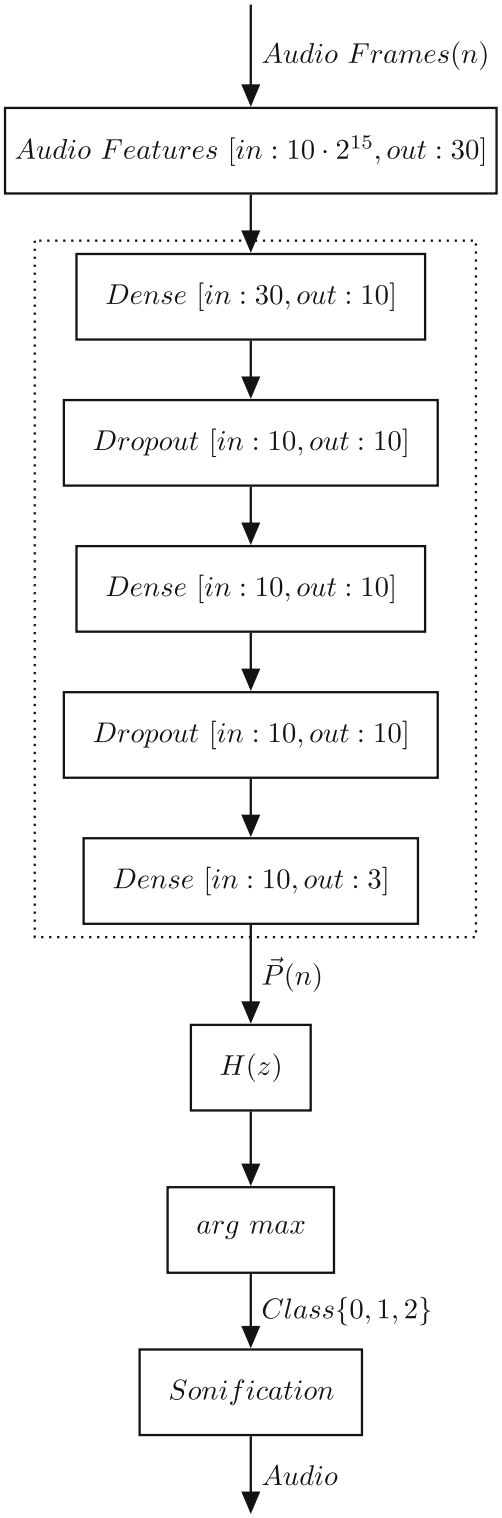

Fig. 8 Complete classification model including feature analysis and selection, custom artificial neural network (ANN), and smoothing filter

common experience for all operators who work there. Therefore, the development of a sonification model that considers listening to music can be expected to fulfill the stated criteria.

Table 5 Confusion matrix for classifiers [0–2] obtained by the application of our custom artificial neural network (ANN)

	[0]: handling	[1]: punching	[2]: re-arranging
[0]: handling	261	2	9
[1]: punching	3	182	0
[2]: re-arranging	15	5	126

All three operation phases ("handling," "punching," "re-arranging") should be displayed on a non-judgmental basis. One way to comply with this condition is the instrumentation of a musical piece. However, other than the application of audio effects, such as patina or tempo fluctuations, instrumentation as a sonification parameter cannot be applied to produced music recordings. For our second proof-of-concept study, we therefore arranged the jazz standard *Autumn Leaves* by Joseph Kosma manually according to the time sequence of phases given by the applied machine learning algorithm. While the plucked double bass and the laid-back drums (including brushes) build a continuous stable basis over the complete scene, the handling phase is represented by a muted trumpet for the melody and a piano for the accompaniment. During the "automatic" operation phases (i.e., "punching" and "re-arranging") of the punching machine, these two instruments were substituted by a lead and a rhythm guitar. In order to distinguish between "punching" and "re-arranging" phases, the latter were instrumented with an additional synthetic male choir (Table 6).

