DIGITAL APPRENTICE FOR CHATTER DETECTION:

AN ON-LINE LEARNING APPROACH TO REGENERATIVE CHATTER

DETECTION IN MACHINING VIA HUMAN-MACHINE INTERACTION

A Thesis
Presented to
The Academic Faculty

by

Xiaoliang Yan

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Mechanical Engineering

Georgia Institute of Technology
December 2020

DIGITAL APPRENTICE FOR CHATTER DETECTION:

AN ON-LINE LEARNING APPROACH TO REGENERATIVE CHATTER

DETECTION IN MACHINING VIA HUMAN-MACHINE INTERACTION

Approved by:

Dr. Shreyes Melkote, Advisor
School of Mechanical Engineering
Georgia Institute of Technology

Dr. Thomas Kurfess
School of Mechanical Engineering
Georgia Institute of Technology

Dr. Matthew Gombolay
School of Interactive Computing
Georgia Institute of Technology

Date Approved: December 2, 2020

ACKNOWLEDGEMENT

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

A = A set of possible actions in an LfD problem

a = Sequence of discrete time instances bounded by t_s and $t_s - \bar{r} - 6 \cdot s_{\bar{r}}$ for the discrete estimation of chatter observation instance t_c

d = Difference between a human signal instance t_s and a chatter mark instance t_a (s)

$\bar{d}$ = Mean difference between human signal instances t_s and chatter mark instances t_a (s)

$f_{chatter}$ = Chatter frequency (Hz)

f_s = Sampling frequency (Hz)

f_{tooth} = Tooth passing frequency (Hz)

f_{react} = Probability density function of reaction time

f_{trace} = Probability density function of the chatter observation instance

k = Probability coefficient for chatter detection threshold

k_x = Modal stiffness matrix in x-axis, element unit (N/m)

k_y = Modal stiffness matrix in y-axis, element unit (N/m)

K_s = Tangential cutting force coefficient (N/m^2)

m_x = Modal mass in x-axis (kg)

m_y = Modal mass in y-axis (kg)

M = A mapping between the states to the observation of the states in an LfD problem

N = Bin size of Fast Fourier Transform

$P_{chatter}$ = Frequency spectrum ratio of band-pass filtered audio signals

$P_{max,stable}$ = Maximum frequency spectrum ratio of a stable cutting process

$P_{max,chatter}$ = Maximum frequency spectrum ratio of a chatter cutting process

P_{notch} = Frequency spectrum ratio of notch filtered audio signals

P_R = Power ratio of the top two largest Fast Fourier Transform peaks

P_{th} = Chatter detection threshold

r = Reaction time between chatter observation instance and human signal instance (s)

$\bar{r}$ = Sample mean reaction time of a human subject (s)

S = States in an LfD problem

$s_{\bar{r}}$ = Standard deviation of the sample mean reaction time of a human subject (s)

T = A Markovian transition function in an LfD Problem

t_a = Chatter mark instance (s)

t_c = Chatter observation instance (s)

$\widehat{t_c}$ = Estimation of the chatter observation instance (s)

t_s = Human signal instance (s)

t_{st} = Time instance between stable period and transient period (s)

t_{tc} = Time instance between transient period and chatter period (s)

X_F = Discrete Fourier Transform of a signal

Z = Observation of the states in an LfD problem

θ = Chatter detection threshold time instance (s)

μ_{stable} = Mean frequency spectrum ratio from the stable cutting region

ξ_x = Modal damping ratio in x-axis

ξ_y = Modal damping ratio in y-axis

π = A policy that selects the actions based on the states in an LfD problem

σ_{stable} = Standard deviation frequency spectrum ratio from the stable cutting region

SUMMARY

Regenerative chatter in machining, which is characterized by self-excited vibration, is a common process anomaly that limits productivity and part quality in machining operations. This thesis proposes an on-line approach for chatter detection via effective human-machine interaction, facilitating knowledge transfer from experienced machinists to the "digital apprentice" through the "learnable skill primitive" (LSP) method that establishes a chatter detection threshold. The research focus is to develop the methodology for chatter-specific knowledge acquisition and a human-machine interface inspired by computing techniques and frameworks such as learning from demonstration, reinforcement learning, and interactive agent shaping. In this work, the milling operation is selected as a case study for the proposed LSP method. Digital audio data is acquired from milling experiments through a studio-style condenser microphone mounted inside a milling machine. The data is pre-processed through various digital filters before Fast Fourier Transform (FFT) is performed to identify the chatter frequency contents. During the training phase, data for the human operator's natural reaction to chatter is collected via a specially designed human-machine interface. The learned chatter detection thresholds are obtained through the "learnable skill primitive" method by temporally mapping the reaction data to the cutting signal. In addition, a variance mitigation strategy is developed to reduce the negative impact of the high variance in the operator's reaction time to chatter. During the testing phase, experiments are conducted to evaluate the detection accuracy, detection speed, and robustness of the learned chatter detection thresholds. Experimental

data support the claim that the learned thresholds can detect chatter with good detection accuracy and detection speed. Finally, the learned threshold is demonstrated to be robust to milling of different workpiece materials under different cutting conditions such as feeds, speeds, axial and radial immersions (depths of cut), and directions of cutting.

CHAPTER 1 INTRODUCTION

1.1 Motivation and Problem Statement

Chatter is a common process anomaly that occurs during machining operations. It is a dynamic instability of the machine tool-cutting tool-workpiece system characterized by unstable vibrations. Poor surface finish, excessive noise, and potential permanent machine tool damage are some of the negative impacts of chatter [1]. Chatter can incur significant additional manufacturing cost. For example, in the automotive industry the average estimated unit cost due to chatter during production of engine cylinder blocks is approximately 0.4 US Dollar; considering the annual production volume of 3 million engines of the same model, the total cost incurred can be very high. Consequently, industry has been in search of a systematic and robust solution to control chatter [2].

The most common type of machining chatter is regenerative chatter, sometimes simply referred to as chatter. This type of chatter is characterized by self-excited vibrations, which are induced by the phase difference between the instantaneous vibratory motion of the cutting edge and the previously generated wave on the workpiece surface [3]. The cutting tool engages the wavy surface left by the previous tooth pass and the phase shift amplifies the instantaneous variation in chip load, which causes the cutting force to vary dynamically and results in large amplitude vibration [4]. The topic of chatter has been studied extensively with the goal of developing a solution for chatter avoidance or suppression. Two major strategies have been introduced in the literature: (1) off-line stability identification of the cutting system, and (2) on-line chatter detection and suppression [1].
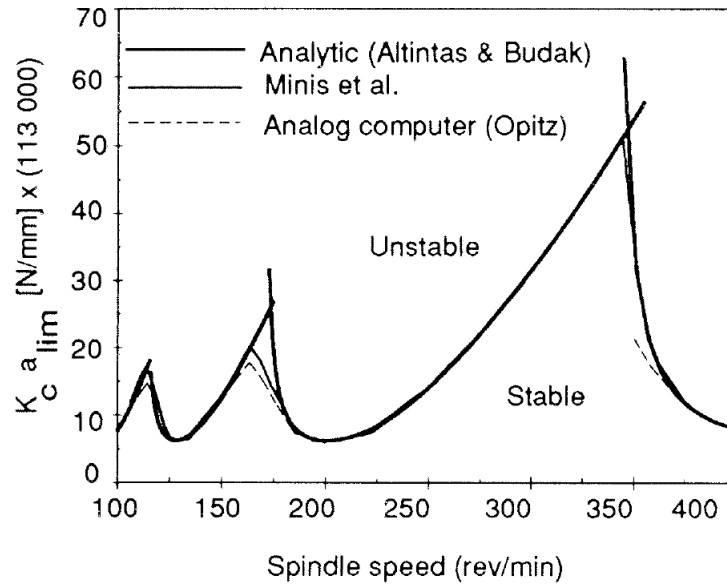
Figure 1. Stability lobe diagram of a single degree of freedom system [5].

The off-line stability identification method seeks to predict the stability of the tool-workpiece-machine tool system (i.e. cutting system) to select ideal process parameters before the machining operation [5]. The most common method for stability identification is experimental modal analysis, which utilizes a modal impact hammer to excite the tip of the cutting tool, where an accelerometer is placed, to obtain the frequency response function (FRF) of the cutting system [6]. By identifying the primary modes of vibration and the cutting coefficient of the workpiece-tool combination, a "stability lobe diagram" (SLD) with spindle speed on the x-axis and axial depth of cut (DOC) on the y-axis (Figure 1) can be obtained for process planning. While the off-line SLD identification method offers an approach for selection of the cutting parameters, it suffers from high uncertainty and a high cost of implementation [1]. The high uncertainty is a result of inaccuracies in the FRF, use of a static cutting coefficient, and inherent inaccuracies in the physical model

2

of the cutting system. While researchers have identified the key sources of uncertainty and created a confidence interval on the SLD [7], the uncertainty in the off-line SLD identification approach can lead to ultra-conservative selection of the cutting parameters, which inhibits productivity. The difficulty in implementation of the SLD approach can also drive manufacturers away from the off-line approach. The processes to obtain the FRF and the cutting coefficients are sophisticated and require well-controlled experiments. As a result, only major manufacturers are incentivized to train staff for SLD identification.

On the other hand, on-line chatter detection and suppression involves monitoring the machining process through sensors. The advantage of on-line detection methods is that the dynamics of the cutting system do not need identification, and the implementation cost is relatively low. However, critics of the on-line detection methods argue that chatter detection algorithms often suffer from their inability to detect chatter before damage is done to the workpiece [1]. Although previous works have proposed algorithms that can detect the onset of fully developed chatter [8, 9], the chatter detection threshold is an engineering parameter that must be tuned by engineers to obtain the desired detection accuracy. Because the cutting signals acquired from the machining operation vary with cutting tools, sensor types, sensor locations, and machining conditions, on-site tuning of the chatter detection threshold is often necessary for the implementation of on-line detection methods. However, tuning of the chatter detection thresholds can be challenging because research engineers who have the experience and skills for tuning cannot always be on the production floor, and production staff typically do not have the necessary time or skills for tuning [10]. Thus, online-chatter detection methods have not found wide acceptance in production.

Ultimately, both strategies for chatter avoidance, detection, or suppression suffer from being case-specific or inflexible in a production setting. It takes too much effort and cost to either re-identify the system dynamics of the off-line SLD method for any changes in the machine tool-tool holder-cutting tool-workpiece combination or re-tune a chatter detection threshold of the on-line detection method to account for variations in tooling and sensor types, sensor location, and machining conditions. As a result, manufacturers remain heavily reliant on the experience of human machine tool operators to monitor and control the machining process. However, the reliance on human operators also has significant disadvantages. Even with years of experience, the human operator's lack of speed and accuracy are limiting factors. In addition, due to shortage of skilled labor, corporations are facing increasing difficulty in replacing the expertise of experienced operators when they leave or retire. Therefore, the manufacturing industry needs a robust, fast, and accurate solution to monitor and control chatter.

1.2   Research Objectives

Based on the above discussion, a practical, effective, robust and easy to implement chatter detection methodology is warranted. This thesis proposes to leverage both chatter detection algorithms and the experience of human operators for effective chatter detection in a machining operation. Specifically, this research aims to answer two major research questions: 1) Is it possible for the machine to learn to detect chatter from a human operator's demonstration? And 2) to what extent can the machine learn to detect chatter early and accurately?

The objective is to create a "Digital Apprentice" for chatter detection that learns from an experienced operator's demonstration through human-machine interaction to accurately detect chatter in a timely manner. While the use of human-machine interaction techniques is novel in the field of manufacturing process monitoring, learning from demonstration and interactive agent shaping are techniques that have been employed for effective interactive computing in the robotics field [11].

Learning from demonstration (LfD) acknowledges that learning from scratch without any prior knowledge is challenging and impractical. As humans, we typically approach learning a task based on instructions or demonstrations from other more experienced humans. The objective of LfD is to allow the machine to learn a particular policy π or a mapping between states S and actions A from one or a few human demonstrations [12]. One way to construct an LfD problem is to define a set of probabilistic transition functions $T(s'|s, a): S \times A \times S \rightarrow [0, 1]$, which form a Markovian chain [13]. If the states are assumed unobservable, the hidden states S are approximated through the mapping $M: S \rightarrow Z$, where Z is the observable state and the policy $\pi: Z \rightarrow A$ selects the actions based on the observations of the world states [13]. This process is characterized as a Hidden Markov Model (HMM). Figure 2 shows that a set of human demonstrations D records the demonstrated states S and actions A, from which a policy π can be derived by evaluating the probabilistic transition functions of the Markovian chain.
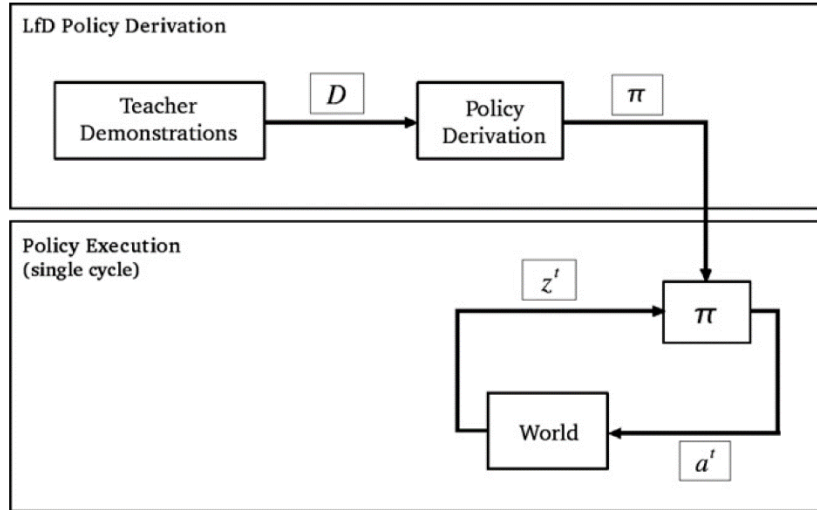
Figure 2. LfD policy derivation [13].

In the context of machining, the human operator is responsible for recognizing chatter and performing the necessary corrective actions, which include adjusting the cutting conditions in real time or halting the process to prevent further damage to the part or machine. The operator perceives the cutting signals using his/her natural senses, and the "Digital Apprentice", which learns from the operator, monitors the cutting signal using sensors. In this work, the policy π learned by the "Digital Apprentice" is a binary chatter classifier that classifies the sensed cutting signals into either chatter or stable cutting; the frequencies and the frequency spectrum ratios computed from the cutting signal correspond to the observed states Z, and the actions A are the classification of the machining process as either chatter or stable cutting. Future work will expand the set of actions to corrective control policy that selects corrective actions to suppress chatter. However, the prerequisite for this is to learn a chatter detection classifier (a threshold) from human demonstration,

6

which can be challenging due to the correspondence problem between the human's perception abilities and the data acquired from artificial sensors [14].

