Journal of Intelligent Manufacturing (2023) 34:3039–3052
https://doi.org/10.1007/s10845-022-01992-3

A digital apprentice for chatter detection in machining via human–machine interaction

Xiaoliang Yan[1] · Shreyes Melkote[1] · Anant Kumar Mishra[2] · Sudhir Rajagopalan[2]

Received: 7 October 2021 / Accepted: 6 July 2022 / Published online: 4 August 2022

Abstract
Regenerative chatter in machining operations such as milling is a common process anomaly that limits productivity and part quality, which in turn lead to increased manufacturing costs. The industrial relevance of the problem has sparked many research efforts over the recent decades, with a growing interest in real-time chatter detection and suppression. Inspired by learning from human demonstration frameworks, this paper proposes a new approach to milling chatter detection via effective human–machine interaction, which facilitates knowledge transfer from an *experienced* machine tool operator to a "Digital Apprentice." The proposed chatter detection approach acquires chatter-specific knowledge through a learnable skill primitive (LSP) algorithm designed to establish a robust chatter detection threshold from few-shot real-time demonstrations by an experienced human operator. In this work, digital audio data were acquired from milling experiments through a microphone mounted inside the milling machine. During the training phase, data for the human operator's natural reaction to chatter were collected via a specially designed human–machine interface. The learned chatter detection thresholds were obtained via the LSP algorithm by temporally mapping the reaction time data to the audio signal. During the testing phase, experiments were conducted to validate the detection accuracy and detection speed of the learned chatter detection thresholds under different cutting conditions. The experimental validation results of the learned thresholds indicate an average chatter detection accuracy of 94.4%, with 55.6% of chatter cases detected before chatter marks are produced on a 4140 Steel workpiece, thus demonstrating the effectiveness of human–machine interaction in chatter detection.

Keywords Real-time process monitoring · Human–machine interaction · Learning from demonstration · Machining chatter

Introduction

Regenerative chatter is a common process anomaly that occurs in machining operations such as milling (Tobias, 1964). Chatter occurs when the cutting tool engages the wavy surface left by the previous tooth pass, where the phase shift under certain conditions amplifies the instantaneous variation in chip load, causing the cutting force to vary dynamically, resulting in large amplitude vibrations (Smith & Tlusty, 1991). Machining chatter negatively impacts part surface quality and productivity in milling operations, which in turn lead to increased manufacturing costs. According to (Quintana & Ciurana, 2011), a major European automotive engine manufacturer estimated the added cost due to chatter during engine cylinder block machining to be 0.35 Euros per engine block, which is significant when scaled over 3 million engines produced in a year. For these reasons, research efforts over the past five decades have focused on developing solutions for chatter avoidance (Altintaş & Budak, 1995; Tlusty & Polacek, 1968) and chatter suppression (Altintas & Chan, 1992; Delio et al., 1992). Chatter avoidance through off-line identification of the system dynamics to construct the stability lobe diagram remains the primary approach (Altintaş & Budak, 1995; Duncan et al., 2000; Smith & Tlusty, 1993), where the oriented transfer function of the machine-tool holder-cutting tool system and the cutting stiffness of the workpiece material are required. However, such system identification requires specialized equipment, knowledge, and skills that may not be readily available to all machine tool users. On the other hand, on-line chatter detection and suppression methods offer an alternative approach to off-line chatter avoidance (Quintana & Ciurana, 2011).

✉ Xiaoliang Yan
mike.yan@gatech.edu

1 George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

2 Siemens Corporate Technology, Charlotte, NC, USA

On-line chatter detection and suppression involves monitoring the dynamic stability of the machining process through sensors and chatter detection algorithms without explicit identification of the machining system dynamics (Caixu et al., 2019; Quintana & Ciurana, 2011). Various sensor signals such as vibration (Kuljanic et al., 2008; Yesilli et al., 2020; Yuqing et al., 2015), strain (Cen et al., 2018; Luo et al., 2018; Ma et al., 2014), sound (Cao et al., 2017; Delio et al., 1992; Schmitz et al., 2002), cutting force (Liu et al., 2017; Ma et al., 2013; Nguyen et al., 2016), and current (Aslan & Altintas, 2018; Lamraoui et al., 2015; H. Liu et al., 2011) have been researched as potential measurands for chatter detection. Chatter detection methods based on frequency domain analysis (Delio et al., 1992; Jardine et al., 2006; Liao & Young, 1996; Nguyen et al., 2016), time domain analysis (Berger et al., 1997; Khorasani et al., 2014; Ma et al., 2013), and machine learning techniques (Rahimi et al., 2021; Shi et al., 2020; Tran et al., 2020; Yao et al., 2010; Zhang et al., 2010) have demonstrated varying levels of success, but can suffer from excessive false positives and late chatter detection after damage to the workpiece or the machine tool has already occurred (Quintana & Ciurana, 2011). Because cutting signals acquired from the machining operation vary with sensor types and locations, cutting tools, and machining conditions, to achieve acceptable detection accuracies and speed, on-line chatter detection algorithms that utilize machine learning methods such as support vector machine (Yesilli et al., 2020) and logistic regression (Ding et al., 2017) require a large number of costly chatter experiments for training; frequency and time domain chatter detection methods require an effective chatter detection threshold, which is an engineering parameter that must be tuned by engineers through similarly costly trial-and-error experiments to obtain the desired detection accuracy, speed, and robustness (Bachrathy et al., 2021; Faassen et al., 2006; Ma et al., 2013; Wright & Bourne, 1988). Tuning of the chatter detection threshold, which involves setting *a numerical value* using specialized equipment, is non-trivial. Research engineers who have the experience and skills for tuning are not always available on the shop floor, and machine tool operators typically do not possess the necessary skills for tuning the threshold (Wright & Bourne, 1988). For these reasons, on-line chatter detection methods have not found widespread acceptance in real production settings.

As a result, manufacturers continue to rely heavily on the experience of human machine tool operators to monitor and control the machining process. However, the operator's expertise and cognitive capabilities are not easily scalable or directly transferrable, and the industry continues to face difficulty in replacing experienced operators when they retire or leave.

Ultimately, the current state-of-the-art methods for on-line chatter detection require either a large number of chatter examples for training (Rahimi et al., 2021; Shi et al., 2020; Tran et al., 2020) or extensive trial and error experiments for tuning the chatter detection threshold, both of which are costly. Inspired by advances in learning from human demonstration in the field of robotics, where the potential to significantly reduce the amount of training data needed for learning a particular skill has been shown (Knox & Stone, 2009), this paper proposes human–machine interaction to effectively reduce the number of chatter examples needed to learn the chatter detection threshold. Analogous to human apprenticeship, an interactive learning agent is proposed, referred to here as the "Digital Apprentice", to capture the underlying chatter detection knowledge from experienced human operators for effective and robust chatter detection via human–machine interaction. The Digital Apprentice comprises a sensor (e.g. microphone,) a Learnable Skill Primitive (LSP) algorithm for chatter detection, and a human–machine interface (HMI). The sensor corresponds to the human apprentice's natural perception ability, the LSP emulates the human's cognitive ability to distinguish chatter from stable cutting, and the HMI facilitates demonstration of the skill by an experienced operator to the Digital Apprentice. The paper seeks to answer two specific questions: (1) Is it possible for the Digital Apprentice to learn to detect chatter from experienced human operator's demonstrations? and (2) What is the accuracy and speed of such a chatter detection method? Because repeated occurrences of machining chatter are costly and may result in damage to the part and/or machine tool, the objective is to learn to detect chatter from few-shot demonstrations by an experienced machine tool operator.

The rest of the paper is organized as follows. "Chatter detection and learning from demonstration (LfD)" section frames the chatter detection threshold training problem as a learning from demonstration problem. "Learnable skill primitive (LSP) for chatter detection" section introduces the Learnable Skill Primitive (LSP) algorithm by which the correspondence problem between the human operator's perception of chatter and the Digital Apprentice's ability to perceive chatter is addressed. "Experimental verification" section presents experimental validation of the LSP algorithm. The paper concludes with an evaluation of the effectiveness of the approach and provides recommendations for future work.