3.8 Results of the 2nd pilot study on operation phase sonification

Our results in detail are as follows:

1. The adjusted model combining feature extraction and a custom artificial neural network appears to be robust against the environmental influences that occurred during the recording phases.
2. The model applied to estimate the probability of three different operation phases indicates an accuracy even

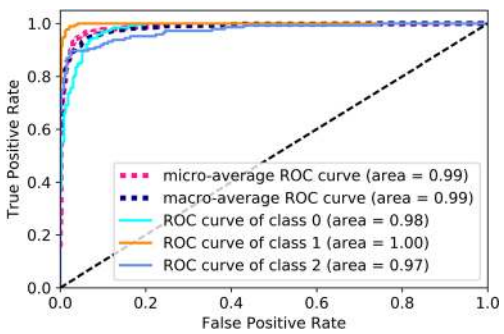

Fig. 9 Receiving operator characteristic

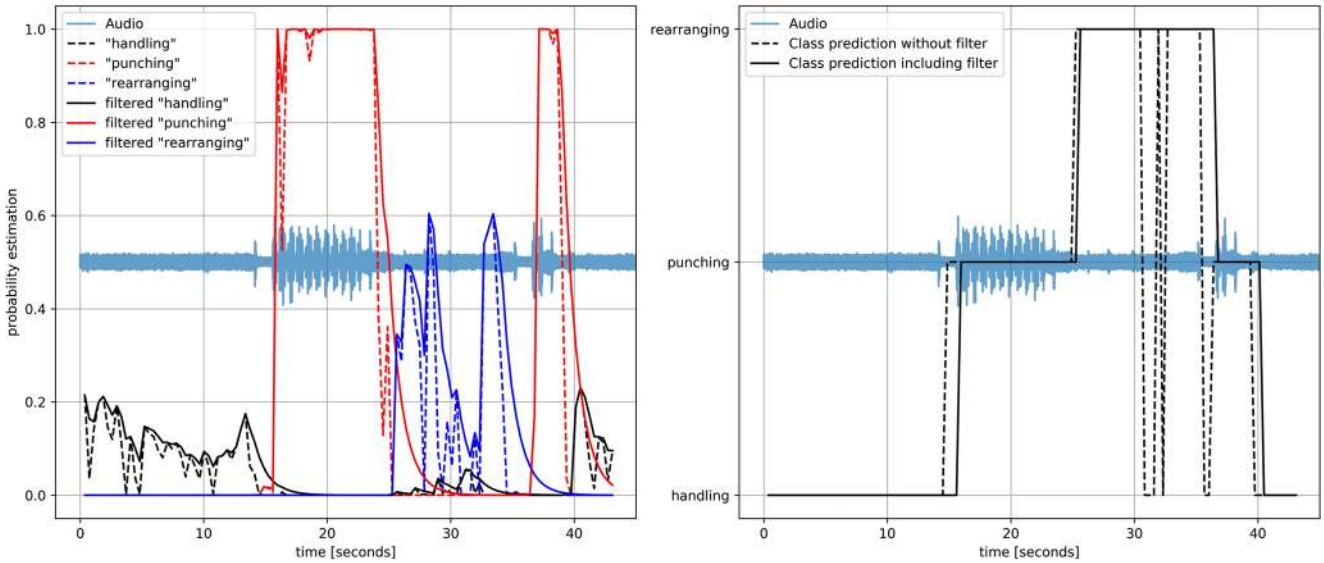

Fig. 10 The diagram on the left side displays the three probability estimates of the three classes over time, filtered and unfiltered ("smooth down"). The diagram on the right side shows the resulting classification with ("phase_f") and without ("phase") filtering (cf. Fig. 8)

higher (> 94%) than the one achieved in the first proof-of-concept study (> 93%). The robustness of the model could be further improved by the implementation of an IIR filter.

3. The three states of this classifier representing the three operation phases were acoustically displayed by characteristic and clearly distinguishable instrumentations of a musical piece. An intrusion into the auditory scene of operators is not expected as long as they are accustomed to listen to music during working hours—as many operators do according to our observations.

4 Discussion and conclusion

We presented two proof-of-concept approaches on the sonification of estimated error conditions of 3D printing processes and operation phase classification of punching processes. The results of these studies form the basis for the development of an "intelligent" noise protection headphone as part of Cyber Physical Production Systems (CPPSs) which provides auditorily augmented information to machine operators and enables radio communication between them. Further application areas for these auditory displays will be their implementation in control rooms (equipped with multi-channel loudspeaker systems) and their utilization for training purposes.

Table 6 Mapping of operation phases to musical instrumentation

Operation phase	Instrumentation	
[0] handling	Muted trumpet, piano	Plucked double bass, drums
[1] punching	Lead and rhythm guitar	
[2] re-arranging	Lead and rhythm guitar, synthetic choir (male voices)	

The focus of our research lies on situation-specific acoustic processing of conditioned machine sounds and operation related data with high information content, considering the often highly auditorily influenced working knowledge of skilled workers. One crucial aspect of continuous sonification for process monitoring in the context of shop floors is the willingness of operators to accept exposures to the provided acoustic information on a day-to-day 8-h basis. Having background in both, manufacturing and auditory display, our observations and experiences let us assume that offering the selection of arbitrary music (which operators are listening to anyway) will have a high acceptance rate and therefore a good chance for real-world implementations. Since our project primarily addresses noise production environments (> 85 dB SPL), where operators are obliged to use noise protection devices anyway, there will be no constraints by wearing additional equipment. According to [42], acoustic features integrated in assistance systems, such as attenuation of generated vibrations or adaptation of sound absorbers, are supportive to the well-being and motivation of employees. However, acceptance and benefit can only be evaluated in long-term studies, which were out of the scope of this exploratory study and which will need a much more robust database for reliable error prediction.