Because chatter is a process anomaly that can quickly develop in a short period of time, and because a human's response to chatter is delayed due to his/her reaction time, by the time a human provides a demonstration that indicates the occurrence of chatter, the machining process state S has already changed significantly. Effective learning from a human demonstration requires knowledge of the state at the time instance when the human recognizes chatter by accounting for his/her delayed reaction. A similar challenge in the correspondence problem is discussed by Knox et al. [15] in their "Training an Agent Manually via Evaluative Reinforcement" (TAMER) framework. TAMER is a framework that addresses the shaping problem in psychology and applies it to computing. It focuses on allowing a human trainer to interactively shape an agent's policy through reinforcement rewards [16]. The human's reward to or critique of the agent's current policy is mapped to a corresponding state-action pair of the learning agent's policy. When the human's critique is applied to quickly varying states and actions, it is demonstrated that the correspondence problem can be resolved by systematically accounting for the delay in the human's reaction time.

Ultimately, the research objective of this thesis is to create a "Digital Apprentice" for machining chatter detection that learns from an expert machine operator's demonstration of chatter detection. The methodology proposed in this thesis can be considered a solution to a simplified learning from demonstration problem where the policy is a binary classifier of machining stability or chatter, the states S are the cutting process audio signal's frequencies and frequency spectrum ratios, and the actions A are the

classification of the cutting signal as stable or chatter. In addition, a solution to the correspondence problem in chatter detection is proposed based on the approach introduced by the TAMER framework.

1.3   Proposed Approach

The overall research approach is summarized in Figure 3. By facilitating effective human-machine interaction during the machining process, a "Learnable Skill Primitive" (LSP) algorithm for chatter detection is developed to take in a human operator's demonstration and rewards for chatter detection, and output an effective and robust chatter detection threshold that is applicable over a range of machining conditions. The value proposition is that effective human-machine interaction reduces the time and cost of learning a good chatter detection threshold, which reduces the implementation cost and eliminates the need for manual tuning of the detection threshold, thereby drastically improving the robustness of chatter detection algorithms, particularly in production settings.

```
┌─────────────────────────────────────┐
│   Human-Machine Interface for Human-  │
│  Machine Interaction in Chatter Detection │
└─────────────────────────────────────┘
                    ▼
┌─────────────────────────────────────┐
│   Evaluation of Human Operators' Reaction │
│          Time to Chatter in Milling       │
└─────────────────────────────────────┘
                    ▼
┌─────────────────────────────────────┐
│    Learnable Skill Primitive for Chatter  │
│                Detection                  │
└─────────────────────────────────────┘
                    ▼
┌─────────────────────────────────────┐
│  Variance Mitigation Strategy of Learnable │
│    Skill Primitive for Chatter Detection  │
└─────────────────────────────────────┘
                    ▼
┌─────────────────────────────────────┐
│    Experimental Validation of Detection   │
│       Accuracy, Speed and Robustness      │
└─────────────────────────────────────┘
```
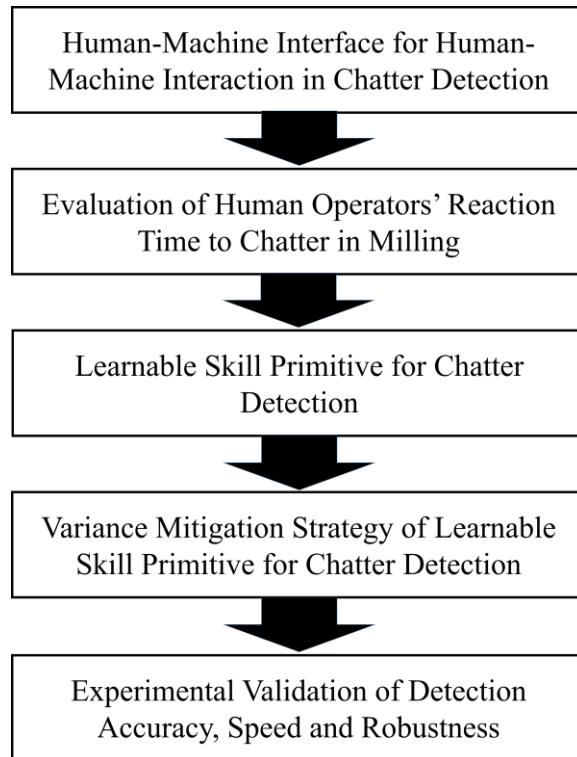
Figure 3. Overall research approach.

1.4   Thesis Outline

Chapter 2 provides an introduction to regenerative chatter in machining and computing techniques that have inspired the proposed methodology. A description of the human-machine interface and the perception mapping problem is provided in Chapter 3, where an experimental method for collecting data for human reaction time to chatter is presented. Chapter 4 focuses on the experimental setup, data acquisition method, and the "Learnable Skill Primitive" algorithm for on-line knowledge transfer from a human operator to the machine. The Variance Mitigation Strategy is introduced, and the experimental performance of the learned chatter detection thresholds is analyzed in Chapter 5. Finally, the thesis concludes with evaluation of the effectiveness of the "Digital Apprentice," and provides recommendations for future work using the LSP approach.

CHAPTER 2 LITERATURE REVIEW

In this chapter, a detailed literature review of chatter in machining and interactive computing frameworks is presented. The review consists of six sections: 1) prior research on regenerative chatter in machining, 2) milling process stability analysis, 3) on-line chatter detection sensors and algorithms, 4) learning from demonstration, 5) reinforcement learning, and 6) interactive agent shaping.

2.1   Regenerative Chatter in Machining Process

Chatter is a common process anomaly that can occur in a machining operation. The most common cause of chatter is the regenerative effect induced by self-excited vibrations [6]. The regenerative effect is induced by the phase shift between the instantaneous vibratory motion of the cutting edge and the previously generated wavy surface on the workpiece [3], which introduces dynamic variations in chip load and large amplitude vibrations. The dynamic instability due to regenerative chatter can result in catastrophic consequences such as poor surface finish, excessive tool wear, loud noise, and machine tool damage [1]. Figure 4 is an example of chatter marks on a 7075-T651 Aluminum workpiece produced by an unstable milling operation.
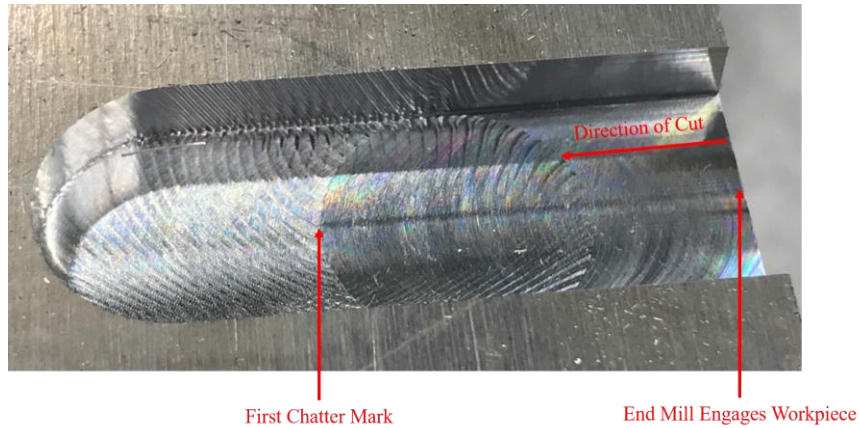
Figure 4. Chatter marks on a 7075-T651 Aluminum workpiece in a slot milling operation.

Previous works have proposed different methods to combat chatter instability in machining processes [1, 5, 8, 9, 17]. The first method proposed was the off-line method for stability identification, which focuses on identifying the stability limit between chatter and stable cutting through analytical or numerical modelling methods [3-5, 18]. On-line chatter detection methods, on the other hand, require no previous knowledge of the cutting system's frequency response function, and instead, utilize various sensors and algorithms for timely chatter detection [8, 9, 17, 19]. Other methods such as passive chatter avoidance through non-standard cutting tools (e.g. variable pitch or variable helix end mills) or external system damping [20, 21], and active chatter suppression through spindle speed variation (SSV) have also been suggested to avoid or suppress chatter [22]; however, some critics argue that these methods offer case-specific solutions, which may not be generally applicable in a flexible manufacturing system, where changes in products and required processes can be frequent [1]. In the following two sections, an in-depth review of off-line stability prediction models and on-line detection methods is presented.

2.2 Off-line Machining Stability Prediction Models

The premise of this line of research is that effective chatter avoidance can be achieved through understanding of the system dynamics of the machine tool-tool holder-cutting tool-workpiece system [6], which permits the selection of cutting parameter values that do not result in chatter. Historically, mathematical analysis of chatter stability in machining operations such as milling was first presented by Tobias and Fishwick [23], and Tlusty and Polacek [24], who modeled regenerative chatter vibration using delayed differential equations. Minis et al. [25] presented a general mathematical model that described the milling dynamics to predict the limiting axial depth of cut for stable milling. Altintas and Budak [5] proposed a zeroth order approximation of the cutting system stability for the milling process, where a zeroth order Fourier term is used as an approximation of the cutting force variation. Other researchers have extended analytical chatter stability models to three dimensions and multiple degrees of freedom for different machining operations [26, 27]. The stability of a milling process can be modelled, provided that the transfer function of the structure at the cutter-workpiece contact, the static cutting coefficients, radial immersion, and the number of cutter teeth are known [6]. Figure 5 shows a picture of an impact hammer experiment used to obtain the frequency response function (FRF) of the cutting system. Based on the cutting coefficient and the FRF, the chatter stability limit can be computed and graphed as in Figure 6, where the spindle speed is plotted on the x-axis and the axial depth of cut on the y-axis. Due to its lobed shape, the figure is commonly referred to as the stability lobe diagram (SLD).

Figure 5. Impact hammer test.
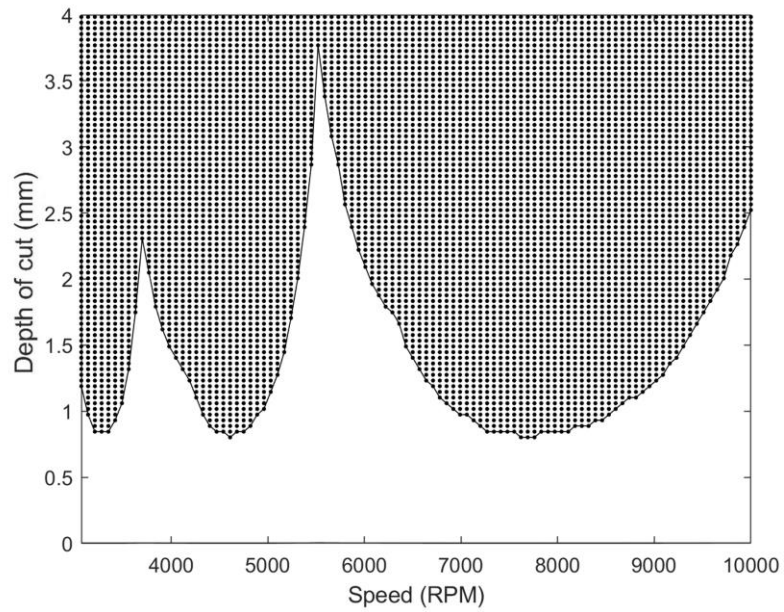
Figure 6. Two degrees of freedom stability lobe diagram generated from time-domain simulation. Cutting coefficient $K_s = 800\ MPa$, modal mass matrix $m_x = m_y = \begin{bmatrix} 4.245 & 0 \\ 0 & 2.711 \end{bmatrix} kg$, modal stiffness matrix $k_x = k_y = 10^7 \cdot \begin{bmatrix} 9.805 & 0 \\ 0 & 4.878 \end{bmatrix} N/m$, modal damping ratio matrix $\xi_x = \xi_y = \begin{bmatrix} 0.0182 & 0 \\ 0 & 0.0429 \end{bmatrix}$.

While the SLD of the system is informative for process planning, the analytical stability model lacks consideration of the machine tool's maximum allowable cutting force

13

[18]. Time domain simulations of the machining process have been used to evaluate the peak-to-peak force and the vibration amplitude in a machining operation. Jemielniak and Widota [28] developed an algorithm for numerical simulation of non-linear chatter vibration in turning. Smith and Tlusty [4, 18, 29] proposed an efficient time domain simulation approach for chatter in milling using a regenerative force dynamic deflection model where the instantaneous tangential force is a function of the cutting coefficient, axial depth of cut, nominal chip load, radial immersion, and dynamic position of the cutting teeth. Li et al. [30] presented a novel chatter stability criterion for time domain simulation of the milling process. The time domain simulation approach provides not only force and displacement information for a given combination of cutting parameters, but, by simulating the process over a range of cutting parameters, the peak-to-peak force diagram with respect to the spindle speed can also be obtained [18]. The detailed prediction of the milling forces and displacements for a specified combination of cutting parameters can lead to better process planning.

Critics of the off-line chatter stability limit identification approach argue that the depth of knowledge required and the amount of training needed to obtain a complete analysis of the machine tool dynamics makes it difficult for industrial users to implement [1]. In addition, the dynamics of the system may suffer from large uncertainty arising from measurements and model parameters. Duncan et al. [7] described the propagation of uncertainty in an analytical milling stability model and proposed a procedure for adding uncertainty bounds to the milling stability limits. Karandikar et al. [31] applied the random walk approach for Bayesian inference of stability in a milling operation. Various other works have applied the Bayesian learning approach to estimate the stability limit and select

optimal cutting parameters [32, 33]. However, the large uncertainty bounds of the stability limit can lead to more conservative cutting parameters in process planning, which can significantly limit productivity. The next section reviews on-line chatter detection methods for timely recognition of chatter in-process.

2.3   On-line Chatter Detection Methods: Sensors and Algorithms

While the off-line SLD approach to chatter stability limit prediction provides great insight into the system dynamics and stability of a machining process, its inflexibility in adapting to changes in the cutting system make it challenging for an average machine shop to employ the off-line SLD approach. Therefore, on-line chatter recognition techniques have been proposed for timely detection and control of machining chatter.