Chatter detection and learning from demonstration (LfD)

While the use of human–machine interaction techniques is comparatively new in the field of manufacturing process monitoring, LfD techniques (Knox & Stone, 2009; Schaal, 2006; Schaal et al., 2003; Warnell et al., 2018) have been successfully employed in the fields of interactive computing and robotics (Chernova & Thomaz, 2014).

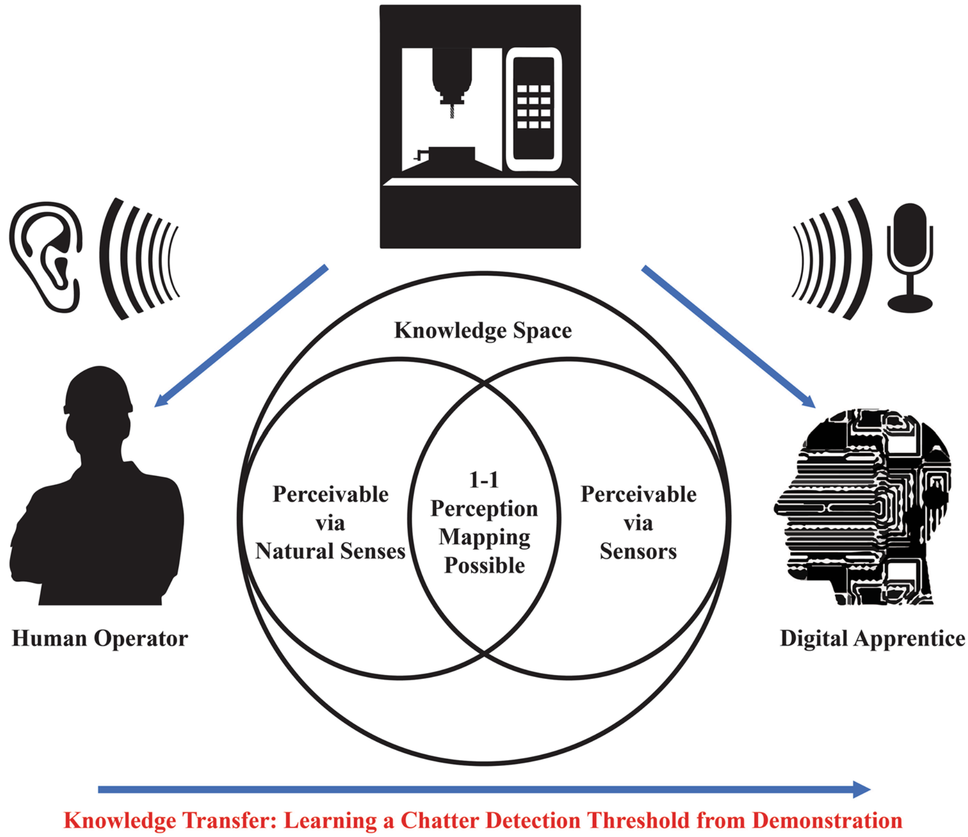


Fig. 1 Chatter detection as an LfD problem: knowledge space and perception mapping between experienced human machine tool operator and the Digital Apprentice

The premise of LfD is that learning a skill from scratch without any prior knowledge is challenging and impractical (Schaal, 1997). The specific objective of LfD is to enable a learning agent to learn a policy π, which is a mapping between states S and actions A, from a single or very few human demonstrations (Schaal, 1997). The policy $\pi : S \rightarrow A$ selects the actions based on the states observed by the learning agent (Argall et al., 2009). In the context of machining, the experienced machine tool operator is responsible for recognizing the state of the process (e.g. chatter or stable cutting) and executing the necessary corrective action, which include adjusting the cutting conditions in real time or halting the process to prevent damage to the part and/or the machine tool. As illustrated in Fig. 1, the operator perceives the cutting signals using his/her natural senses, and the learning agent i.e. the Digital Apprentice learns from the human operator and perceives the state of the machining process through suitable sensors. In this paper, the policy π learned by the Digital Apprentice is a value of the chatter detection threshold of the audio signal that discriminates the audio signals emitted by the milling process into either unstable (chatter) or stable cutting; the frequency and amplitude of the audio signal represent features of states S, and the actions A correspond to classification of the milling process as either unstable (chatter) or stable. An extension of this policy would be to expand the set of actions to include a corrective control policy to suppress chatter. The focus of this paper, however, is to learn a chatter detection threshold from human demonstration, which is a prerequisite for chatter suppression.

Another challenge in modelling chatter detection as a LfD problem is the temporal correspondence problem between the human's perception of chatter and the learning agent's perception via sensors and suitable signal processing algorithms (Nehaniv & Dautenhahn, 2002). Since chatter is a process anomaly that can develop and grow rapidly, and because a human operator's response to chatter is delayed due to his/her reaction time, by the time the human operator perceives chatter and reacts to it, the process state S has already changed significantly. Effective LfD requires mapping the human operator's perception of chatter to the corresponding states of the sensor signal. This requires the learning process to account for the operator's delayed reaction. A similar correspondence problem is discussed by Knox and Stone (2008) who address the shaping problem in psychology and apply it to interactive learning in robotics. In their work, the human trainer provides a positive or a negative reward using a clicker to evaluate the robot's actions. By accounting for the delay in the human trainer's clicker signal via a credit assigner, the human reward is mapped to a set of corresponding state-action pairs, which allows the human to interactively shape a learning agent's policy (Knox & Stone, 2009). In the context of machining chatter, the Digital Apprentice must pinpoint the time instance at which the operator first perceives chatter. The learning agent can then temporally identify the corresponding sensor signal and set a chatter detection threshold by accounting for the delay in the operator's demonstration.

Evaluating the correspondence problem in chatter detection

A pre-requisite for accurate and timely detection of chatter when learning from the operator's demonstration is to determine the correspondence between the occurrence of chatter marks on the workpiece and the human operator's observation of chatter. To solve this correspondence problem, we first estimated human operators' reaction times to the onset of chatter by measuring the distance between the start of cut at the edge of the workpiece to the location of the chatter marks produced in actual milling experiments. These reaction times were then compared to the same operators' reaction times to synthetically generated sounds that emulate chatter. The experimental procedures employed are as follows.

Milling experiments were conducted to obtain the operators' reaction times to chatter marks during milling operations. The experiments were performed on a 3-axis CNC milling machine (Okuma MILLAC 44 V) using a 12.7 mm diameter four flute solid carbide end mill of 25.4 mm cutting length (Niagara Cutter Series C430, Single End, TiAlN Finish, Spiral Flute, 30° Helix), which was held in an Iscar CAT 40 toolholder. A USB studio style condenser microphone (Blue Yeti model number 988-000101) was mounted inside the milling machine as shown in Fig. 2. Multiple passes of dry slot end milling experiments, i.e. 100% radial immersion, were conducted on a AISI 4140 Steel workpiece (Cold Finished ASTM A108 Steel bar, 84.1 HRB Rockwell B Hardness) at different axial depths of cut (a_p) and spindle speeds (N) to generate various chatter and stable cutting signals during operator demonstrations. The feed per tooth was fixed at 0.0330 mm. The selected microphone provides comparable frequency responses across a wide range of frequencies from 20 Hz to 20 kHz, which is suitable for chatter monitoring. The HMI for operator demonstrations comprises control buttons, a microcomputer, and a display for visualization. A Raspberry Pi 4 Model B (CanaKit) was selected as the microcontroller to process signals acquired from the microphone. The Raspberry Pi Unit in this work has 4 GB of RAM and a quad-core processor, making it suitable for complex on-line signal processing and visualization. The 16-bit digital audio signals sampled at 48 kHz were acquired through the Raspberry Pi's default advanced Linux sound architecture (ALSA) and PyAudio, an open-source python audio processing package (Pham, 2006). The 16-bit digital signals were converted into a series of integers ranging between -2^{15} and $\left(2^{15}-1\right)$.