The results of the presented studies indicate the feasibility of our long-term proposition to develop an "intelligent" headphone to be used in industrial environments. This concerns the identification of error conditions and operation phases (or states) as well as the design of meaningful and at the same time unobtrusive auditory displays. While audio effects representing the gradual impact of the errors can be applied rather simple to existing music playlists, an

appropriate solution for the representation of operation phases faces major challenges, since operators should be able to select music according to their listening preferences. A feasible solution could be the implementation of music-related artificial intelligence that is capable of creating genre and style-specific tunes without being too repetitive and, at the same time, is capable of providing characteristic musical attributes that are non-judgmental and do not affect the sonic quality.

The development and comparison of sonification models themselves was not a primary focus of our project. The unique selling proposition of the presented project is the combination of process analysis based on acoustic emission and machine learning with auditory display. We used machine learning algorithms to simplify highly complex data to a one-dimensional data stream that could easily be transformed to a stream of auditory information. As a next step, we will extend the approach of general error identification and pursue a comprehensive identification of distinctive machine and operation states and conditions at classification rates similarly high to the ones achieved within our proof-of-concept studies, also considering alternative algorithms [50]. In order to gain more knowledge about the flexibility, stability, and reliability of our custom-built classification models, a large database is required for reliable evaluations. Therefore, long-term monitoring and recording facilities must be installed for data collection in industrial environments. In addition, sound source separation issues [51], which may be needed in more complex shop floor scenarios, will be taken into account as well as aspects of sound spatialization for a position-adjusted display of auditory scenes.

Acknowledgments We wish to thank our colleagues Matthias Zeppelzauer and Djordje Slijepčević for their feedback on deep learning methodologies as well as Franziska Bruckner and Georg Vogt for their support on the project.

Funding information Open access funding provided by FH St. Pölten - University of Applied Sciences. Our research is funded by the Austrian Ministry of Digital and Economic Affairs within the framework of the FFG COIN project *Immersive Media Lab* (866856).

References

1. Ullrich C, Hauser-Ditz A, Kreggenfeld N, Prinz C, Igel C (2018) Assistenz und Wissensvermittlung am Beispiel von Montage- und Instandhaltungstätigkeiten. In: Wischmann S, Hartmann EA (eds) Zukunft der Arbeit – Eine praxisnahe Betrachtung. Springer Berlin Heidelberg, Berlin, pp 107–122
2. Stocker A, Brandl P, Michalczuk R, Rosenberger M (2014) Mensch-zentrierte IKT-Lösungen in einer Smart Factory. E Elektrotechnik Informationstechnik 131(7):207–211. https://doi.org/10.1007/s00502-014-0215-z
3. Stocker A, Spitzer M, Kaiser C, Rosenberger M, Fellmann M (2017) Datenbrillengestützte Checklisten in der Fahrzeugmontage: Eine empirische Untersuchung. Inform.-Spektrum 40(3):255–263. https://doi.org/10.1007/s00287-016-0965-6
4. Fantini P, Pinzone M, Taisch M (2018) Placing the operator at the centre of Industry 4.0 Design: modelling and assessing human activities within cyber-physical systems. Comput Ind Eng: S0360835218300329. https://doi.org/10.1016/j.cie.2018.01.025
5. Harteis C, Fischer C (2018) Wissensmanagement unter Bedingungen von Arbeit 4.0. In: Maier GW, Engels G, Steffen E (eds) Handbuch Gestaltung digitaler und vernetzter Arbeitswelten. Springer Berlin Heidelberg, Berlin, pp 1–18
6. Klueber S, Wolf E, Grundgeiger T, Brecknell B, Mohamed I, Sanderson P (2019) Supporting multiple patient monitoring with head-worn displays and spearcons. Appl Ergon 78:86–96. https://doi.org/10.1016/j.apergo.2019.01.009
7. Iber M (2020) Auditory display in workspace environments. In: Filimowicz M (ed) Foundations in sound design for embedded media: a multidisciplinary approach. Routledge, New York, pp 131–154
8. Hildebrandt T, Hermann T, Rinderle-Ma S (2016) Continuous sonification enhances adequacy of interactions in peripheral process monitoring. Int J Hum Comput Stud. 95:54–65. https://doi.org/10.1016/j.ijhcs.2016.06.002
9. Yan R, Gao RX (2006) Hilbert–Huang transform-based vibration signal analysis for machine health monitoring. IEEE Trans Instrum Meas 55(6):2320–2329. https://doi.org/10.1109/TIM.2006.887042
10. Elmaleeh MAA, Saad N, and Awan M (2010) Condition monitoring of industrial process plant using acoustic emission techniques, in 2010 International Conference on Intelligent and Advanced Systems, Manila, Philippines, pp. 1–6, doi: https://doi.org/10.1109/ICIAS.2010.5716110
11. Goel S, Ghosh R, Kumar S, Akula A (2014) A methodical review of condition monitoring techniques for electrical equipment. NDE-India:8
12. Zafar T, Kamal K, Sheikh Z, Mathavan S, Jehanghir A, and Ali U (2015) Tool health monitoring for wood milling process using airborne acoustic emission, in 2015 IEEE International Conference on Automation Science and Engineering (CASE), pp. 1521–1526, doi: https://doi.org/10.1109/CoASE.2015.7294315
13. Pasha S, Ritz C, Stirling D, Zulli P, Pinson D, and Chew S (2018) A deep learning approach to the acoustic condition monitoring of a sintering plant, in 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Honolulu, Hawaii, USA, pp. 1803–1809, doi: https://doi.org/10.23919/APSIPA.2018.8659486
14. Industrial Media Applications - Fraunhofer IDMT, Fraunhofer Institute for Digital Media Technology IDMT. [Online]. Available: https://www.idmt.fraunhofer.de/en/business_units/ima.html. [Accessed: 12-May-2019]
15. Sodnik J, Tomažič S (2015) Auditory interfaces. Spat Audit Hum Comput Interfaces:33–44. https://doi.org/10.1007/978-3-319-22111-3_3