There are two major objectives of on-line chatter detection research, namely, identifying and collecting information-rich cutting signals and developing associated chatter detection algorithms. Vibration, strain, sound, cutting force and power signals have been extensively researched as potential measurands for effective chatter detection [8, 17, 34-38]. Altintas and Chan [39] presented an in-process chatter detection and suppression method for milling, where chatter was detected by analyzing the cutting force acquired from a piezoelectric cutting force dynamometer. Liao and Young [40] proposed a spindle speed regulation method that relies on effective detection of chatter by evaluating the dynamic cutting force measured with a cutting force dynamometer. Smith and Delio [41] presented an optimal spindle speed selection algorithm that detects chatter based on analysis of the time domain signal acquired from an acoustic microphone. Kuljanic et al. [42] compared the sensitivity of different sensors for detecting chatter, and investigated

15

accuracy and robustness of single and multi-sensor systems. Ma et al. [34] developed a PVDF sensor based wireless monitoring system for chatter detection in the milling process. The extensive research works on sensors and signals have proven their effectiveness for chatter detection in a controlled research environment [36, 38].

Figure 7. A microphone used in a milling operation for chatter detection.

Developing chatter detection algorithms is the other emphasis of this line of work. Faassen et al. [9] proposed a chatter detection algorithm based on the power spectral density of vibration signals acquired from an accelerometer that recognizes chatter at its onset i.e. before chatter marks appear on the workpiece. Yao et al. [43] proposed a chatter detection method based on the wavelet transform and support vector machines. Zhang et al. [44] developed a hybrid hidden Markov model and artificial neural networks using vibration signals acquired from an accelerometer for classification of temporal cutting signal states (normal cutting, normal and transition, transition, transition to chatter, and chatter states). Ma et al. [8] developed a computationally efficient method for on-line

detection of chatter in milling using time-series data and a autoregressive model. Nguyen et al. [19] compared the effectiveness of chatter detection algorithms that utilize autocorrelation, wavelet transform, and the Fast Fourier Transform (FFT).

A common criticism of the on-line chatter detection methods reviewed above is that untimely detection of severe chatter may result in irreversible damage to the part and machine tool components [1]. To ensure timely detection of chatter, statistical machine learning methods such as support vector machines and artificial neural networks require a relatively large training data set of chatter signals gathered from the cutting process and the unique cutting system, namely the machine tool-tool holder-cutting tool-workpiece combination, which is costly when the purpose of the algorithm is to control chatter. On the other hand, algorithms that utilize time series autocorrelation or spectral analysis are commonly designed with an engineering threshold or control limits that need to be manually tuned by research engineers. Tuning of such thresholds can be difficult because the machine tool condition is dynamically changing, and the tooling, sensors, and sensor locations can be different from one application to another [10]. Therefore, it is unrealistic to expect manual tuning of thresholds for different processes and machine tools since research engineers cannot always be on the production floor, and the production staff such as machine tool operators typically do not have the necessary knowledge or experience to tune the chatter detection thresholds used in these algorithms.

Ultimately, the uncertainty and complexity of SLD identification makes it difficult to implement; the manual tuning of chatter detection thresholds or control limits for individual processes is challenging and costly. Although decades of work have advanced the mechanistic understanding of chatter in machining, the multivariate nature of the metal

removal process is a significant barrier that inhibits cost-effective prediction, detection, or suppression of chatter.

In this thesis, we focus on developing a robust solution for chatter detection in milling that can adapt to the multivariate flexible manufacturing environment by not only employing the previously developed chatter detection theories and methods, but also learning from the greatest asset in manufacturing, namely the human operator's perception and experience, via effective human machine interaction. In the following sections, relevant literature in the field of interactive computing and robotics that have inspired the work documented in this thesis are reviewed.

2.4   Learning from Demonstration

In robotics, Learning from Demonstration (LfD) is an interactive computing technique that seeks to replace the manual programming of a robot by an automatic programming process [12]. This process is driven by showing the robot the procedures associated with a specific task. The involvement of human demonstration requires effective facilitation of human-machine interaction that specifically encodes the human demonstrations of a task. While different approaches to LfD may contain unique elements, they share certain key traits. According to Chernova and Thomaz [11], all LfD works assume the existence of a human teacher, who performs demonstrations of the desired task that are fed into specific learning algorithms. The learned task is then reproduced by the learning agent through a repetitive reinforcement loop that enables constant refinement of the performance.

Specifically, in the context of robotics control applications, all LfD methods contain a teacher demonstration D, which consists of state-action pairs S-A, from which a policy π can be derived for the robot to execute the same task as demonstrated by the teacher. To solve an LfD problem is to determine S, A, and π. Different methods to acquire π from S-A have been investigated in this line of research.

What separates LfD from most other machine learning techniques is its objective to mimic the human's learning scheme, where learning is not an explicit activity that happens at a particular time instance, but rather is part of a continuous activity [11]. According to "the Correspondence Problem" by Nehaniv and Dautenhahn [14], temporal correspondence can be a challenge that inhibits the learning agent's ability to imitate and learn from a teacher. In the context of LfD, the correspondence must be facilitated by the mapping between the teacher and the learner's perceptions that enables transfer of knowledge from one to the other. In other words, the S-A pairs must be accurately mapped from a human's demonstrations to the robot's interpretation of π.

An important line of research in LfD concerns trajectory learning, and one way to represent a trajectory is through a time-series of the robot's end effector locations [45-48]. The trajectory determines the motion of a robot parameterized by time when performing a task such as moving an object from one location to another. The correspondence problem can be solved in trajectory learning by integrating the sensory perception of a human with the sensor inputs of the robot. For example, teleoperation in robotics via a joystick is a direct transfer of information that minimizes the impact of the correspondence problem. Through a human's demonstration, the trajectory information can be recorded as sequential states or state-action pairs. There are two major categories of modern approaches to LfD:

Dynamic Movement Primitives (DMPs) and probabilistic modeling methods such as Hidden Markov Models (HMMs) [11].

Dynamic Movement Primitive (DMP) is a framework built upon a series of papers authored by Ijspeert et al. [49]. The focus of DMP is to learn the attractor landscape of a controller from a single demonstration. The objective of DMP is to reach a target by the end of the action, where the system is modelled as a damped spring attached to the goal position with a non-linear modulation term $f(\omega, s)$:

$$\ddot{x} = K(g - x) - D\dot{x} + f(\omega, s) \tag{1}$$

where x, $\dot{x}$, and $\ddot{x}$ are the position, velocity and acceleration of the system, respectively, K is the spring constant, D is the damping coefficient, and g is the goal position. $f(\omega, s)$ is the non-linear modulation term that adds acceleration to the system. If $f(\omega, s) = 0$, g is the point attractor, and K and D are the proportional derivative (PD) gains of the system, respectively, that are ideally set such that the robot moves to the goal position without overshoot. The objective of the DMP is then reduced to tuning $f(\omega, s)$ based on a single demonstration, from which x, $\dot{x}$, and $\ddot{x}$ can be obtained [11].

Another popular LfD approach uses probabilistic models. For example, the first order Hidden Markov Model (HMM) is the most basic Dynamic Bayes Net. In the context of LfD, the skill to learn (e.g. trajectory) is constructed as a chain of hidden states, with prior probabilities, transition probabilities and observation probabilities [11]. Specifically, a set of probabilistic transition functions, $T(s'|s, a): S \times A \times S \rightarrow [0, 1]$, form a Markovian chain. Because the states S are hidden, they can only be approximated through the mapping $M:S \rightarrow Z$, where Z is the observable state and the policy $\pi: Z \rightarrow A$ selects the actions based on the observations of the world states. The human demonstration is fit into

the chain of transition functions and observation probabilities, and the parameters that contribute to the probabilities can be learned through expectation-maximization algorithms. Kulic et al. [50] described the pipeline for applying HMM to learn a trajectory from human demonstration. Lee et al. [51] stressed the importance of compensating missing information from discretizing continuous observations.

There are two key takeaways from the common LfD approaches in robotics. First, an effective human-machine interface must be designed and implemented to solve the correspondence problem between demonstration and the acquired information. Effective acquisition of signal and compensation for missing information are critical to generate discrete states that closely represent the continuous perception of a human. The second takeaway is that a primitive which facilitates learnable information must first be developed for the demonstration to be learned. For example, in the context of robot trajectory learning, DMP is based on control theory where the motion is learned via training the non-linear modulation term through the demonstration [52]. To draw an analogy with machining, the acquired signal from the machining process must be synchronous with the human operator's observation abilities and must provide sufficient resolution to minimize loss from discretization; a "learnable skill primitive" for chatter detection that enables learning from human demonstration must be developed based on fundamental theory and understanding of machining chatter.

2.5   Reinforcement Learning

A single demonstration can be enough to yield the desired outcome of an event. However, the policy π that selects the action to reach the desired outcome may not perform

well after a single demonstration. A common approach is to refine the learned policy through reinforcement learning (RL). RL refines a newly trained policy through environmental reward and exploration [53]. The learning agent is connected to the environment through perception and action. While the state is changed by the action, a reward signal based on the learning agent's current policy is fed back to the learning agent. RL aims to optimize the agent's policy to maximize the long-term sum of rewards. Systematic trial and error can be performed over time to improve the policy performance [54].

The disadvantage of traditional reinforcement learning is clear. First, the reward function for policy iteration is hard to determine, and it is traditionally hard coded by the developer who may not always select an optimal reward function for the learning task. Second, the number of iterations necessary for reinforcement learning methods to achieve acceptable accuracy is large. Its applicability in physical processes such as machining chatter detection where the negative impacts of error outweigh the gain from accuracy enhancement is limited. In other words, it is impractical and costly to iterate on chatter experiments hundreds of times before obtaining a sufficiently accurate chatter detection algorithm. In addition, constantly changing machining process parameters and cutting tool combinations represent a dynamic environment for the reinforcement learning framework, which adds another layer of complexity. Therefore, adapting reinforcement learning to machining process monitoring is challenging. Experts in the field of computing have also realized this limitation of traditional reinforcement learning and have therefore proposed an interactive policy shaping technique that integrates user input with RL. In the next section, relevant techniques of interactive agent shaping are discussed.

2.6   Interactive Agent Shaping

Interactive agent shaping is defined as the problem of learning from human reward. According to Knox and Stone [16], the fundamental problem solved by interactive agent shaping is how an agent can learn by leveraging the human's observation and critique of the learning agent's behavior. Teaching an Agent Manually via Evaluative Reinforcement (TAMER) is a framework developed to formulate the human's critique as a human reward function that instructs the agent to greedily select actions that maximize the expected immediate reward from the human [15]. Various applications of the TAMER framework in the virtual world and in robotics have proven its performance in training an agent to quickly learn an effective policy.
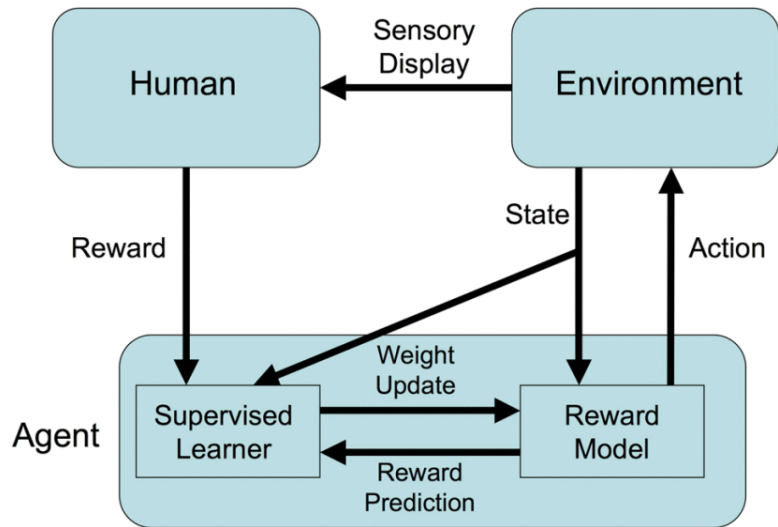
Figure 8. TAMER framework [15].

As shown in Figure 8, the credit assignment problem in traditional reinforcement learning where a reward must be assigned to the appropriate state-action pair is resolved by allowing an attentive human trainer to provide immediate temporal rewards to the

23

corresponding states. The TAMER framework assumes that the trainer has taken into consideration the long term reward when teaching the policy, and therefore, results in a significantly reduced number of iterations for policy refinement [15].

One of the major breakthroughs in the TAMER framework is the development of a credit assigner that assigns the human trainer's response to the corresponding states. The necessity of an effective credit assigner stems from the high frequency of recorded states for temporal tasks and the natural delay of the human's reaction. The TAMER framework addresses this challenge by modelling the expected delay and mapping the feedback to a set of recorded temporal ranges/states. The delay was modeled as a uniform distribution in an example of robot learning [55]. As a result, the TAMER framework effectively assigns the human input to the corresponding states as rewards.

In addition to human rewards, it has been reported that a human's input may directly modify the action selection mechanism of the reinforcement learning algorithm. Thomaz and Breazeal [56, 57] integrated the environmental reward with human reward and a human guidance input with which the human teacher provides anticipatory reward for a possible future event. Smart and Kaelbling [58] used demonstration to indicate the area of interest in state-space, and combined reinforcement learning to achieve better-than-human performance. The supervised actor-critic reinforcement learning algorithm integrates the human reward and demonstration to allow teachers to improve the performance of a robotic assembly task.

The various research works that have been conducted in the field of interactive agent shaping and reinforcement learning show that the performance of the learning agents obtained using this approach is promising. Chatter monitoring shares key elements of the

24

interactive agent shaping problem, where an expert human operator can provide feedback and demonstrations by mapping them to the corresponding temporal states of a machining operation.

2.7   Summary

In summary, this chapter reviewed the relevant literature in machining chatter and interactive computing. On one hand, although previous works in chatter prediction and recognition have advanced our understanding of chatter, and to a certain extent provided case-specific solutions to control chatter in machining, the general applicability of these methods in a flexible manufacturing system remains a significant challenge. On the other hand, this chapter reviewed influential papers in learning from demonstration, reinforcement learning, and interactive agent shaping that have advanced the performance of human-machine interaction in virtual and robotics applications. In the following chapters, this thesis aims to propose a step-by-step solution that facilitates human-machine interaction in a machining chatter monitoring application.

CHAPTER 3 PERCEPTION MAPPING FOR HUMAN-MACHINE INTERACTION IN CHATTER DETECTION

3.1   Introduction

In this chapter, the problem of on-line chatter detection in a machining operation is addressed, followed by a solution approach that draws inspiration from interactive computing techniques in robotics. In the following sections, an analogy is drawn between human-machine interaction in chatter detection and learning from demonstration in robotics, humanistic intelligence and the modes of human-machine interaction in chatter detection are defined and modelled, a human-machine interface for monitoring of a machining operation is presented, and the delayed human response to chatter is introduced as a basis for perception mapping between a human operator and the chatter detection algorithm.