Here we denote the true time instance when a chatter mark first appeared on the workpiece as t_c. The distance between the first chatter mark and the start of cut was measured and t_c was calculated from the feed rate and the time instance of tool entry into the workpiece. Figure 3 shows the correspondence between t_c and the chatter marks.

During the slot milling experiments, human operators were instructed to react to the sound of chatter by providing a pushbutton signal, from which a corresponding time instance t_s was recorded (see Fig. 2). The experiments were conducted per the Georgia Institute of Technology Institutional Review Board (IRB) Protocol H20340 (Melkote, 2020). In total, three human operators with different levels of machining experience were recruited (see Table 7 in Appendix). Prior to each milling experiment, the human subjects had no knowledge of the selected process parameters i.e. spindle speed and depth of cut, or whether chatter would occur or not. The order of milling experiments was randomized. Table 1 lists the operators' experimental reaction times to occurrence of chatter, r_c, and its corresponding sample mean and standard deviation; r_c is defined as:

$$r_c = t_s - t_c \tag{1}$$

Based on the experimental data given in Table 1, the mean reaction time determined from actual slot milling experiments is 0.345 s with a sample standard deviation of 0.240 s.

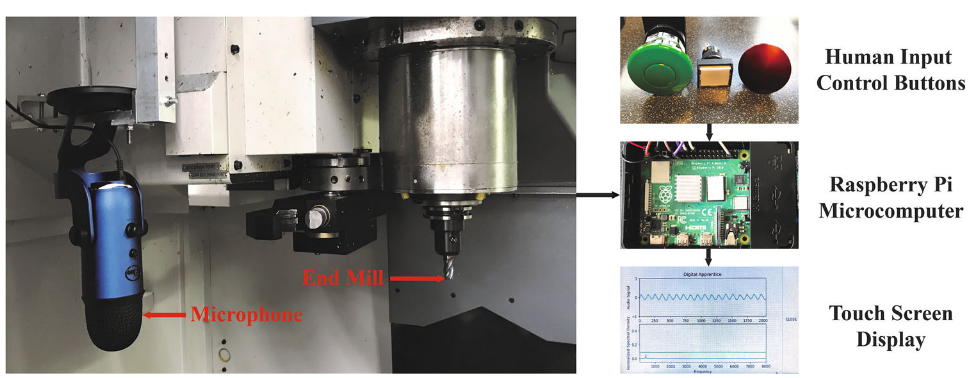


Fig. 2 Experimental setup, HMI, and data flow

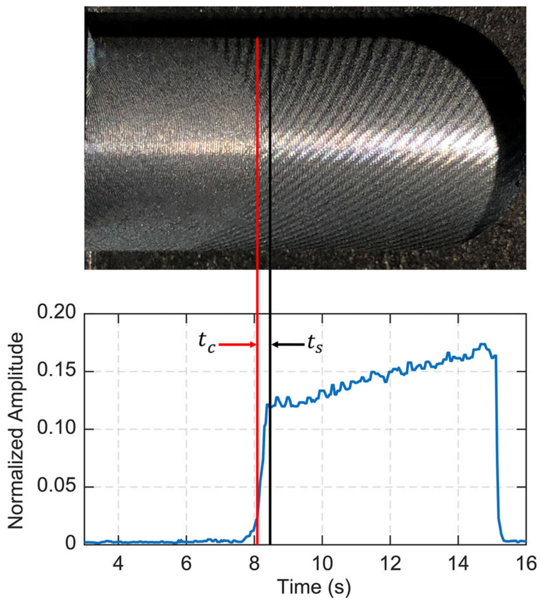


Fig. 3 Chatter mark time instance t_c relative to the time instance the operator signals chatter t_s; the time series shown is bandpass-filtered and normalized amplitude, $P_{band}(t)$, as described in "Filter, transform, and normalize the audio signal" section

Table 1 Experimental operator reaction times to occurrence of chatter

Operator	N(RPM)	a_p(mm)	r_c(s)
A	750	3.5	− 0.199
A	2000	2	0.370
A	2000	2.5	0.675
B	750	3.5	0.206
B	2000	3	0.447
C	1000	2.5	0.376
C	1000	3	0.430
C	1750	2	0.484
C	1750	2.5	0.319
Mean			0.345
Standard deviation			0.240

Although the above experiments can be conducted in a laboratory environment, in practice, the experiments required to evaluate the operator's reaction time to chatter could be time consuming and costly since the chatter marks need to be physically measured to determine t_c. Assuming operators primarily utilize their hearing ability to detect chatter, a simpler approach is to estimate the operator's mean reaction time to chatter offline through synthetically generated sounds at the relevant chatter frequencies as described next.

Here we denote r as the human operator's reaction time to a synthetically generated sound simulating specific chatter frequencies. Research on human subjects reaction times in visual search problems has shown that it can be described by a probability density function such as Gaussian, ex-Gaussian, ex-Wald, or the Gamma distribution (Palmer et al., 2011). The auditory reaction time r in this paper is assumed to be described by a Gaussian distribution. The operators were instructed to wear a headphone, through which samples of computer-generated sounds mimicking different chatter frequencies lasting one second each were played. The synthetic sounds were generated using PyAudio (Pham, 2006), which is an audio processing library for the Python programming language, with a specified frequency and duration. The operators were asked to respond to the sounds by depressing a control button like that shown in Fig. 2 as soon as they heard the sound. The exact time instance of the synthetically generated sound was recorded. Each operator responded to 90 samples of synthetically generated sounds that simulated three chatter frequencies of 500 Hz, 2000 Hz, and 3500 Hz, which are in the range of milling chatter frequencies reported in the literature (Delio, 1989). The samples of synthetically generated sounds were played in randomized order with randomized pause intervals ranging from 2 to 4 s between samples. This randomization ensured that the operators had no anticipation of when a synthetic sound would be played. Figure 4 is an example of the computer-generated time domain audio signal and the corresponding Fast Fourier Transform (FFT). Figure 5 shows a histogram of the reaction times obtained from the 90 samples collected from a representative operator.

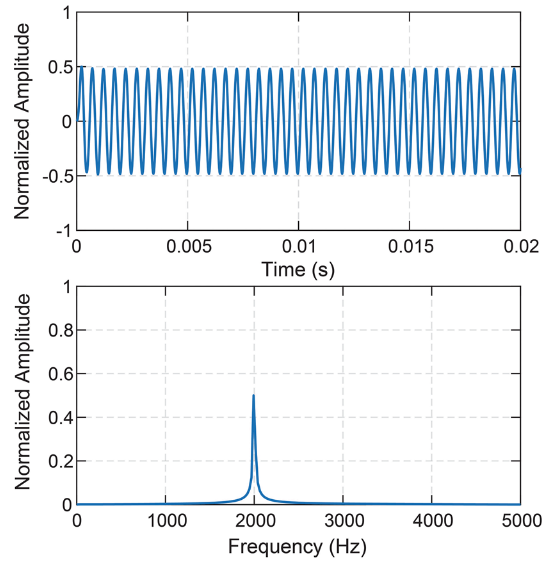


Fig. 4 Time and frequency domain representations of the 2000 Hz synthetically generated sound

Based on the experimental data given in Table 2, the mean sample reaction time (for all operators) obtained from the synthetically generated sound experiments was 0.341 s with a sample standard deviation of 0.070 s. These statistics

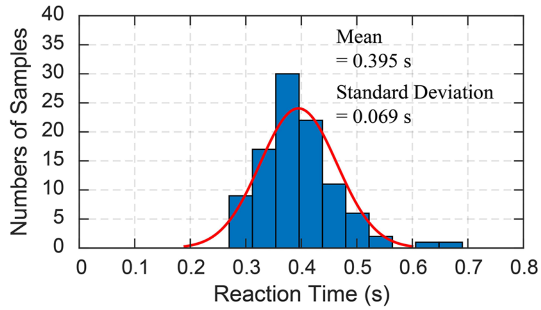


Fig. 5 Reaction time distribution based on 90 reaction time samples from Operator C

compare favorably with the mean operator reaction time of 0.345 s and sample standard deviation of 0.240 s obtained from the actual slot milling chatter experiments reported in Table 1. We therefore conclude that human subject experiments with simulated chatter sounds offer an inexpensive method to obtain a sufficiently accurate estimate of an experienced operator's mean reaction time to chatter, which we employ in our methodology for learning the chatter detection threshold.