16. Horsak B, Dlapka R, Iber M, Gorgas AM, Kiselka A, Gradl C, Siragy T, Doppler J (2016) SONIGait: a wireless instrumented insole device for real-time sonification of gait. J Multimodal User Interfaces 10(3):195–206. https://doi.org/10.1007/s12193-016-0216-9
17. Gorgas A-M et al (2016) Short-term effects of real-time auditory display (sonification) on gait parameters in people with Parkinsons' disease—a pilot study. In: Converging Clinical and Engineering Research on Neurorehabilitation II, Segovia, pp 855–859. https://doi.org/10.1007/978-3-319-46669-9_139
18. Berner B (2008) Working knowledge as performance: on the practical understanding of machines. Work Employ Soc 22(2):319–336. https://doi.org/10.1177/0950017008089107
19. Reeves BN, Shipman F (1996) Tacit knowledge: icebergs in collaborative design. SIGOIS Bull 17(3):24–33. https://doi.org/10.1145/242206.242212
20. Vickers P (2011) Sonification for process monitoring. In: Hermann T, Hunt A, Neuhoff JG (eds) The sonification handbook. Logos, Berlin, pp 455–491
21. Maglio PP and Campbell CS (2000) Tradeoffs in displaying peripheral information, in Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp. 241–248
22. Patterson RD, Mayfield TF (1990) Auditory warning sounds in the work environment [and discussion]. Philos Trans R Soc B Biol Sci 327(1241):485–492. https://doi.org/10.1098/rstb.1990.0091
23. Edworthy J et al (2017) The recognizability and localizability of auditory alarms: setting global medical device standards. Hum Factors 59(17):1108–1127. https://doi.org/10.1177/0018720817712004
24. Paterson E, Sanderson PM, Paterson NAB, Liu D, Loeb RG (2016) The effectiveness of pulse oximetry sonification enhanced with tremolo and brightness for distinguishing clinically important oxygen saturation ranges: a laboratory study. Anaesthesia 71(5):565–572. https://doi.org/10.1111/anae.13424
25. Viraldo J, Caldwell B (2013) Sonification as sensemaking in control room applications. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol 57, Los Angeles, pp 1423–1426. https://doi.org/10.1177/1541931213571318
26. Johannsen G (2004) Auditory displays in human-machine interfaces. Proc IEEE 92(4):742–758. https://doi.org/10.1109/JPROC.2004.825905
27. Hearst MA (1997) Dissonance on audio interfaces. IEEE Expert 12(5):10–16. https://doi.org/10.1109/64.621221
28. Chamberland C, Hodgetts HM, Vallières BR, Vachon F, Tremblay S (2017) The benefits and the costs of using auditory warning messages in dynamic decision making settings. J Cogn Eng Decis Mak. https://doi.org/10.1177/1555343417735398
29. Watson M (2006) Scalable earcons: bridging the gap between intermittent and continuous auditory displays. In: Proceedings of the 12th International Conference on Auditory Display, London
30. Sanderson P, Anderson J, and Watson M (2000) Extending ecological interface design to auditory displays, in Proceedings of the 10th Australian Conference on Computer-Human Interaction, pp. 259–266
31. Baldwin CL et al (2012) Multimodal cueing: the relative benefits of the auditory, visual, and tactile channels in complex environments. Proc Hum Factors Ergon Soc Annu Meet 56(1):1431–1435. https://doi.org/10.1177/1071181312561404
32. Hildebrandt T, Hermann T, Rinderle-Ma S (2014) A sonification system for process monitoring as secondary task. In: 2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom), pp 191–196. https://doi.org/10.1109/CogInfoCom.2014.7020444