3.2   Chatter Detection and Learning from Demonstration

Chatter can result in serious consequences such as poor surface finish, excessive tool wear and machine tool spindle damage [1]. On-line chatter detection focuses on detecting chatter accurately at its onset before chatter is fully developed and causes severe damage to the workpiece. To a certain extent, research works on on-line chatter detection have shown success in applying chatter detection algorithms that satisfy the accuracy and detection speed requirements [8, 9, 19]. However, a common drawback of existing on-line chatter detection algorithms is that the chatter detection threshold must be tuned manually by engineers who have deep knowledge, skills, and experience in programming the chatter

detection algorithm. When research engineers are onsite, adjustments can be made to account for variations in tooling, fixtures and workpiece materials, etc.; however, after the engineer departs, the production staff typically do not have the knowledge or time to make necessary adjustments to the chatter detection threshold [10]. Because onsite tuning of the chatter detection thresholds is crucial for achieving the specified detection accuracy and speed, the on-line chatter detection approach has not yet found wide acceptance in production settings.

Instead, over the years, manufacturers have relied on experienced human operators for chatter detection and control in a machining operation. While the experienced operator's perception is robust in detecting chatter despite variations in tooling types, fixtures or workpiece materials, human operators typically lack the necessary reaction speed to control chatter in a timely manner. In addition, the increasing shortage of experienced operators and waves of retirements have motivated the manufacturing industry to invest in robust alternatives that can detect chatter accurately and timely.

In robotics, learning from demonstration (LfD) has been employed as a methodology for learning a policy, i.e., a decision mapping between states and actions, or what actions the robot should perform when a particular state is present [13]. As a simple example, in a robotic trajectory learning application, the states can be a combination of joint angle, joint velocity, and the target position; the action can be the joint effort for the robot to reach the target position. The motivation behind LfD is that the development of policies by hand is challenging. With increasing levels of complexity, and the required degree of adaptation to a new application, the robot's policy must be highly customizable [11]. Hard coding the policy by hand in a dynamically changing environment is inefficient

and sometimes impossible due to the complexity of the task. To draw an analogy between robotics and chatter detection, a chatter detection threshold is a binary classifier, which distinguishes unstable cutting (chatter) from stable cutting, which represent two states of interest in a machining operation. The chatter detection algorithm can be considered as part of a policy that classifies the states of cutting (stable versus unstable), and commands chatter suppression. For the purpose of this thesis, the control actions that suppress chatter are not investigated. Instead, developing an accurate, timely, robust, and easy to implement chatter detection algorithm is a prerequisite for any effective control action. Chatter detection thresholds as part of a policy have often been hard coded by manufacturing engineers. However, because the production environment is dynamically changing due to variations in tooling, fixtures and workpiece materials, hard coding the threshold is inefficient and costly. Therefore, the success of LfD in robotics provides motivation for this work to investigate a methodology that can learn the chatter detection thresholds from a human operator, who may not have the skills to program the thresholds, but is always onsite and has the experience and perceptive knowledge for chatter detection.

A common LfD technique involves the development and implementation of dynamic movement primitives (DMPs). The fundamental inspiration for DMPs is the theory that biological systems such as humans perform a movement task as a combination of overlapping movement primitives also known as units of actions [59]. The fact that DMPs of a robot can be learned as stable non-linear attractor systems from demonstrations that can adapt to the dynamically changing, stochastic environment makes DMPs a common policy learning approach in motion control. Similarly, in chatter detection, human operators recognize chatter as an emerging abnormal and dominant pitch of sound

28

produced by the cutting process, which is a dynamically changing environment. Based on experimental data [17], the chatter frequency is a function of the vibration mode that has been excited by the cutting process. For a given cutting system, namely, the machine tool-tool holder-end mill assembly, there are multiple modes of vibrations that can result in different chatter frequencies. Figure 9 shows three distinct chatter frequencies with different amplitudes produced by the same cutting system. In this work, which is inspired by DMPs that can be learned from human demonstrations and serve as building blocks of a more complex task, we hypothesize that each chatter frequency from a cutting system, induced by a vibration mode, is a dynamic primitive that can be learned from human demonstrations of chatter detection. Furthermore, for each chatter frequency to be detected accurately and in a timely manner, a frequency-specific chatter detection threshold must be established. A more in-depth analysis of whether different chatter frequencies affect the chatter detection thresholds is presented in Chapters 4 and 5. Ultimately, the production process is a complex task that involves changing cutting systems and workpiece materials. A chatter threshold for a specific chatter frequency may not be applicable to another chatter frequency, and hence the detection thresholds for different chatter frequencies can be treated as building blocks of a more complex process monitoring task. Chapter 4 presents a detailed description of the methodology used to develop a "learnable skill primitive" for chatter detection that facilitates learning the detection threshold for a chatter frequency from human demonstrations.
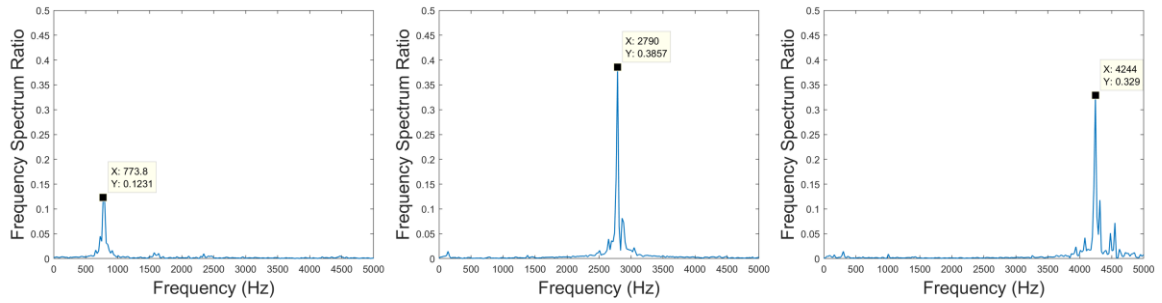
Figure 9. Different chatter frequencies of a cutting system.

While the analogy between LfD in robotics and LfD in machining chatter detection is evident, it is important to note that there are still significant barriers for direct implementation of learning from demonstration techniques in a machining chatter detection application. First, machining chatter is a process anomaly that ideally should be avoided and not reproduced, which implies that demonstrations of chatter detection should be limited to one or at most a few shots in a production setting. Second, because chatter develops quickly, the delay associated with a human's reaction time prohibits timely demonstrations that directly correspond to the time instance of chatter occurrence. A perception mapping algorithm is therefore necessary to correlate the time instance when the human operator signals chatter based on his/her perception and reasoning, for instance by pressing a button, with the actual time instance when chatter occurs. Finally, the variability in a human operator's reaction time, which is commonly described by a distribution, can result in inconsistent demonstration qualities. The learning from demonstration methodology for chatter detection must account for these factors.

As described above, the presence of a human operator in a chatter detection application may entail higher complexity than in some robotics applications. Therefore, to understand the operator's perception and reasoning in a chatter detection setting and to

investigate the relationship between the operator and the "digital apprentice", the next section presents an in-depth description of the humanistic intelligence approach and the modes of human-machine interaction in chatter detection.

3.3 Humanistic Intelligence and Modes of Human-Machine Interaction

In this section, instead of a one-way demonstration from a human operator to the "digital apprentice", which is simply referred to as the machine in this section, the presence of the human operator in chatter detection is modelled as a computational loop where the human and machine share perception and reasoning. This relationship is characterized as humanistic intelligence.

Humanistic intelligence, according to Mann and Minsky [60, 61], is a form of intelligence created by integrating a human being in the feedback loop of a computational process. Essentially, the focus of this thesis is to effectively integrate a human operator into the computational loop of an on-line chatter detection algorithm. By combining the processing power of a micro-controller with the associative decision-making ability of a human operator via effective human-machine interaction, the humanistic intelligence approach offers the potential to collectively generate high quality classifiers, i.e. thresholds for accurate and timely chatter detection.
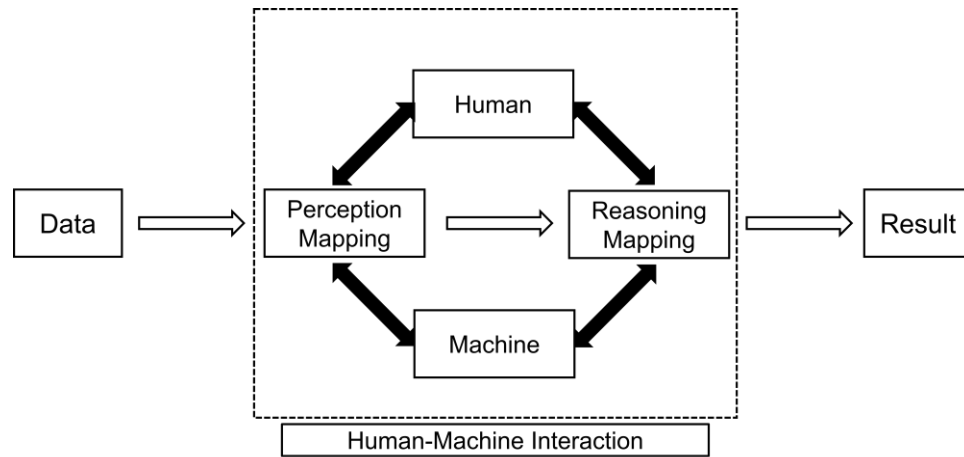
Figure 10. Perception and reasoning mapping in human-machine interaction.

In a humanistic intelligence system for chatter detection, the human and the machine exchange perception and reasoning through an interface. Figure 10 shows the perception and reasoning mapping between human and machine that enables collective decision making. In the context of chatter detection, the human is an expert operator who is experienced in the characteristics of machining chatter, and the machine is a computational system that monitors the manufacturing process via sensors and algorithms. The value proposition is that the perception and reasoning mapping between the human and the machine can leverage the strengths of the human and the machine, thereby avoiding the negative impacts of their respective disadvantages to achieve accurate and timely detection of chatter.

Human perception is enabled by the natural senses, primarily hearing, sight, and touch in the context of chatter detection. Human reasoning in the context of chatter detection is the cognitive ability to recognize chatter based on prior machining experience. The assumption here is that the human expert can distinguish between chatter and stable cutting based on past occurrences of chatter. Similarly, the perception ability of the

32

machine is derived from signals gathered through sensors such as microphones, accelerometers, or piezoelectric force dynamometers. The reasoning ability of the machine consists of an algorithm that determines the sensor signal threshold for chatter detection. The natural senses of the human and sensor signals of the machine can be temporally mapped to capture information about the machining process. Figure 11 is a Venn diagram that depicts the knowledge space of a task. Perception mapping is possible when a piece of information is perceivable by both human and machine. There are two modes of human-machine interaction enabled by perception and reasoning mapping for chatter detection: 1) machine learns from the human's demonstration to establish a threshold for chatter detection, and 2) machine serves as an advisor to the human by communicating critical information that may not be accessible or quantifiable by the human. The focus of this thesis is to develop a pathway for the first mode of interaction in the context of the chatter detection problem.
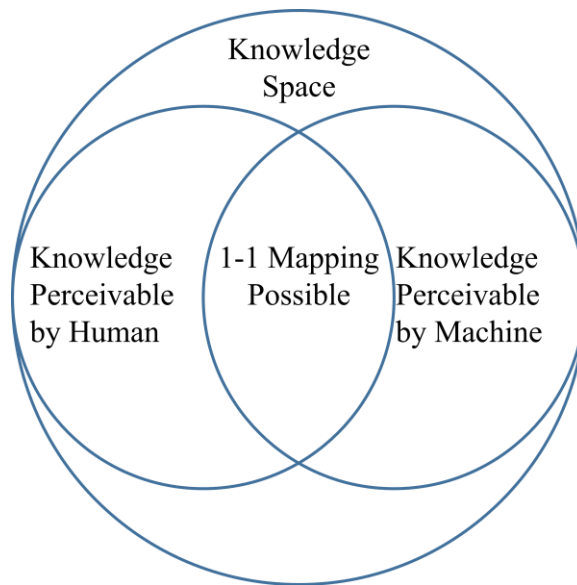
Figure 11. Knowledge space and perception mapping of learning agents.

The first mode of interaction improves the robustness of the machine's detection algorithm by allowing threshold adjustments based on human demonstration while the second mode of interaction enhances the accessibility of data related to recognition of chatter by other human operators or by a computational system. Table 1 lists the advantages and disadvantages of the human's cognitive capabilities for recognition of chatter and those of a machine based algorithm for chatter detection. It is worth noting that a machine's chatter detection algorithm is analogous to a human's ability to recognize chatter. Without specific knowledge of chatter, it is difficult for a lay person to distinguish between chatter and stable cutting. A human's lack of accessible/sharable memory can be improved by the machine's accessible data, and the inflexibility of a machine's detection algorithm can be complemented by the robustness of a human's cognitive ability to recognize chatter under different machining conditions.

Table 1. Comparison of a machine-based chatter detection algorithm with a human's cognitive ability.

	Machine's Chatter Detection Algorithm	Human's Cognitive Ability to Recognize Chatter
Advantages	Accessible memory Quantifiable data	Self-adaptive Robust in different operations
Disadvantages	Vulnerable to variations in cutting system	Difficult to access and share

The challenge remains in creating an efficient means for perception mapping in the context of chatter detection. Because of the inaccessibility of human memories and perceptions, continuous real-time exchange of natural perception data is impossible based on currently available technology. The alternative is to enable perception mapping as soon as a critical event is captured by the human's perception through physical exchanges. In

the next section, an example of a human-machine interface for chatter detection is presented as a means for perception mapping.