Specifically, the quantities used in the LSP algorithm presented next are the sample mean reaction time,$\bar{r}$, and its standard deviation,s_r, obtained from each human operator. In the training phase, the objective of the learning agent or digital apprentice is to learn the key features of chatter from *few-shot demonstrations* given by an experienced operator; in the testing phase, the learning agent applies the learned features to detect chatter on-line. The two features for the learning agent to learn from the audio signal are the chatter frequency and the chatter detection threshold.

Learnable skill primitive (LSP) for chatter detection

Figure 6 is a flowchart of the LSP algorithm. The LSP takes inputs from the audio microphone and an experienced operator's demonstration of chatter detection and outputs the dominant chatter frequency $f_{chatter}$ and the chatter detection threshold P_{th}. The key steps of the LSP algorithm are described next.

Determining the dominant chatter frequency

The first step of the LSP algorithm is to determine $f_{chatter}$ from the audio signal and the operator's demonstration. A sampling rate of 48 kHz was utilized to ensure that the entire frequency range of human hearing (20 Hz–20 kHz) is covered. The audio signal was transformed into the frequency domain using FFT, which enabled the extraction of frequencies and their amplitudes as features. The frequencies and amplitudes correspond to the human operator's perception of pitch and volume, respectively.

As noted earlier, the operator provided a demonstration by pressing a push button switch, upon perceiving chatter in the milling operation. Because our approach relies on an experienced operator's reaction to chatter, the learning algorithm only attempts to search for a chatter frequency after receiving the push button signal from the operator. It thus eliminates the need to filter out the tooth passing frequency and its harmonics to isolate the chatter frequency. At the human signal time instance t_s, the highest peak in the FFT of the corresponding audio signal is taken as the dominant chatter frequency. Figure 7 shows the FFT of the audio signal at t_s during a dry slot milling experiment, which clearly identified the dominant chatter frequency of approximately 774 Hz as the highest peak in the frequency spectrum of the audio signal at t_s. Note that the chatter frequency amplitude has increased significantly at time t_s due to the natural delay in the human operator's reaction to perception of chatter.

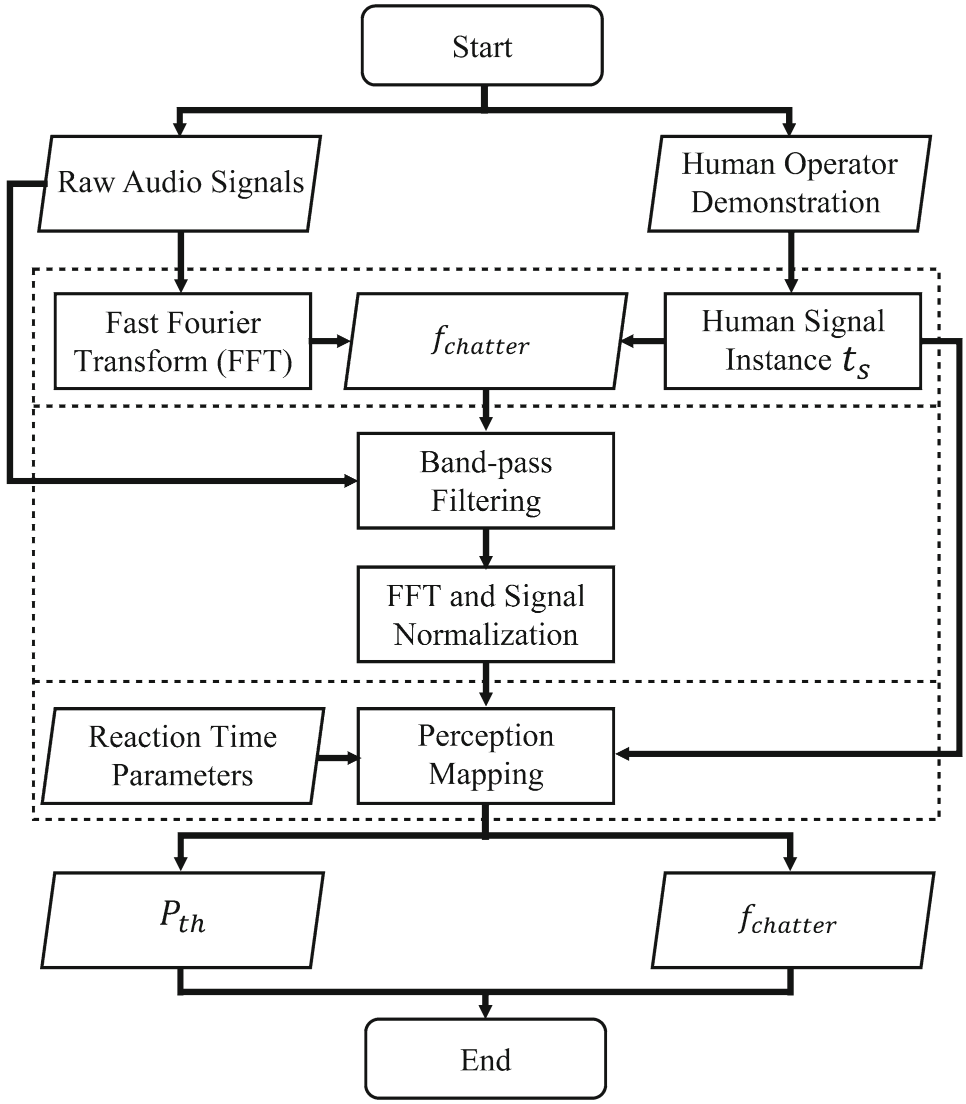
Filter, transform, and normalize the audio signal

Analogous to short-term memory, in this work approximately 1 s duration of the audio signal was stored in the microcontroller of the HMI at any time instant. The stored time-domain audio signal was filtered by a bandpass filter with a bandwidth of 200 Hz centered around $f_{chatter}$ identified in the previous step. An FFT with a bin size of $N = 2048$ was sequentially applied to the filtered signal to extract the frequency decompositions and amplitudes. Finally, the maximum value in each bin of the FFT was collected as a time-series, which was normalized by dividing it by $2^{15} \cdot N$, where 2^{15} is the maximum amplitude of the 16-bit digital audio signal. The normalized amplitude time-series is labelled as $P_{band}(t)$. Figure 8 shows a representative raw audio signal obtained in the milling

Table 2 Operator reaction time statistics from synthetically generated chatter sound experiments

	Operator A r (s)	Operator B r (s)	Operator C r (s)	Average
Mean	0.306	0.323	0.395	0.341
Standard deviation	0.061	0.079	0.069	0.070

Fig. 6 LSP algorithm flowchart



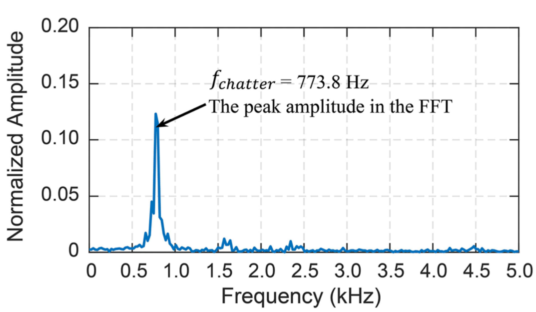


Fig. 7 Identifying chatter frequency at time instance t_s. Dry slot end milling (four flute solid carbide end mill, 100% radial immersion) of 4140 Steel, spindle speed $N = 1000$ RPM, axial depth of cut $a_p = 2.5$ mm, and feed rate $= 132$ mm/min. The machine tool, workpiece and cutting tool specifications are provided in the "Experimental verification" section

experiment of Fig. 7, the corresponding bandpass-filtered signal, and their respective FFTs.