33. Haas EC, van Erp JBF (2014) Multimodal warnings to enhance risk communication and safety. Saf Sci 61:29–35. https://doi.org/10.1016/j.ssci.2013.07.011
34. Barra M et al. (2001) Personal webmelody: customized sonification of web servers, in Proceedings of 2001 Conference on Auditory Display, Espoo
35. Mynatt ED, Back M, Want R, Frederick R (1997) Audio aura: lightweight audio augmented reality. In: Proceedings of the 10th Annual ACM Symposium on User Interface Software and Technology, New York, pp 211–212. https://doi.org/10.1145/263407.264218
36. Bovermann T, Hermann T et al (2010) Auditory augmentation. Int J Ambient Comput Intell IJACI 2(2):27–41. https://doi.org/10.4018/jaci.2010040102
37. Gross-Vogt K, Weger M, and Höldrich R (2018) Exploration of Auditory Augmentation in an Interdisciplinary Prototyping Workshop, presented at the Proceedings of the 11th Forum Media Technology and 4th All Around Audio Symposium, St. Pölten, pp. 10–16
38. Grosshauser T, Hermann T (2010) Multimodal closed-loop human machine interaction. In: Proceedings of the 3rd International workshop on Interactive Sonification, Stockholm, pp 59–63. https://doi.org/10.4119/unibi/2698347
39. Liang J and Wang K (2017) Vibration feature extraction using audio spectrum analyzer based machine learning, in 2017 International conference on information, Communication and Engineering (ICICE), pp. 381–384, doi: https://doi.org/10.1109/ICICE.2017.8479273
40. Grebenik J, Zhang Y, and Bingham C (2016) Roller element bearing acoustic fault detection using smartphone and consumer microphones. 2016 17th International Conference on Mechatronics - Mechatronika (ME)
41. Yang B-S, Hwang W-W, Kim D-J, Tan AC (2005) Condition classification of small reciprocating compressor for refrigerators using artificial neural networks and support vector machines. Mech Syst Signal Process 19(2):371–390
42. McFee B et al (2015) Librosa: audio and music signal analysis in Python, presented at the Python in Science Conference, Austin, pp 18–24. https://doi.org/10.25080/Majora-7b98e3ed-003
43. Pedregosa F et al (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830
44. Liu H, Setiono R (1995) Chi2: Feature selection and discretization of numeric attributes. In: Proceedings of 7th IEEE International Conference on Tools with Artificial Intelligence, Herndon, pp 388–391. https://doi.org/10.1109/TAI.1995.479783
45. Keras Team, Keras. GitHub (2015)
46. Abadi M et al. (2015) TensorFlow: large-scale machine learning on heterogeneous systems
47. De Campo A (2007) Toward a data sonification design space map, in Proceedings of the 13th Conference on Auditory Display, Montreal, pp. 342–347
48. Farnell A (2010) Designing sound. Mit Press
49. Schnell N, Röbel A, Schwarz D, Peeters G, and Borghesi R MuBu & Friends - assembling tools for content based real-time interactive audio processing in Max/MSP, p. 4
50. Schröder J, Anemüller J, and Goetze S (2016) Performance comparison of Gmm, Hmm and Dnn based approaches for acoustic event detection within task 3 of the Dcase 2016 Challenge, in Proc. Workshop Detect. Classification Acoust. Scenes Events, pp. 80–84
51. Cano E, Nowak J, and Grollmisch S (2017) Exploring sound source separation for acoustic condition monitoring in industrial scenarios, in 2017 25th European Signal Processing Conference (EUSIPCO), pp. 2264–2268