3.4   Human-Machine Interface for Perception Mapping

The objectives of perception mapping in a machining chatter detection application are to 1) enable the machine to learn a chatter detection threshold from a human operator's demonstrations, and 2) display critical chatter detection data obtained from the machine's sensors. The following functional requirements have been identified for designing the human-machine interface for chatter detection:

1. The interface enables both machine input and human input.

2. The interface allows simultaneous signal input from the machine and human.

3. The interface should confirm the receipt of human input.

4. The interface provides a visualization of the mapped perception.

5. The interface should provide a reset option.

6. The interface should be intuitive and easy to operate.

7. The interface takes minimal space for accessibility in a production setting.
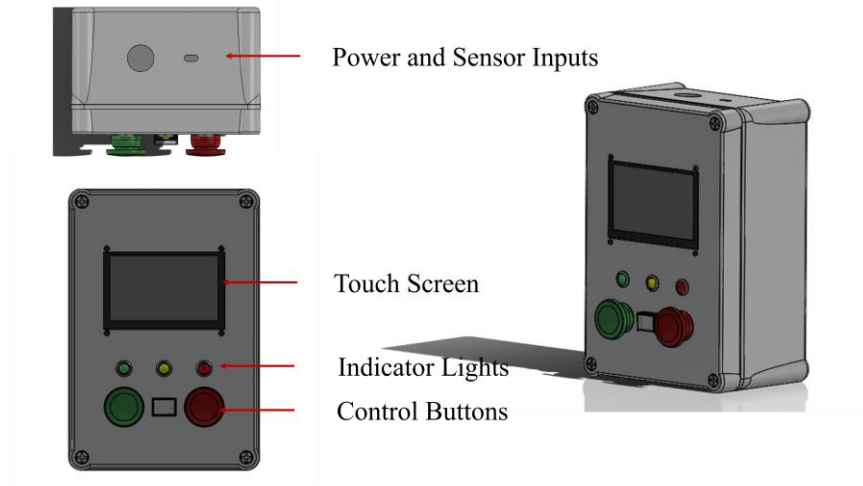
Figure 12. The human-machine interface (HMI) design.

Figure 12 is a CAD model of the HMI designed for the experiments in this work. In the top view, input ports for the power and sensor signals are indicated. In the front view, the touch screen satisfies the basic functions for data collection and visualization, where the machine's input and the human's input can be displayed. The control buttons at the bottom of the interface resemble those on a Computer Numerical Control (CNC) machine tool. Whenever a control button is pressed the corresponding indicator light flashes once to indicate the operation. When chatter is detected by the algorithm, all three indicator lights stay on. It is also intuitive to operate this interface; as soon as chatter is observed, the operator simply presses the red control button; if chatter is falsely detected by the algorithm and the indicator lights stay on, the operator presses the green control button to indicate a stable cutting signal. The grey control button (rectangular button in the middle) is a reset button, which can reset the detection algorithm to its default setting.

Using this HMI, an operator can effectively communicate his/her observations to the machine by pressing the appropriate control buttons. The interface is portable and can

be mounted on the CNC machine tool to facilitate timely human-machine interaction. The specific data acquisition method and chatter detection algorithm developed through the HMI are described in Chapter 4. In the next section, the delayed human response to chatter is discussed and a method to collect a human subject's reaction time is presented.

3.5 Delayed Human Responses in Perception Mapping

The human-machine interface described in the previous section enables in-process perception mapping from human to machine. Here we assume discrete-time signals are represented as a time series, t, where the nth time instance is denoted by t_n. The time instance when a chatter mark first appears on the workpiece is denoted as t_a. In perception mapping for chatter detection, the human should ideally provide a signal via the control buttons at the time instance the human first observes chatter, t_c. Whether t_a coincides with t_c is evaluated in Chapter 4. However, there is a natural delay associated with the human signaling chatter after observing it, and therefore the signal is received by the interface at time instance t_s. This delay is determined by the equation:

$$r = t_s - t_c, \qquad t_s \geq t_c \tag{2}$$

where r is the human's reaction time to signal the occurrence of chatter. There are multiple sources that can contribute to the delayed chatter signal from the human operator. These include:

1. The time needed to recognize chatter.

2. The time between recognition of chatter and physically pressing the button.

3. The time between pressing the button and the controller receiving the signal.
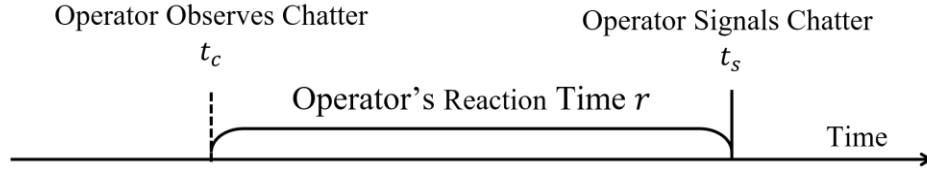
Figure 13. Graphical representation of human observation of and reaction to chatter.

It is likely that the reaction time is unique to each human subject. Other sources of delay may also contribute to the reaction time. For the purpose of this research, the sources of delay are not investigated. The reaction time to chatter, r, is treated as a lumped parameter that represents the overall temporal delay in receiving human input.

Research on human subjects' reaction time in visual search problems has shown that it can be described by a probability density function such as Gaussian, ex-Gaussian, ex-Wald, and the Gamma distribution [62]. By establishing a reference timeframe for the observation, t_c, the reaction time r can be characterized by a probability density function, f_{react}, where the probability of r falling in a specified range is defined as:

$$Pr[\alpha \leq r \leq \beta] = \int_{\alpha}^{\beta} f_{react}(r) \, dr, \qquad \alpha, \beta > 0 \qquad (3)$$

where α and β are positive numbers that specify the range of reaction times. The typical objective of human subject reaction time research is to evaluate whether any physical impairment or disease may have a negative impact on the reaction time. The observation time instance t_c is known and r can be obtained from the sampled signal time instance t_s.

In machining chatter detection, however, t_c is unknown and can only be estimated from known values of t_s. To trace back the original observation of the event from the recorded signal, we define a probability density function, f_{trace}, as follows:

$$Pr[t_s - \beta \leq t_c \leq t_s - \alpha] = \int_{t_s-\beta}^{t_s-\alpha} f_{trace}(\widehat{t_c}) \, d\widehat{t_c}, \qquad \alpha, \beta > 0 \qquad (4)$$

where $\widehat{t_c}$ is an estimate of the true observation time t_c, and α and β are positive numbers that specify a range of the integral of the same size as equation (3). A similar problem of assigning credit to train a learning agent with only the signal time instance has been discussed by Knox et al. who proposed the TAMER framework for high action frequencies training where multiple states can be observed within a second [15, 16]. One example of f_{trace} provided by Knox et al. is a uniform distribution ranging from -0.8 to -0.2, which is adopted from a paper on human reaction times in visual search problems [63]. Because operators primarily observe chatter through their auditory sense of hearing, and a subject-specific human-machine interface has been designed and developed for chatter detection, we performed controlled reaction time experiments on human subjects using the HMI for chatter detection to determine f_{react}. This probability density function is then utilized as the basis for developing the learnable skill primitive algorithm in Chapter 4.
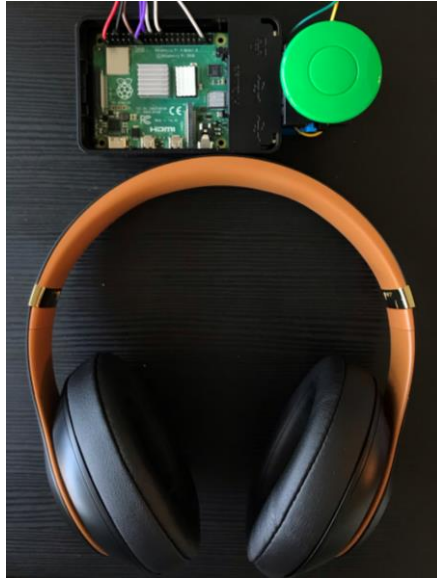
Figure 14. Reaction time collection apparatus.

Figure 14 shows the HMI apparatus for the reaction time experiments. The experiments were conducted per the Georgia Institute of Technology IRB Protocol H20340 [64], which was specifically drafted for human subject research in chatter detection. The human subjects were instructed to wear the headphones before the experiment. Samples of synthetically generated sounds of different frequencies that lasted one second each were played, and the human subjects were asked to respond to the sound by depressing the control button as soon as they heard the sound. In total, each human subject responded to 90 samples of synthetically generated tones that respectively simulated three different frequencies of chatter. Table 2 lists the number of samples generated for the simulated chatter frequencies of 500 Hz, 2000 Hz, and 3500 Hz, respectively. These frequencies were selected based on a reported range of chatter frequencies observed in a well-cited milling study [17]. One of three silence intervals (2, 3, or 4 seconds) between two consecutive

sound samples were assigned to 30 samples, and the samples were fully randomized to avoid the subject's anticipation of the sound.

Table 2. Human subject reaction time experiments.

Frequency	2 seconds silence interval	3 seconds silence interval	4 seconds silence interval	Subtotal
500 Hz	10 samples	10 samples	10 samples	**30** samples
2000 Hz	10 samples	10 samples	10 samples	**30** samples
3500 Hz	10 samples	10 samples	10 samples	**30** samples
Subtotal	**30** samples	**30** samples	**30** samples	**90 samples**

Figure 15 is a histogram of the reaction times obtained from the 90 samples collected from a human subject. The samples are distributed in 10 equally spaced bins, and a Gaussian distribution is fit to the samples for analysis. The sample mean reaction time $\bar{r}$ is then obtained as:

$$\bar{r} = \frac{1}{n}\sum_{i=1}^{n} r = \frac{1}{n}\sum_{i=1}^{n}(t_{s,i} - t_{c,i}), \qquad n = 90 \tag{5}$$

where n is the number of samples. For the reaction time of operator C shown in Figure 15, $\bar{r}$ is 0.395 seconds and the sample standard deviation, $s_{\bar{r}}$, is 0.069 seconds. Note that in this work the Gaussian distribution is used to model f_{react} in terms of the mean and standard deviation of the data set since it fits the data reasonably well. Other distributions may fit the reaction data more accurately; however, for our purposes, f_{react} only serves as a basis

for determining the parameters in the perception mapping algorithm. In Chapter 4, the model of f_{trace} is further relaxed to be a uniform distribution for simplicity and processing speed. Therefore, the fitting accuracy of the f_{react} distribution is less significant.
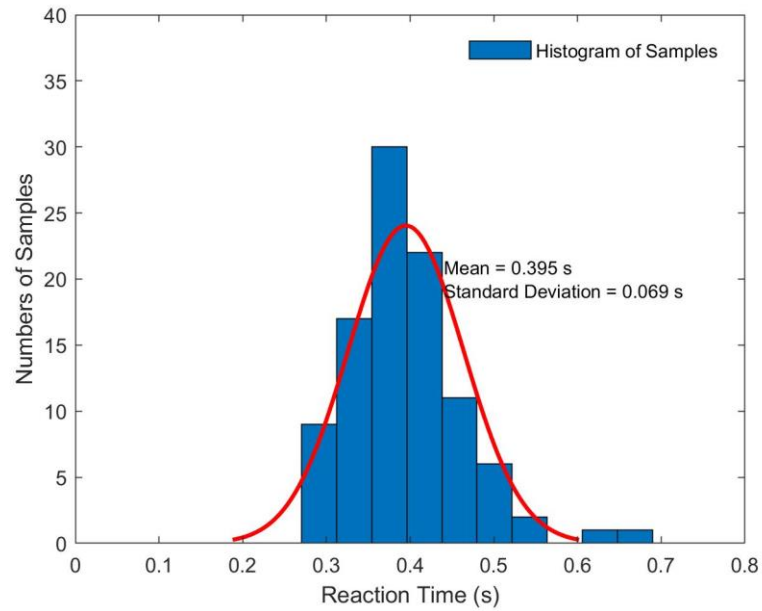
Figure 15. Reaction time histogram and distribution of 90 samples from a human subject.
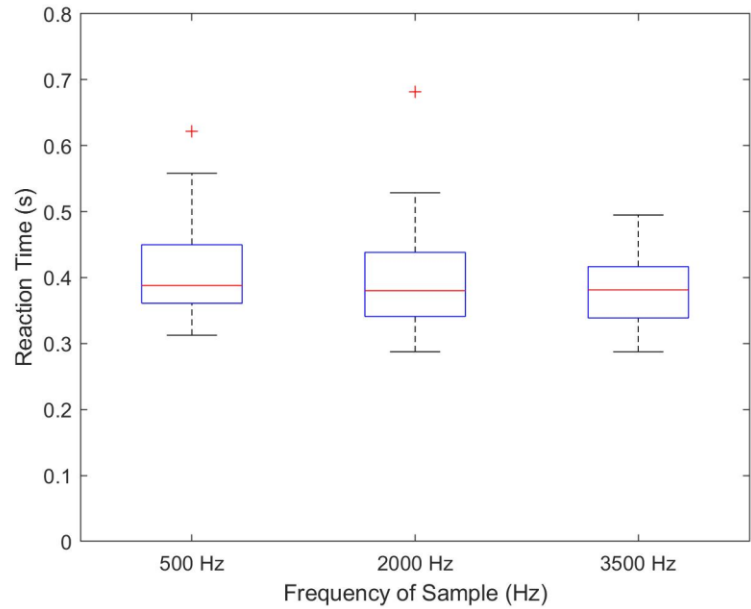
Figure 16. Reaction time boxplot for different chatter frequencies.

To evaluate the effect of different chatter frequencies on a human subject's reaction time, assuming the data sets follow a Gaussian distribution, one-way ANOVA was performed on the data sets corresponding to the three simulated chatter frequencies.

Table 3. Statistics of reaction time data from human subject experiments.

	Mean Reaction Time (s)	Standard Deviation (s)	Reaction to Different Chatter Frequencies ANOVA P-Value
Operator A	0.306	0.061	3.35×10^{-7}
Operator B	0.323	0.079	9.43×10^{-5}
Operator C	0.395	0.069	0.1455

Table 3 lists the mean and standard deviation of the reaction times for each human subject and the corresponding P-value of the ANOVA. As expected, each operator has a unique reaction time to the synthetically generated sound. The largest difference between

the mean reaction times is 0.089 seconds, which is large since it is 30% of the fastest mean reaction time. There is also an approximately 30% difference between the largest and smallest standard deviations. Finally, the P-values of the ANOVA tests indicate that for some operators the reaction time is significantly different based on the chatter frequency, while other operators do not seem to be affected as much. Therefore, the dependence of the subject's reaction time to the chatter frequency is inconclusive. In Chapter 4, the chatter detection algorithm is calibrated using the unique reaction time data for each individual operator.

3.6   Summary

In this chapter, the analogy between LfD in robotics and its application to machining chatter detection was discussed. The human-machine interaction in chatter detection was modelled as a humanistic computational loop, a human-machine interface for the machining process was designed, and a methodology to evaluate a human operator's reaction time for use in perception mapping of the human's chatter detection ability to the machine's ability to detect chatter was presented.