Determining the chatter detection threshold via perception mapping

The next task is to determine the chatter detection threshold P_{th} based on the operator signal instance t_s and $P_{band}(t)$. As discussed in "Evaluating the correspondence problem in chatter detection" section, because humans are subject to natural delay in reacting to an observed signal,t_s corresponds to the time instance when chatter has already developed significantly. Therefore, setting $P_{th} = P_{band}(t_s)$ will result in late detection of chatter.

Figure 9a shows a graphical representation of the temporal correspondence problem introduced by the operator reaction time r. A solution is to evaluate $\widehat{t_c}$, a naïve estimate of the true t_c, by subtracting the sample mean reaction time $\overline{r}$ (obtained from the synthetic sound experiments described earlier) from t_s, as shown in Fig. 9b. Therefore, the chatter detection threshold P_{th} can be set to the value of $P_{band}(t_c) \cong P_{band}(\widehat{t_c}) =$

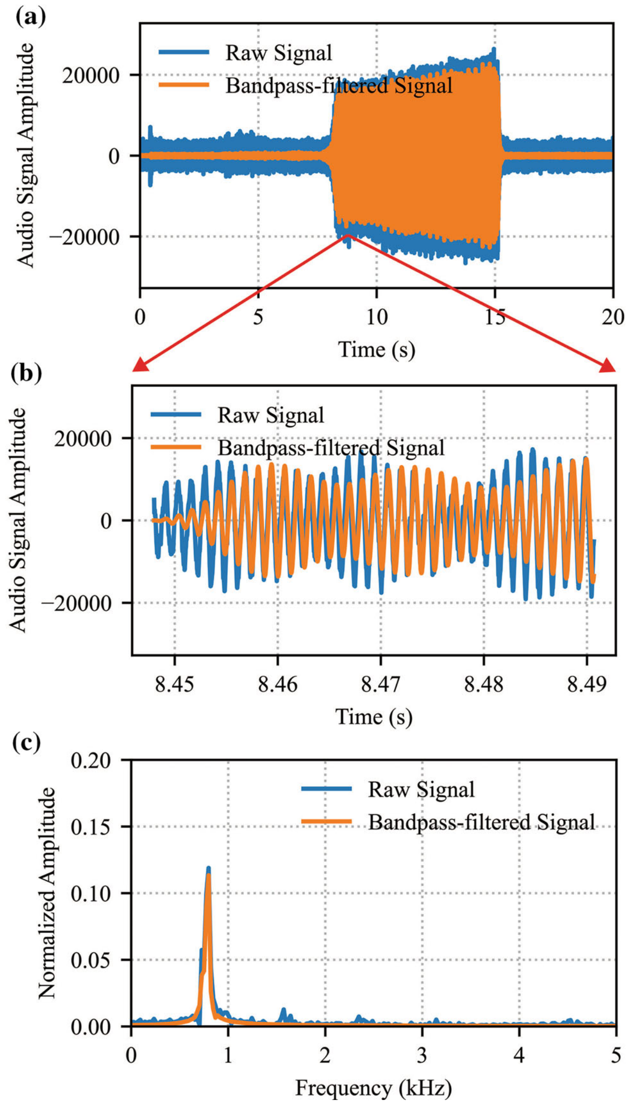


Fig. 8 Signal processing example of the milling experiment in Fig. 7, a raw audio signal and bandpass-filtered signal, b raw and bandpass-filtered signal at t_s, and c FFT of raw and bandpass-filtered signal at t_s

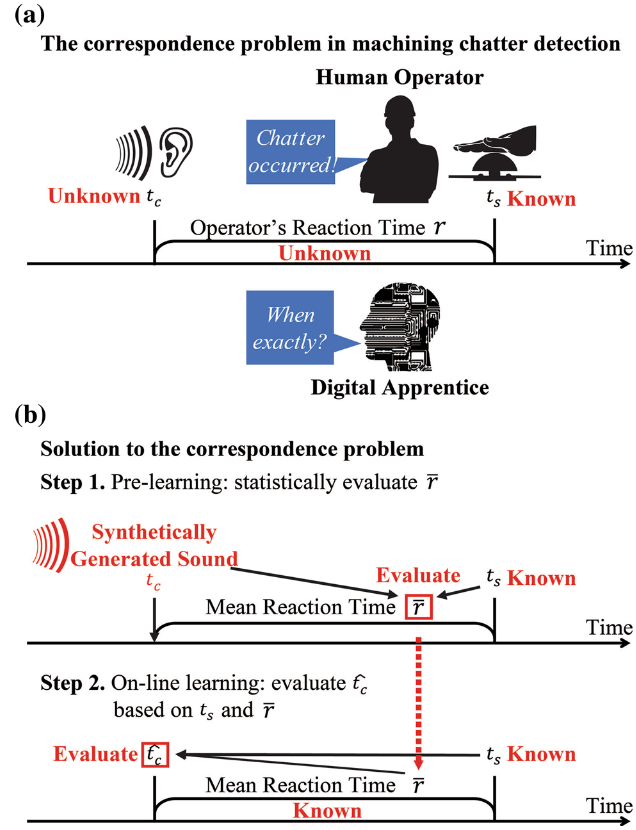


Fig. 9 Graphical representation of the correspondence problem in chatter detection and its solution: **a** two unknown variables r and t_c in the correspondence problem, **b** a naïve estimate to solve the correspondence problem by (step 1) evaluating the mean reaction time $\bar{r}$ to synthetically generated sounds and (step 2) evaluating the estimated chatter mark instance $\widehat{t_c}$

$P_{band}(t_s - \bar{r})$. The simplicity of the naïve estimate is appealing; however, because chatter marks have already appeared on the workpiece at t_c, the chatter detection threshold P_{th} should ideally be adjusted such that the likelihood of detecting chatter before chatter marks appear on the workpiece is higher. Figure 10 illustrates an adjusted time instance θ such that $P_{th} = P_{band}(\theta)$, where $\theta < \widehat{t_c}$. Specifically, instead of using $\widehat{t_c}$, which was calculated from the mean reaction time, the learning agent assumes a uniform probability distribution $f_{trace}(t_c)$ centered around $\widehat{t_c}$. The range of the uniform distribution is specified as $t_s - \bar{r} - 6 \cdot s_r \leq t_c \leq t_s - \bar{r} + 6 \cdot s_r$, where s_r is the standard deviation of the reaction time r. The uniform distribution can be expressed as:

$$Pr(t_s - \bar{r} - 6 \cdot s_r \leq t_c \leq t_s - \bar{r} + 6 \cdot s_r) = \int_{t_s-\bar{r}-6\cdot s_r}^{t_s-\bar{r}+6\cdot s_r} f_{trace}(t_c)dt_c \cong 1 \tag{2}$$

The rationale for the range of uniform distribution in Eq. (2) is as follows. Time instance $t_s - \bar{r} + 6 \cdot s_r$ is approximately equivalent to the operator signal time instance, which occurs too late and therefore corresponds to a cumulative probability of chatter of 1. On the other hand, time instance $t_s - \bar{r} - 6 \cdot s_r$ is too early such that the cumulative probability of chatter before this time instance is approximately 0. Therefore, for time instance θ that lies in the range of the uniform distribution, the cumulative probability of t_c can be rewritten as:

$$Pr(t_c < \theta) = \int_{t_s-\bar{r}-6\cdot s_r}^{\theta} f_{trace}(t_c)dt_c \tag{3}$$

A chatter detection threshold can now be established by assigning it the value of $P_{band}(t)$ at time instance θ to achieve the desired cumulative probability of early chatter detection:

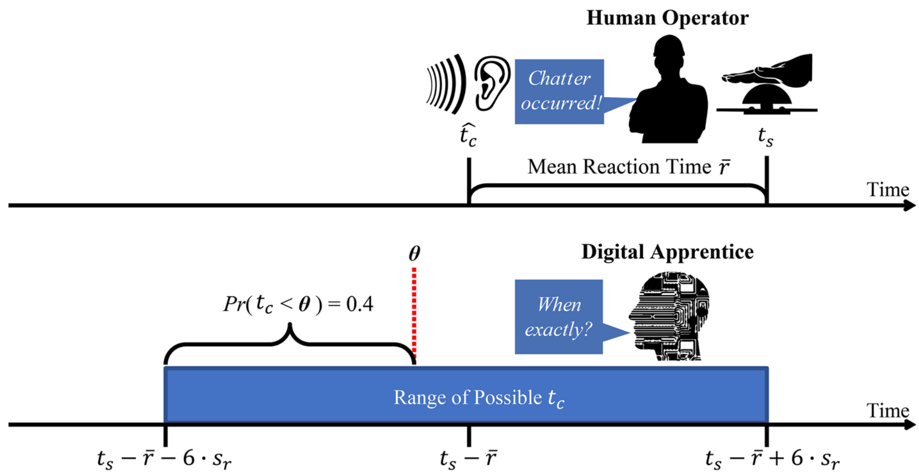


Fig. 10 $P_{band}(\theta)$ as the chatter detection threshold P_{th}

$$Pr(t_c < \theta) = k$$

$$P_{th} = P_{band}(\theta) \tag{4}$$

where k is a tunable parameter for the desired cumulative probability of early chatter detection. A cumulative probability of 0 will result in a chatter detection threshold that detects chatter early but also triggers more false positives while a cumulative probability of 0.5 is equivalent to the unadjusted naïve estimate where $P_{th} = P_{band}\left(\widehat{t_c}\right)$. In this work, since we have deemed that chatter should ideally be detected before chatter marks appear on the machined surface, k was set to 0.4. The mean values of P_{th} were obtained from demonstrations given by each human operator and were used for experimental verification of the LSP algorithm, which is described next.

Table 3 Chatter detection thresholds obtained from training experiments

Operator	N(RPM)	a_p(mm)	P_{th}
A	750	3.5	0.020
A	2000	2	0.038
A	2000	2.5	0.209
Operator A Mean Threshold			0.089
B	750	3.5	0.036
B	2000	3	0.026
Operator B Mean Threshold			0.031
C	1000	2.5	0.013
C	1000	3	0.019
C	1750	2	0.035
C	1750	2.5	0.014
Operator C Mean Threshold			0.020

Experimental verification

Training and validation experiments

The training experiments were designed such that the human operator would encounter both stable and unstable (chatter) cutting conditions. When the operator judged a process to be stable, he/she was instructed to do nothing; conversely, when the operator detected chatter, he/she was instructed to depress the push button switch that recorded the time instance of chatter t_s signaled by the operator. Regardless of the operator's reactions, the training experiments were not interrupted, which enabled post-mortem evaluation of the experimental evidence of chatter marks on the workpiece. Although all three operators were able to accurately discriminate between chatter and stable cuts over the range of milling parameters used in the training experiments, it should be noted that the possibility of incorrect classification by the operator can be accounted for through post-mortem evaluations of the workpiece for visual evidence of chatter marks, which serve as the ground truth. Incorrect classifications by the operator can be discarded and only verified classifications considered as valid operator demonstrations for training the LSP algorithm.

Using the LSP algorithm presented earlier, nine P_{th} values were obtained from the training demonstration experiments and are given in Table 3. The mean thresholds obtained from each operator's demonstrations were subsequently computed. The chatter detection thresholds obtained from demonstration are assumed to work for a representative cutting tool condition. This underlying assumption is valid because the chatter frequencies are a function of the primary modes of vibration based on the dynamics of the system. If the dynamics of the machine tool-cutting tool system is significantly altered (e.g. due to tool wear), the chatter detection thresholds must be relearned and updated using the proposed approach, which only requires a few demonstrations. After

Table 4 Validation experiments

Workpiece material	N(RPM)	a_p(mm)	$P_{band}(t_c)$	Unstable cut max(P_{band})	Stable cut max(P_{band})
Unstable cutting conditions					
4140 Steel	750	3.5	0.038	0.143	N/A
4140 Steel	750	3.5	0.032	0.170	N/A
4140 Steel	1000	2.5	0.024	0.174	N/A
4140 Steel	1000	3	0.033	0.191	N/A
4140 Steel	1000	3	0.051	0.215	N/A
4140 Steel	1000	3.5	0.044	0.282	N/A
4140 Steel	1000	4	0.068	0.279	N/A
4140 Steel	1250	3.5	0.017	0.055	N/A
4140 Steel	1500	3	0.041	0.234	N/A
4140 Steel	1750	2	0.038	0.044	N/A
4140 Steel	1750	2.5	0.015	0.307	N/A
4140 Steel	1750	2.5	0.085	0.456	N/A
4140 Steel	2000	2	0.056	0.092	N/A
4140 Steel	2000	2.5	0.039	0.451	N/A
4140 Steel	2000	3	0.036	0.386	N/A
Stable cutting conditions					
4140 Steel	750	3	N/A	N/A	0.013
4140 Steel	750	3	N/A	N/A	0.021
4140 Steel	1000	2	N/A	N/A	0.004
4140 Steel	1000	2.5	N/A	N/A	0.003
4140 Steel	1250	3	N/A	N/A	0.004
4140 Steel	1500	2.5	N/A	N/A	0.025
4140 Steel	1750	2	N/A	N/A	0.009
4140 Steel	2000	1.5	N/A	N/A	0.003
4140 Steel	2000	2	N/A	N/A	0.016

obtaining the mean chatter detection thresholds from the slot milling experiments, validation slot milling experiments listed in Table 4 were conducted. The milling experiments presented in Tables 3 and 4 were conducted in randomized order.

Performance evaluation

Four outcomes are possible when evaluating the performance of the Operator Mean Thresholds established in the training experiments. These outcomes include early chatter detection, late chatter detection, false positive, and false negative, which are illustrated via an example validation test case shown in Fig. 11. While both early and late chatter detection cases indicate that chatter is accurately detected, it is ideal to detect chatter before chatter marks appear on the workpiece so that chatter can be suppressed before the workpiece is damaged. When Operator Mean Threshold $< P_{band}(t_c)$, chatter was detected before chatter marks appeared on the workpiece, resulting in an early chatter detection classification. When Operator Mean Threshold $> P_{band}(t_c)$, chatter was detected after chatter marks appear on the workpiece surface, which was classified as late chatter detection.

False negative and false positive classifications are both undesirable and should ideally be eliminated. False negative means that the milling process was falsely classified as a stable process, which implies that Operator Mean Threshold $> \max(P_{band})$. On the other hand, a false positive classification occurred when a stable milling process was falsely classified as unstable, which implies that Operator Mean Threshold $< \max(P_{band})$. Each Operator Mean Threshold was first compared with $P_{band}(t_c)$ and the $\max(P_{band})$ obtained in the unstable cuts, and then the Operator Mean Threshold was compared with the $\max(P_{band})$ obtained in the stable cuts. Table 5 summarizes the experimental performance of the Operator Mean Thresholds for each operator and the average performance when applied to the respective validation cases given in Table 4.