CHAPTER 4 ON-LINE CHATTER DETECTION THROUGH LEARNABLE SKILL PRIMITIVE

4.1   Introduction

This chapter presents the development of the "learnable skill primitive" (LSP) approach for chatter-specific knowledge transfer from a human operator to a machine. Using the perception mapping methodology described in Chapter 3, the LSP is developed as a learning algorithm for identification of a chatter detection threshold based on temporal input signals from a human operator and the delayed reaction time distribution obtained from human reaction time experiments. The LSP method stems from on-line chatter detection algorithms that are based on frequency spectrum analysis. The frequency decomposition of the audio signal generated during machining is leveraged by the human demonstration through LSP method. Because the LSP method is developed as a generic algorithm for extracting a chatter detection threshold inversely from demonstration, the objective is to improve robustness (better accuracy and speed across a range of cutting conditions and workpiece materials) in comparison to deterministically setting chatter detection thresholds.

In the following sections, a detailed description of the data acquisition and signal processing method is presented, the LSP methodology for establishing the chatter detection threshold is introduced, and the on-line learning performance of the LSP is experimentally evaluated.

4.2   Experimental Setup and Data Acquisition

4.2.1   Milling Experimental Setup

A milling process is selected for the experiments. Figure 17 shows the experimental setup. The experiments were conducted on an Okuma MILLAC 44V 3-axis CNC milling machine using a four flute solid carbide end mill with 12.7 mm diameter and 25.4 mm cutting length, which was held in an Iscar CAT 40 end mill holder. A USB Blue Yeti studio style condenser microphone (model number 988-000101) was mounted inside the milling machine as seen in the figure. Multiple passes of 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) under different axial depths of cut and spindle speeds to generate various chatter and stable cutting signals. As lot end milling process was selected to temporarily isolate chatter signals from large amplitude variations at the tooth passing frequency in the cutting signal that are typical of a peripheral end milling process. A more in-depth analysis of the effectiveness of LSP under radial immersions representative of peripheral milling is presented in Chapter 5.

Figure 17. Experimental setup of machine tool-tool holder-end mill system and microphone.

Figure 18. AISI 4140 Steel workpiece with test cuts.

4.2.2   Sensor Selection and Signal Processing Method

The objective of the LSP is to learn a chatter detection threshold from a human operator's demonstration. A fundamental hypothesis of this work is that the human operator recognizes chatter as an emerging dominant frequency of sound in the milling process. At chatter onset, a significant increase in amplitude of a specific chatter frequency is observed. Therefore, to obtain useful information that can be gained from a human operator's demonstrations, the sensor and signal processing methodology must emulate the perception and reasoning ability of a human operator during chatter detection.
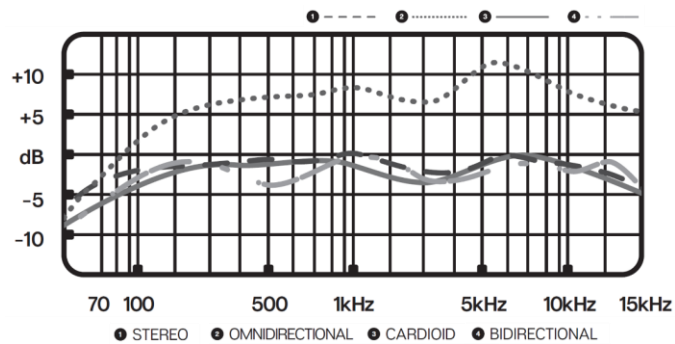
Figure 19. Frequency response curves of the blue yeti microphone [65].

Acoustic sensors such as microphones sample audio signals that mimic a human's hearing ability. In this work, a Blue Yeti USB studio-style condenser microphone was selected for its relatively flat frequency response and the integrated analog-to-digital signal converter. The cardioid mode was chosen to isolate the direction of the audio signal by facing the microphone toward the end mill. The frequency response of the microphone is suitable for chatter monitoring because it provides comparable frequency responses across a wide range of frequencies from 20 Hz to 20 kHz. The microphone's integrated analog-to-digital converter enables direct processing of signals by the micro-controller without an additional signal processing interface. A Raspberry Pi 4 Model B (from CanaKit) was

selected as the micro-controller to process signals acquired from the USB microphone. The Raspberry Pi Unit in this work has 4 GB of RAM and a quad-core processor, making it suitable for complex online signal processing and visualization. In addition to processing power, the Raspberry Pi has a Linux-based desktop interface for programming in Python. 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 [66]. The 16-bit digital signals were converted into a series of integers ranging between -32768 and 32767, or -2^{15} and $(2^{15} - 1)$.
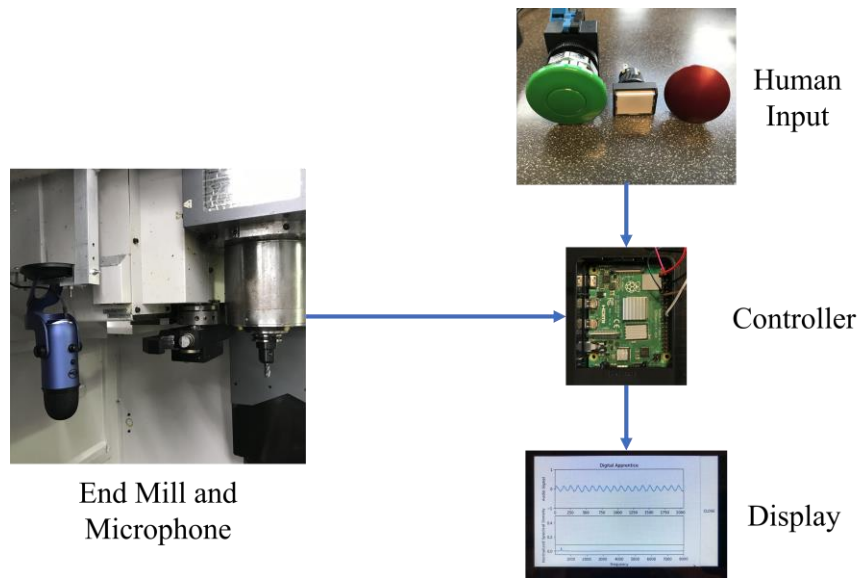
Figure 20. Experimental workflow for data acquisition.

The raw data was then processed for chatter detection. Although other computationally efficient chatter detection methods exist [8, 19], the Fast Fourier Transform (FFT) is a frequency spectrum analysis method that identifies the frequency

content in an manner similar to the human's hearing ability [67, 68]. The FFT is a computationally efficient method of Discrete Fourier Transform (DFT) suitable for real-time signal analysis. The DFT is given as:

$$X_F(m) = \sum_{n=0}^{N-1} x_n e^{-i2\pi m \frac{n}{N}}, \quad m = 0, \cdots, N-1 \tag{6}$$

Frequency spectrum analysis using FFT enables a range of frequencies and their amplitudes to be identified, which correspond to human perception's pitch and volume, respectively. Applying FFT to the audio signals results in data that closely resembles the human's perception of chatter. Although chatter frequencies are typically lower than 10 kHz, a human ear can detect frequencies between 20 Hz to 20 kHz. Therefore, the sampling rate of the audio signal was 48 kHz, which means frequency content up to 24 kHz can be identified by the FFT. The selected FFT bin size, N, is 2048; i.e. for each set of 2048 data points collected at a rate of 48000 data points per second, the FFT is performed once, which results in a frequency resolution of approximately 23.43 Hz. The chosen bin size of 2048 is a compromise between good FFT resolution and high update speed. With the FFT updated every 0.043 seconds, the changes in amplitudes of the frequencies are captured quickly, while an acceptable FFT resolution is maintained.

4.3 Learnable Skill Primitive for Chatter Detection—Methodology

4.3.1 On-line Chatter Detection Methods based on Frequency Spectrum Analysis

The audio signals acquired from an unstable milling operation consist of the periodic spindle rotation frequency, the tooth passing frequency and its harmonics, the chatter frequency, and other environmental and measurement noises. Nguyen et al. [19]

reported an efficient chatter detection method in turning that compares the magnitudes of the two largest peaks in the FFT. The peak ratio threshold P_R is defined as the threshold for chatter detection:

$$P_R = \frac{P_1}{P_2} \qquad (7)$$

where P_1 and P_2 are the amplitudes of the two largest peaks in the FFT. The assumption here is that chatter frequency is dominant when chatter occurs, and P_1 is the amplitude of the chatter frequency. When the ratio of the peaks exceeds P_R, the algorithm determines that chatter has occurred. However, the value of P_R must be experimentally determined and manually tuned by research engineers. The pre-determined threshold can be difficult to set due to variations in sensors, cutting tool, and cutting conditions.

Therefore, instead of relying on research engineers to tune the chatter detection thresholds manually, the LSP learns the chatter detection threshold from a human operator's demonstration, which leverages the skill and experience of the operator in recognizing chatter. To accurately detect chatter in a milling operation as early as possible, there are two parameters that need to be learned from the human operator's demonstration: 1) the dominant chatter frequency $f_{chatter}$, and 2) the chatter detection threshold P_{th}.

4.3.2 Identifying the Dominant Chatter Frequency

Based on the human operator's demonstration, the dominant chatter frequency is identified as the highest peak in the FFT at time instance t_s, which is the time instance when the human operator signals chatter by pressing a push button switch. Because of the natural delay in the human's operator's reaction to detection of chatter, the hypothesis is that the amplitude of the chatter frequency has increased significantly by time instance t_s

51

such that the chatter frequency has the highest amplitude among all frequencies in the spectral decomposition obtained through the FFT. This hypothesis is confirmed by experimental data. Figure 21 shows the chatter frequency at approximately 770 Hz as the highest peak in the FFT. The process to identify the chatter frequency is represented symbolically as:

$$f_{chatter} = argmax_m(|X_F(m)|) \cdot \frac{f_s}{N} \qquad (8)$$

where f_s is the sampling frequency of the audio signal, and N is the bin size of the FFT.
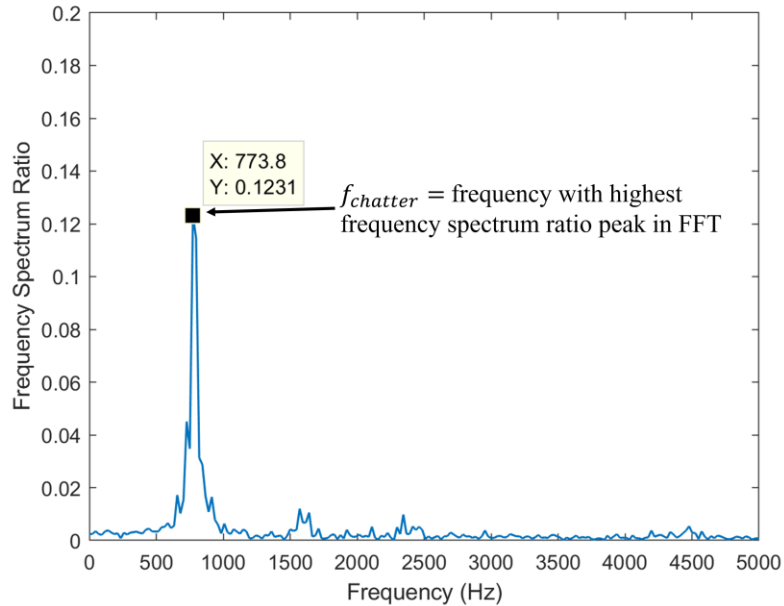
Figure 21. Chatter frequency at time instance t_s. 4140 Steel, 1000 RPM, 2.5 mm DOC and 132 mm/min feed rate.

Knowing the chatter frequency, $f_{chatter}$, we define $P_{chatter}(t)$ as a time-series of the chatter frequency's magnitude divided by $2^{15} \cdot N$ evaluated at each discrete time instance, where 2^{15} is the maximum possible value of the 16-bit digital audio signal and N is the bin size of the FFT. Similarly, $P_{notch}(t)$ is defined as the time series of the maximum amplitude of notch filtered signals from each bin of the FFT divided by $2^{15} \cdot N$. Both

$P_{notch}(t)$ and $P_{chatter}(t)$ are identified in this work as frequency spectrum ratio, as they are the ratios of the magnitude of particular frequency content and its maximum possible value. In the following section, the methods for pre-processing the audio signal to isolate $P_{notch}(t)$ and $P_{chatter}(t)$ are presented.

4.3.3   Pre-processing of Audio Signals

As mentioned in Section 4.3.1, the audio signal acquired via the microphone is a combination of the periodic spindle rotation frequency, the tooth passing frequency and its harmonics, and any potential noise in the production environment. When chatter occurs, a dominant chatter frequency emerges in the frequency domain. In order to detect chatter as soon as possible, the periodic component of the cutting signal must be first removed using notch filters. Although slot end milling operations are not significantly affected by the spindle frequency and its harmonics, as in a peripheral end milling operation, removal of the periodic components is crucial to reduce the impact of large-amplitude periodic signals on the ability to detect the chatter frequency. In this work, twelve harmonics of the spindle rotation frequency (up to three times the tooth passing frequency) are removed by notch filters designed to isolate the chatter frequency. When the chatter frequency is unknown, the notch filters provide the best isolation of the chatter frequency for setting the chatter detection threshold.

However, because the audio signals acquired from the condenser microphone are subject to interference from noise in a production environment, a chatter frequency specific band-pass filter is applied to isolate the known chatter frequency identified from demonstrations.
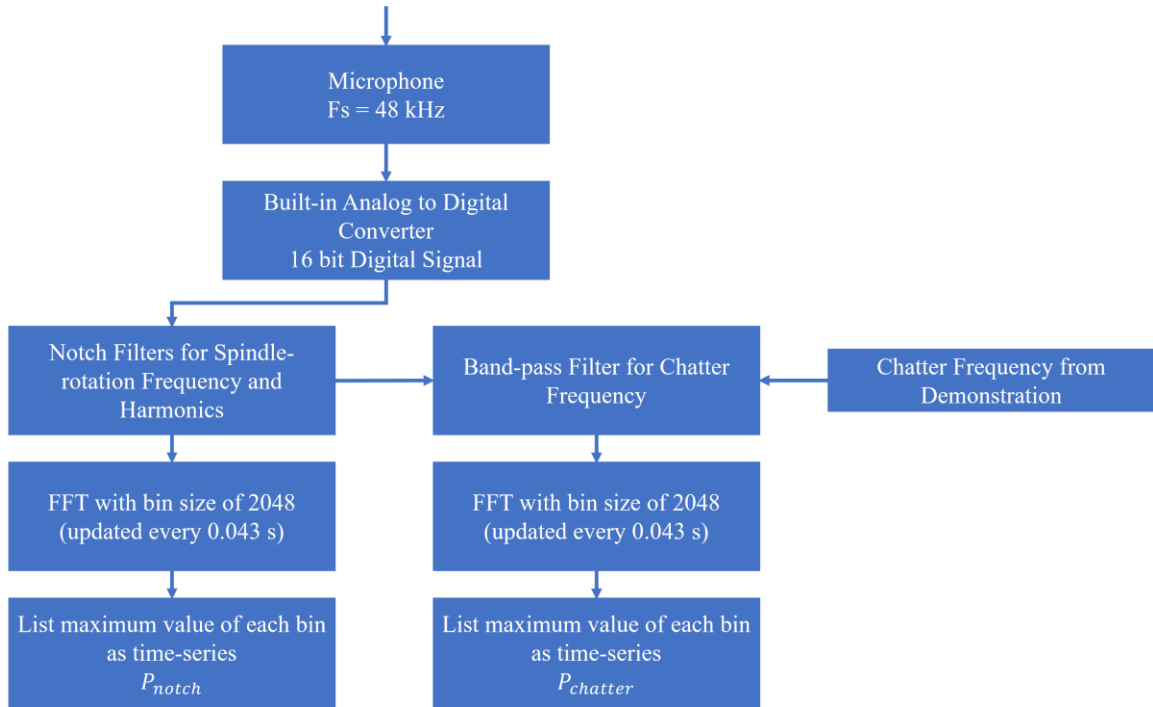
Figure 22. Signal pre-processing approach.

Figure 22 summarizes the signal pre-processing workflow of the LSP method, where an additional band-pass filter serves to isolate the known chatter frequency from previous demonstrations. If a previous demonstration does not exist, the band-pass filter is by-passed. The pre-processed data is then used to establish a chatter detection threshold. The next section presents details of the process.

4.3.4 LSP Method for Chatter Detection—Approach Overview

The LSP method seeks to learn a chatter detection threshold P_{th}. This section provides an overview of the training and testing phases of the LSP method for a milling process. Note that the LSP is developed for continuous monitoring of the milling process, which means that in the real application the difference between training and testing is decided by whether a human operator demonstration is provided or not.
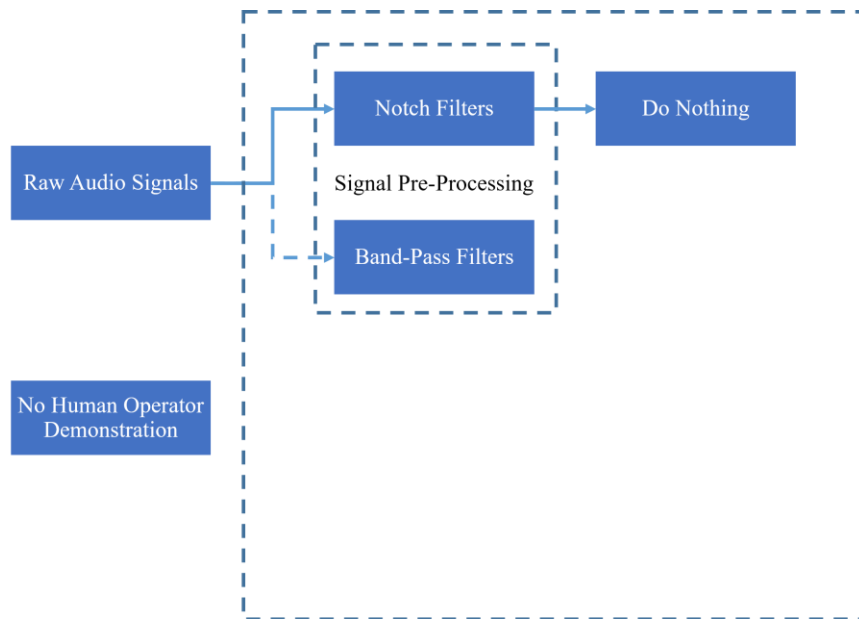
Figure 23. Initialization of the LSP method.

Figure 23 shows the initial state of the LSP algorithm without any prior of the chatter detection threshold, P_{th}, and chatter frequency, $f_{chatter}$. In addition, the figure shows that without human demonstrations, the band-pass filter is by-passed. In this case, no chatter detection threshold is established from demonstration, which means that the machine does not detect chatter. To draw an analogy, in its initialization state the machine is similar to a novice operator who is clueless about chatter.
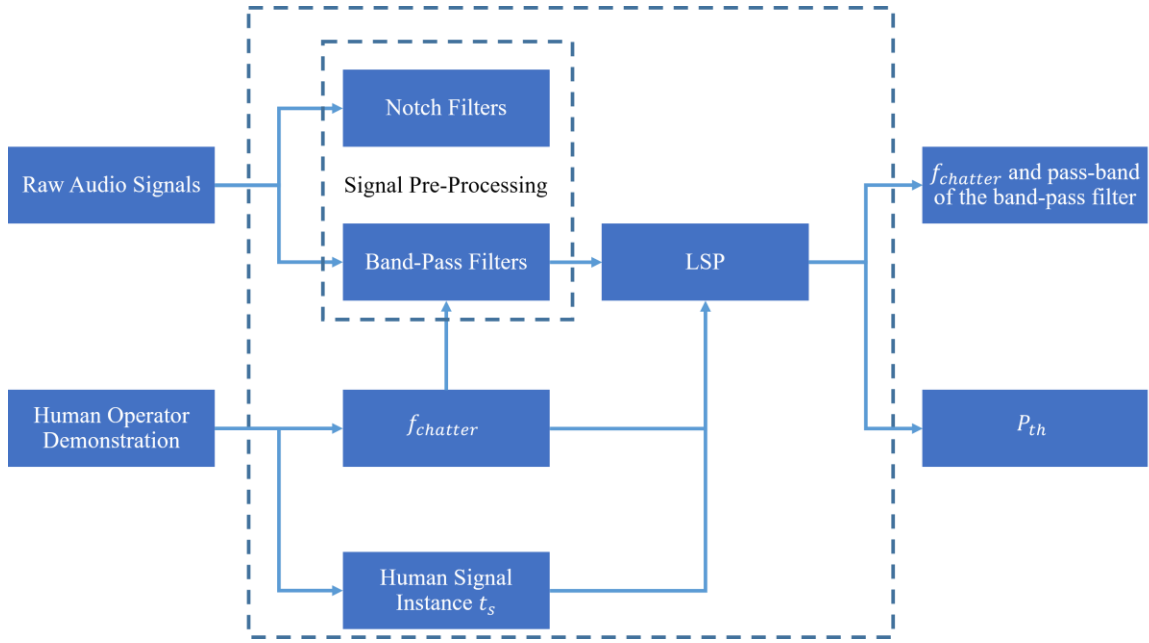
Figure 24. Training the LSP with a human operator demonstration.

Figure 24 shows how the LSP is trained by a single human operator demonstration. The inputs to the algorithm are the digital audio signals acquired from the microphone and the time instance the human signals chatter, t_s, from the human demonstration, and the outputs are $f_{chatter}$ and P_{th}. Note that to obtain any information from the demonstration, a time period of signals preceding the demonstration must be recorded and dynamically updated such that the chatter frequency and the chatter detection threshold can be identified from the demonstration. An example of chatter is demonstrated to the machine by the operator, and the chatter frequency $f_{chatter}$ is identified as the frequency with the highest amplitude at t_s. The pass-band of the band-pass filter is set to $f_{chatter} \pm 100\ Hz$. The recorded data preceding the demonstration is filtered by the band-pass filter, resulting in $P_{chatter}(t)$ as introduced in Figure 22 and P_{th} is established through the LSP.
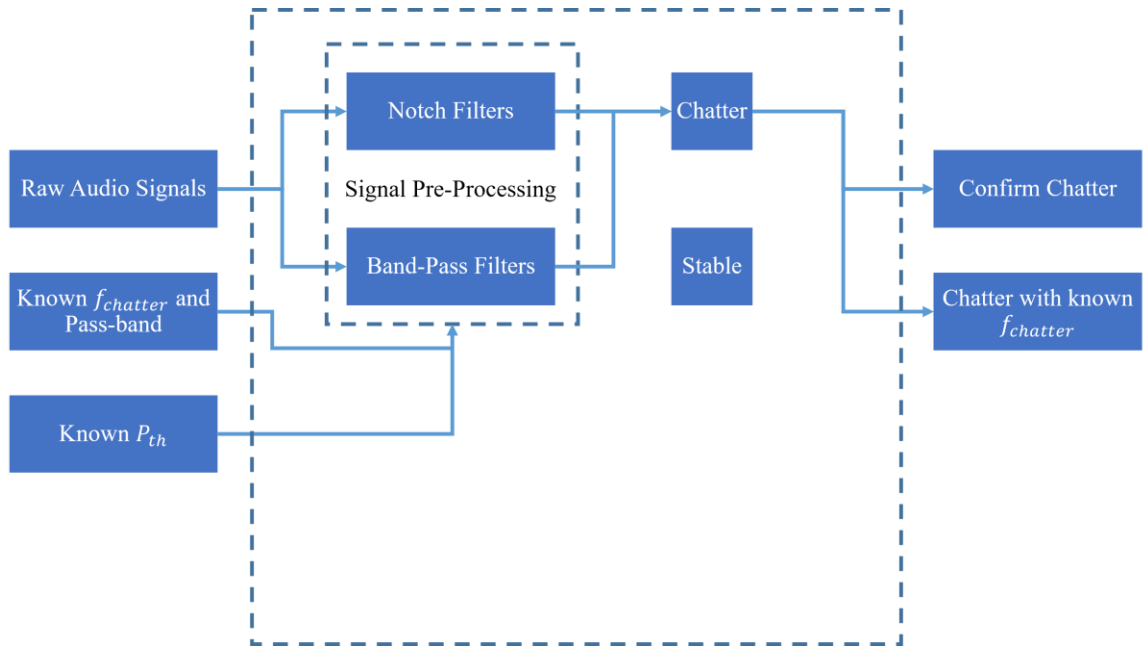
Figure 25. Implementing $\boldsymbol{P_{th}}$ – confirmation of chatter.

In the scenario illustrated in Figure 25, a chatter detection threshold has been established from a previous demonstration, and chatter is confirmed when the data processed through the band-pass filter $P_{chatter}(t)$ exceed P_{th}. Although in this case the notch filtered data $P_{notch}(t)$ is redundant for chatter detection, it is useful for the scenario depicted in Figure 26 where $P_{notch}(t)$ produces different results than $P_{chatter}(t)$.
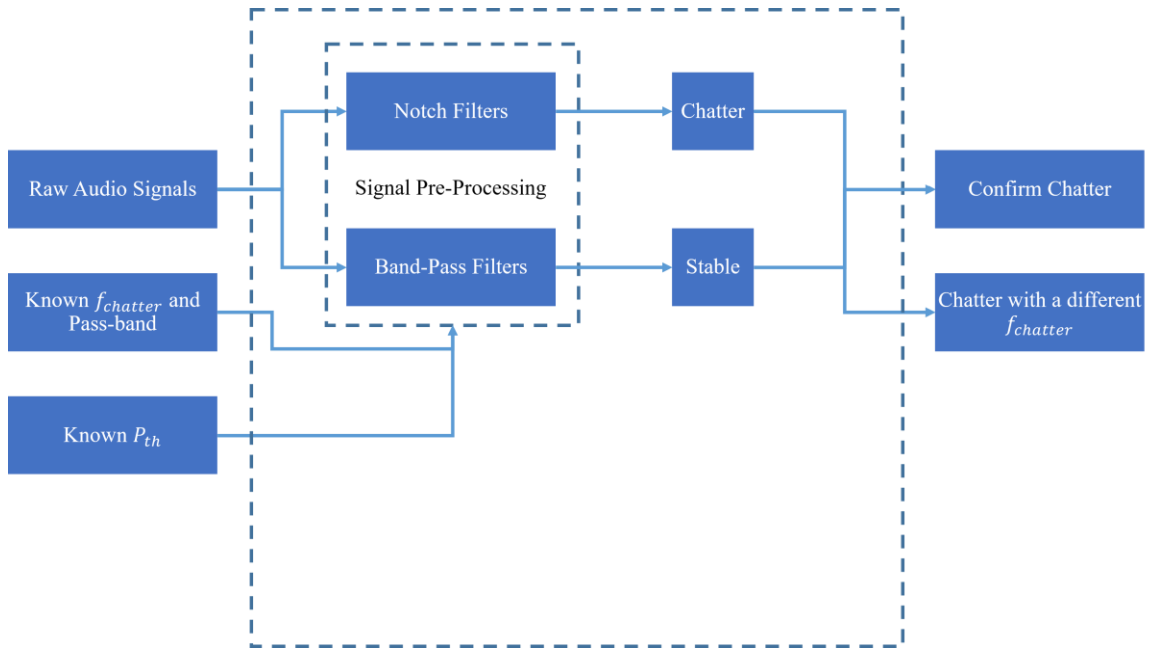
Figure 26. Implementing P_{th} – disagreement in chatter detection.

In the scenario depicted in Figure 26, $P_{notch}(t)$ exceeds P_{th} but $P_{chatter}(t)$ from the bandpass filter does not, which indicates that the chatter frequency detected from the data is different from the known chatter frequency. Note that in this scenario, because the notch filtered data is more prone to noise, it is possible that spikes of noise can accidentally exceed P_{th}. It is then important to make sure that $P_{notch}(t)$ correctly reflects chatter. The adjustment to $P_{th}(t)$ based on $P_{notch}(t)$ can be extensive and is not discussed in this work; instead, the same P_{th} obtained from $P_{chatter}(t)$ is also applied to $P_{notch}(t)$. Finally, while it is likely for $P_{notch}(t)$ to detect chatter when $P_{chatter}(t)$ indicates stable cutting, theoretically it cannot be the other way around because the data is filtered through the band-pass filter after the notch filters.
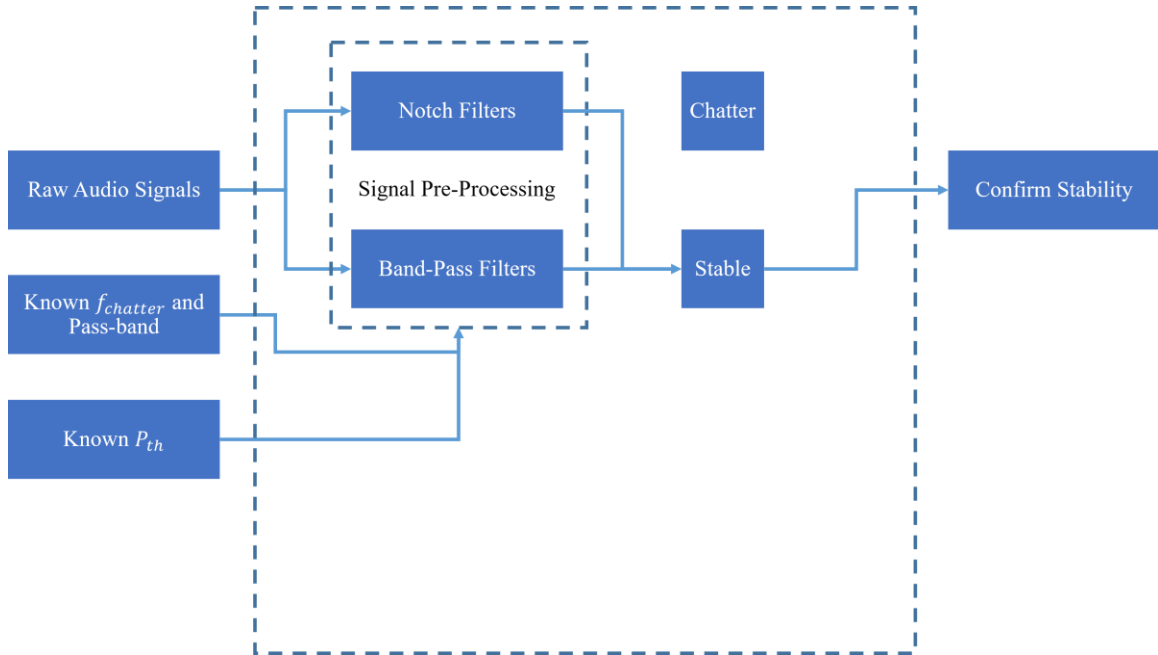
Figure 27. Implementing P_{th} – confirmation of stable cutting condition.