It can be seen that the chatter detection thresholds obtained from the LSP algorithm are capable of detecting chatter

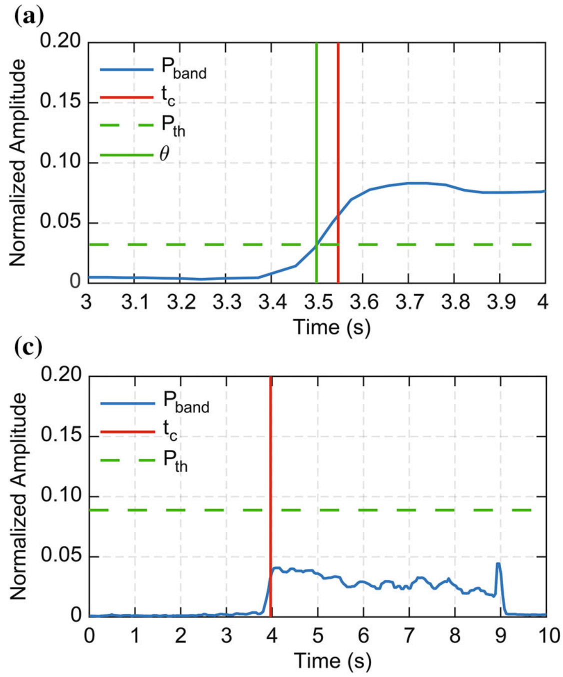


Fig. 11 Examples of classification outcomes when validating the trained chatter detection thresholds: **a** *early detection* with a threshold of 0.031 (Table 3, Operator B Mean Threshold) in a 4140 steel validation test at 2000 RPM, 2 mm axial depth of cut, **b** *late detection* with a threshold of 0.089 (Table 3, Operator A Mean Threshold) in a 4140 steel validation test at 2000 RPM, 2.5 mm axial depth of cut, **c** *false negative* with a threshold of 0.089 (Table 3, Operator A Mean Threshold) in a 4140 steel validation test at 1750 RPM, 2 mm axial depth of cut, **d** *false positive* with a threshold of 0.020 (Table 3, Operator C Mean Threshold) in a 4140 steel validation test at 750 RPM, 3 mm axial depth of cut

Table 5 Experimental performance of the chatter detection thresholds

	Operator A Mean Threshold (%)	Operator B Mean Threshold (%)	Operator C Mean Threshold (%)	Average performance (%)
Unstable cutting conditions				
Early detection	0.0	80.0	86.7	55.6
Late detection	86.7	20.0	13.3	40.0
False negative	13.3	0.0	0.0	4.4
Stable cutting conditions				
True negative	100	100	77.8	92.6
False positive	0.0	0.0	22.2	7.4
Overall				
Accuracy	91.7	100	91.7	94.4

accurately with 94.4% average overall accuracy, 4.4% false negative rate, and 7.4% false positive rate. It is also evident that the chatter detection thresholds result in more early detections (55.6%). It is evident that the worst performing Operator Mean Threshold still resulted in a high accuracy of 91.7% but resulted in a trade-off between a false positive classification and the detection speed. A lower value of the chatter detection threshold typically implies early detection but produces a higher rate of false positives. Taking the average of more demonstrations from each operator can potentially reduce the impact of outliers from poor demonstrations. However, as shown here, even with a limited number of demonstrations (≤ 4), a high chatter detection accuracy can be achieved.

Table 6 Comparison of the LSP algorithm with other machine learning based chatter detection methods reported in the literature

Methodology	Number of training samples	Testing accuracy	Processes and materials
Learning from Demonstration (this paper)	≤ 4	94.4%	Slot Milling of 4140 Steel
ON-LSTM and PBT (Shi et al., 2020)	64	100%	Milling of 2A12 Aluminum
Hybrid Machine Learning and Physics-based model (Rahimi et al., 2021)	355	96.86%	Milling of Aluminum and Steel
Scalogram and Deep Convolutional Neural Network (Tran et al., 2020)	28	100% Unstable Cutting Conditions, 98.01% Stable Cutting Conditions	Slot Milling of Aluminum

To further demonstrate the advantage of the proposed methodology for chatter detection, a comparison of its performance against other machine learning based chatter detection methods reported in the literature is shown in Table 6. The metrics compared are the number of training samples required and the corresponding testing accuracy. It is evident that the proposed method demonstrates similar performance compared to the state of the art albeit with significantly fewer training samples.

The overall chatter detection accuracy of the human operator trained LSP algorithm is promising, especially considering that each chatter detection threshold is obtained from few-shot demonstrations by an experienced operator, which is beneficial in a production setting. One potential limitation of the LSP algorithm is that the learned chatter detection thresholds are only applicable to the chatter frequency demonstrated by the operator. To enable the detection of chatter frequencies not previously demonstrated by the operator during training, one potential solution is to apply notch filters to the audio signals such that the noise and tooth passing frequencies and harmonics are filtered out. In such a case, the learned thresholds can be applied to the given cutting system without additional training demonstrations.

Conclusion

This paper proposed a new approach to milling chatter detection that utilizes an interactive learning agent, termed the Digital Apprentice in this paper, to efficiently learn a chatter detection threshold from few-shot demonstrations provided by an experienced human operator. The paper showed that (1) it is possible to learn to detect chatter from only a few (≤ 4) human operator demonstrations, and (2) the proposed LSP algorithm is able to account for the human operator's reaction time to resolve the temporal correspondence problem, yielding chatter detection thresholds that demonstrate high average chatter detection accuracies (94.4% for milling of 4140 steel). These results indicate that learning from an experienced operator's demonstration to detect milling chatter is effective. The fact that on-line chatter detection methods have not seen widespread industrial acceptance due to the specialized expertise required and the high cost of manually tuning chatter detection thresholds makes the proposed method appealing, especially when experienced operator demonstrations on the production floor are readily available. As a natural extension of the Digital Apprentice proposed herein, an interesting future research direction is to render the Digital Apprentice more practical for industrial applications where preventing chatter is of utmost importance. In such cases, a certain level of false classification must be acceptable, so long as chatter marks are not produced on the workpiece. By coupling chatter suppression techniques with the proposed method for chatter detection, future work can focus on preemptively adjusting the cutting conditions (e.g. spindle speed and/or feed) utilizing thresholds optimized for early detection of chatter in milling operations.

Acknowledgements This work was supported by Siemens Corporation, Corporate Technology through the Siemens Fellowship.

Declarations

Competing interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

See Table 7.

Table 7 Experience level of human subjects used in the training experiments

	Machining experience	Experience with chatter
Operator A	2 years	0 year
Operator B	6 years	6 years
Operator C	3 years overall, 2 years of professional experience	2 years

References

Altintaş, Y., & Budak, E. (1995). Analytical prediction of stability lobes in milling. *CIRP Annals, 44*(1), 357–362.

Altintas, Y., & Chan, P. K. (1992). In-process detection and suppression of chatter in milling. *International Journal of Machine Tools and Manufacture, 32*(3), 329–347.

Argall, B. D., Chernova, S., Veloso, M., & Browning, B. (2009). A survey of robot learning from demonstration. *Robotics and Autonomous Systems, 57*(5), 469–483.

Aslan, D., & Altintas, Y. (2018). On-line chatter detection in milling using drive motor current commands extracted from CNC. *International Journal of Machine Tools and Manufacture, 132*, 64–80.

Bachrathy, D., Sykora, H. T., Hajdu, D., Beri, B., & Stepan, G. (2021). Why is it hard to identify the onset of chatter? A stochastic resonance perspective. *CIRP Annals, 70*, 329.

Berger, B., Minis, I., Rokni, M., Papadopoulos, M., Deng, K., & Chavalli, A. (1997). Cutting state identification. *Journal of Sound and Vibration, 200*(1), 15–29.

Caixu, Y., Haining, G., Xianli, L., Liang, S. Y., & Lihui, W. (2019). A review of chatter vibration research in milling. *Chinese Journal of Aeronautics, 32*(2), 215–242.

Cao, H., Yue, Y., Chen, X., & Zhang, X. (2017). Chatter detection in milling process based on synchrosqueezing transform of sound signals. *The International Journal of Advanced Manufacturing Technology, 89*(9), 2747–2755.

Cen, L., Melkote, S. N., Castle, J., & Appelman, H. (2018). A method for mode coupling chatter detection and suppression in robotic milling. *Journal of Manufacturing Science and Engineering*. https://doi.org/10.1115/1.4040161

Chernova, S., & Thomaz, A. L. (2014). Robot learning from human teachers. *Synthesis Lectures on Artificial Intelligence and Machine Learning, 8*(3), 1–121.

Delio, T. (1989). *A sensor-based adaptive control constraint system for automatic spindle speed regulation to obtain highly stable milling.* (Doctor of Philosophy Dissertation), University of Florida.

Delio, T., Tlusty, J., & Smith, S. (1992). Use of audio signals for chatter detection and control. *Journal of manufacturing science and engineering, 114*(2).

Ding, L., Sun, Y., & Xiong, Z. (2017). *Early chatter detection based on logistic regression with time and frequency domain features.* Paper presented at the 2017 IEEE International Conference on Advanced Intelligent Mechatronics (AIM).

Duncan, G., Schmitz, T. L., & Kurdi, M. H. (2000). *Uncertainty propagation for selected analytical milling stability limit analyses*: Society of Manufacturing Engineers.

Faassen, R., Doppenberg, E., van de Wouw, N., Oosterling, J., & Nijmeijer, H. (2006). *Online detection of the onset and occurrence of machine tool chatter in the milling process.* Paper presented at the CIRP 2nd International Conference on High Performance Cutting.

Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing, 20*(7), 1483–1510.

Khorasani, A. M., Littlefair, G., & Goldberg, M. (2014). Time domain vibration signal processing on milling process for chatter detection. *Journal of Machining and Forming Technologies, 6*(1/2), 45.

Knox, W. B., & Stone, P. (2008). *Tamer: Training an agent manually via evaluative reinforcement.* Paper presented at the 2008 7th IEEE International Conference on Development and Learning.

Knox, W. B., & Stone, P. (2009). *Interactively shaping agents via human reinforcement: The TAMER framework.* Paper presented at the Proceedings of the Fifth International Conference on Knowledge Capture.

Kuljanic, E., Sortino, M., & Totis, G. (2008). Multisensor approaches for chatter detection in milling. *Journal of Sound and Vibration, 312*(4–5), 672–693.

Lamraoui, M., El Badaoui, M., & Guillet, F. (2015). Chatter detection in CNC milling processes based on Wiener-SVM approach and using only motor current signals. In *Vibration engineering and technology of machinery* (pp. 567–578). Springer.

Liao, Y., & Young, Y. (1996). A new on-line spindle speed regulation strategy for chatter control. *International Journal of Machine Tools and Manufacture, 36*(5), 651–660.

Liu, C., Zhu, L., & Ni, C. (2017). The chatter identification in end milling based on combining EMD and WPD. *The International Journal of Advanced Manufacturing Technology, 91*(9), 3339–3348.

Liu, H., Chen, Q., Li, B., Mao, X., Mao, K., & Peng, F. (2011). On-line chatter detection using servo motor current signal in turning. *Science China Technological Sciences, 54*(12), 3119–3129.

Luo, M., Luo, H., Axinte, D., Liu, D., Mei, J., & Liao, Z. (2018). A wireless instrumented milling cutter system with embedded PVDF sensors. *Mechanical Systems and Signal Processing, 110*, 556–568.

Ma, L., Melkote, S. N., & Castle, J. B. (2013). A model-based computationally efficient method for on-line detection of chatter in milling. *Journal of Manufacturing Science and Engineering, 135*(3), 031007.

Ma, L., Melkote, S. N., & Castle, J. B. (2014). PVDF sensor-based monitoring of milling torque. *The International Journal of Advanced Manufacturing Technology, 70*(9–12), 1603–1614.

Melkote, S. (2020). Georgia Institute of Technology Institutional Review Board Protocol H20340. In *Study of Human's Reaction Time to Anomalies in Manufacturing Operations*. Georgia Institute of Technology: Georgia Institute of Technology.

Nehaniv, C. L., & Dautenhahn, K. (2002). The correspondence problem. In *Imitation in animals and artifacts* (p. 41).

Nguyen, V., Melkote, S., Deshamudre, A., Khanna, M., & Walker, D. (2016). *Comparison of on-line chatter detection algorithms in turning.* Paper presented at the 2016 International Symposium on Flexible Automation (ISFA).

Palmer, E. M., Horowitz, T. S., Torralba, A., & Wolfe, J. M. (2011). What are the shapes of response time distributions in visual search? *Journal of Experimental Psychology: Human Perception and Performance, 37*(1), 58.

Pham, H. (2006, 2017). PyAudio 0.2.11. https://pypi.org/project/PyAudio/

Quintana, G., & Ciurana, J. (2011). Chatter in machining processes: A review. *International Journal of Machine Tools and Manufacture, 51*(5), 363–376.

Rahimi, M. H., Huynh, H. N., & Altintas, Y. (2021). On-line chatter detection in milling with hybrid machine learning and physics-based model. *CIRP Journal of Manufacturing Science and Technology, 35*, 25–40.

Schaal, S. (1997). *Learning from demonstration.* Paper presented at the Advances in Neural Information Processing Systems.

Schaal, S. (2006). Dynamic movement primitives-a framework for motor control in humans and humanoid robotics. In *Adaptive motion of animals and machines* (pp. 261–280). Springer.

Schaal, S., Peters, J., Nakanishi, J., & Ijspeert, A. (2003). *Control, planning, learning, and imitation with dynamic movement primitives.* Paper presented at the Workshop on Bilateral Paradigms on Humans and Humanoids: IEEE International Conference on Intelligent Robots and Systems (IROS 2003).

Schmitz, T. L., Medicus, K., & Dutterer, B. (2002). Exploring once-per-revolution audio signal variance as a chatter indicator. *Machining Science and Technology, 6*(2), 215–233.

Shi, F., Cao, H., Wang, Y., Feng, B., & Ding, Y. (2020). Chatter detection in high-speed milling processes based on ON-LSTM and PBT.

The International Journal of Advanced Manufacturing Technology, 111(11), 3361–3378.
Smith, S., & Tlusty, J. (1991). An overview of modeling and simulation of the milling process. *Journal of Manufacturing Science and Engineering, 113*(2), 169.
Smith, S., & Tlusty, J. (1993). Efficient simulation programs for chatter in milling. *CIRP Annals, 42*(1), 463–466.
Tlusty, J., & Polacek, M. (1968). *Experience with analysing stability of machine tool against chatter.* Paper presented at the 9th MTDR Conference.
Tobias, S. (1964). Vibration of machine tools. *Production Engineer, 43*(12), 599–608.
Tran, M.-Q., Liu, M.-K., & Tran, Q.-V. (2020). Milling chatter detection using scalogram and deep convolutional neural network. *The International Journal of Advanced Manufacturing Technology, 107*(3), 1505–1516.
Warnell, G., Waytowich, N., Lawhern, V., & Stone, P. (2018). *Deep tamer: Interactive agent shaping in high-dimensional state spaces*. Paper presented at the Proceedings of the AAAI Conference on Artificial Intelligence.
Wright, P. K., & Bourne, D. A. (1988). *Manufacturing intelligence*. Addison-Wesley Longman Publishing Co., Inc.
Yao, Z., Mei, D., & Chen, Z. (2010). On-line chatter detection and identification based on wavelet and support vector machine. *Journal of Materials Processing Technology, 210*(5), 713–719.
Yesilli, M. C., Khasawneh, F. A., & Otto, A. (2020). On transfer learning for chatter detection in turning using wavelet packet transform and ensemble empirical mode decomposition. *CIRP Journal of Manufacturing Science and Technology, 28*, 118–135.
Yuqing, Z., Xinfang, L., Fengping, L., Bingtao, S., & Wei, X. (2015). An online damage identification approach for numerical control machine tools based on data fusion using vibration signals. *Journal of Vibration and Control, 21*(15), 2925–2936.
Zhang, C. L., Yue, X., Jiang, Y. T., & Zheng, W. (2010). *A hybrid approach of ANN and HMM for cutting chatter monitoring*. Paper presented at the Advanced Materials Research.