Similarly, Figure 27 shows that if both $P_{notch}(t)$ and $P_{chatter}(t)$ indicate stable cutting, the stability of the cutting process is confirmed. The next section presents a probabilistic analysis that serves as a basis for the LSP algorithm to determine P_{th} from $P_{chatter}(t)$.

4.3.5 Probabilistic Analysis of the Time Instance of Chatter Mark Appearance

With the chatter frequency known, the threshold P_{th} is defined as the threshold for chatter detection. When $P_{chatter}(t)$ exceeds P_{th}, the algorithm classifies the cutting process as unstable due to chatter. An ideal threshold value of P_{th} is a value of $P_{chatter}(t)$ obtained a short time period before the time instance t_a when chatter marks first appear on the workpiece.
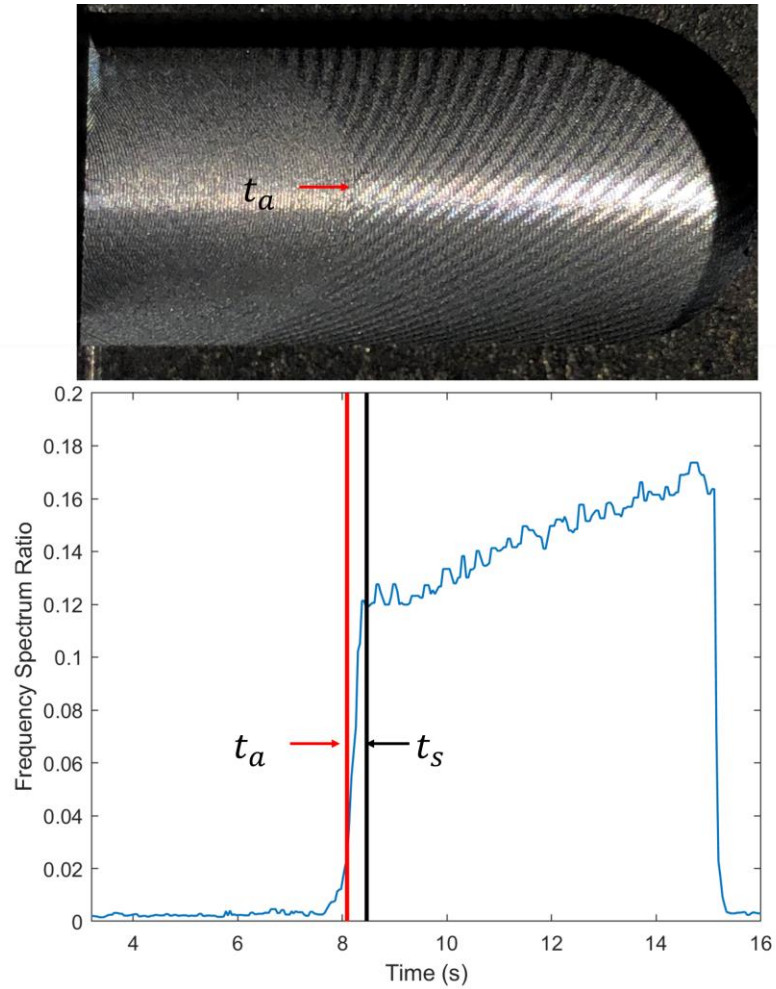
Figure 28. Chatter mark time instance t_a relative to time instance the operator signals chatter t_s; the time series shown is $P_{chatter}(t)$.

Figure 28 shows the correspondence between chatter marks on the workpiece at the chatter mark instance t_a and the frequency spectrum ratio.

In the experiment, the distance between the first chatter mark and the start of cut at the workpiece edge is measured as shown in Figure 29, and t_a is calculated from the feed rate of the milling process and the time instance of tool entry into the workpiece.

Figure 29. Caliper measurement of the chatter mark from point of tool entry into the workpiece.

However, in an actual production setting, t_a is hard to obtain in-process, and even if out-of-process measurement is possible, interacting cutting tool paths can make it hard to determine the exact value of t_a.

A better alternative is to evaluate the time instance t_c when the human operator first observes chatter, as defined in Chapter 3. While t_c cannot be measured precisely because of the variable reaction times of human operators, the probability of its occurrence in time can be estimated via the tracing distribution f_{trace} derived from perception mapping as follows:

$$Pr[t_s - \beta \le t_c \le t_s - \alpha] = \int_{t_s - \beta}^{t_s - \alpha} f_{trace}(\hat{t_c}) \, d\hat{t_c}, \qquad a, b > 0 \qquad (9)$$

Note that the probability distribution function is continuous; however, the audio signal acquired from the microphone is discretized as a time-series. The above equation is

61

modified to calculate the discrete probability of t_c. To do this, we define a finite sequence a where $a_1, \cdots, a_n$ are all discrete time instances separated by $\frac{N}{f_s}$, where N is the FFT bin size and f_s is the sampling frequency, and bounded as $t_s - \bar{r} - 6 \cdot s_{\bar{r}} \leq a_i \leq t_s$. The hypothesis underlying this approach is that t_s is the time instance when the probability of chatter is 100%, which is later confirmed by the human operator's chatter detection accuracy. Six times $s_{\bar{r}}$ represents six times the standard deviation of $\bar{r}$ based on the human reaction time distribution experiment discussed in Chapter 3.
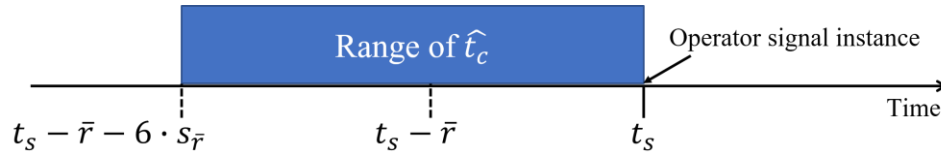
Figure 30. Discrete probability of $\boldsymbol{t_c}$.

As shown in Figure 30, the probability of t_c falling within the specified range of n data instances can be approximated as 1. Then the equation 8 is transformed into:

$$Pr(a_1 \leq \hat{t}_c \leq a_i) = \sum_{i=1}^{j} Pr(\hat{t}_c = a_i), \quad j \in 1, 2, \dots, n \tag{10}$$

In other words, as time progresses and i increases, there is a greater likelihood of the estimated chatter observation time instance to be within the range $a_1 \leq \hat{t}_c \leq a_i$, and the probability for the chatter observation instance $\hat{t}_c$ to be in the range of $a_1 \leq \hat{t}_c \leq a_n$ is approximated as 1. If we assume that $\hat{t}_c$ has uniform probability at each discrete time instance:

$$Pr(\hat{t_c} = a_i) = Pr(\hat{t_c} = a_j), \quad i, j \in 1, 2, \cdots, n \tag{11}$$

Based on the above, Equation (10) can be further simplified to:

$$Pr(a_1 \le \hat{t_c} \le a_i) = \frac{i}{n}, \quad i \in 1, 2 \dots, n \tag{12}$$

In summary, the probability of t_c to be in the range of $a_1 \le \hat{t_c} \le a_i$ is reduced to Equation (12), which states that the likelihood of the operator observing chatter at any time domain instance a_i is equal.

4.3.6 Defining the Chatter Detection Threshold via Cumulative Chatter Probability

Recall that $P_{chatter}(t)$ is defined as a time-series of the frequency spectrum ratios derived from the FFT of the audio signal, and chatter is detected when the value of $P_{chatter}(t)$ exceeds P_{th}, which ideally equals a value of $P_{chatter}(t)$ that is smaller than $P_{chatter}(t_a)$ so that chatter is detected before chatter marks appear on the workpiece:

$$P_{th} < P_{chatter}(t_a) \tag{13}$$

However, t_a is almost always unknown till after the machining operation, and therefore the threshold can be set as the value of $P_{chatter}(t)$ corresponding to the estimated chatter observation time instance $\hat{t_c}$ when the human operator first observes chatter:

$$P_{th} = P_{chatter}(\hat{t_c}) \tag{14}$$

However, the only information known from Equation (11) is that there is equal likelihood for t_c to fall on any data point inside the sequence a, and the cumulative probability for t_c to be in the range $a_1 \le \hat{t_c} \le a_i$ increases linearly from 0 to 1. Thus, we define a decision boundary θ, where the cumulative probability of chatter equals a preset probability threshold k:

$$C = \min\{i | Pr(a_1 \leq \hat{t}_c \leq a_i) > k; \ i \in 1, 2 \dots, n\}, \qquad 0 \leq k \leq 1$$

$$a_{C-1} < \theta < a_C$$

(15)

where C is the index of the first time instance of a for which the cumulative probability exceeds k and θ is defined as the time instance interpolate between a_C and a_{C-1} such that $Pr(a_1 \leq \hat{t}_c \leq \theta) = k$. The value of k can be deterministically set or optimized based on experimental data to obtain a desired false detection rate, and the chatter detection threshold becomes:

$$P_{th} = P_{chatter}(\theta)$$

(16)

In the next sections, chatter detection thresholds are obtained from milling experiments, and the on-line learning performance of the LSP method is evaluated and discussed.

4.4 On-line Learning Experiments

In this section, the quality of a human operator's demonstrations is analyzed, and the performance of the LSP method in setting the chatter detection threshold is assessed.

4.4.1 Human Demonstrations

A set of milling experiments was conducted on a block of AISI 4140 steel to capture the human operator's demonstrations. The tests were designed such that the human operator would encounter both stable and unstable (chatter) cutting conditions. When the human operator judged a process to be stable, the operator was instructed to do nothing during the cutting experiment; conversely, when the human operator detected chatter, the operator was instructed to depress a push button switch that enabled the time instance of

chatter signaled by the operator, t_s, to be recorded. Regardless of the human operator's reactions, the cutting experiments were not interrupted, which enabled post-mortem evaluation of the experimental evidence of chatter marks on the workpiece.

There are two evaluation metrics for the human demonstrations, namely, a human operator's chatter recognition accuracy and his/her chatter recognition speed. The chatter recognition accuracy refers to the ability of the human operator to correctly differentiate between a stable and an unstable process. A human operator's demonstration is accurate if the operator correctly signals chatter for a process that yields visual evidence of chatter marks on the workpiece, or if the operator does not signal chatter for a stable process that does not produce any chatter marks on the workpiece. The chatter recognition speed refers to whether a human operator observes chatter before or after chatter marks appear on the workpiece.
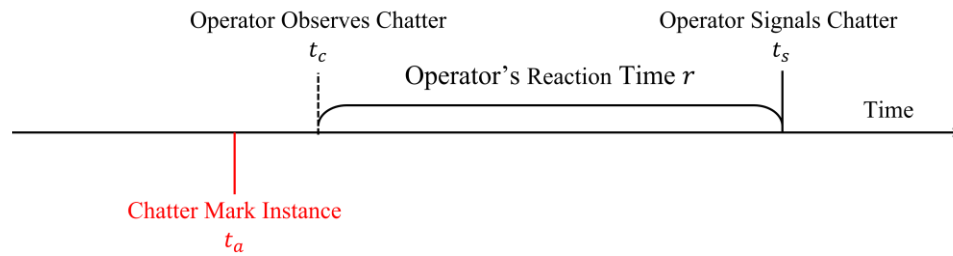
Figure 31. Chatter mark instance in a milling experiment.

The speed of chatter recognition is more difficult to determine because the time instance the operator signals chatter, t_s, is distinct from the time instance the operator observes chatter, t_c. While t_s is almost always too late compared to t_a, it does not represent the true instance of human chatter observation, t_c. Therefore, the operator's chatter recognition speed is defined as $t_a - t_c$. If $t_a - t_c > 0$, the human operator observes chatter before chatter marks appear on the workpiece; if $t_a - t_c < 0$, the human operator observes

chatter after the chatter marks appear on the workpiece. Because the exact value of t_c in each milling experiment is unknown, a direct calculation of $t_a - t_c$ is not possible. However, recall that from Equation (5) in 3.5 the sample mean reaction time $\bar{r}$ of the human operator was determined from synthetically generated audio signals that emulate chatter. Similarly, the difference, d, between the time instance of the operator's signal and the time instance chatter mark appears on the workpiece can be calculated as:

$$d = t_s - t_a \tag{17}$$

Therefore, $t_a - t_c$ can be rewritten as:

$$t_a - t_c = (t_s - t_c) - (t_s - t_a) = r - d \tag{18}$$

Equation (18) represents the difference between the chatter mark time instance and the time instance of human observation in a single demonstration. Again, r for each demonstration is unknown because t_c is not measurable precisely, but from Equation (5) the sample mean $\bar{r}$ can be obtained from the operator's reaction to synthetically generated chatter sound, which hypothetically is an estimate of the actual mean reaction time of the human operator to detect chatter in a milling process. Therefore, the following equation can be deduced:

$$\bar{d} = \frac{1}{q}\sum_{i=1}^{q} d = \frac{1}{q}\sum_{i=1}^{q}\left(t_{s,i} - t_{a,i}\right) \tag{19}$$

$$\bar{t_a} - \bar{t_c} \cong \bar{r} - \bar{d}$$

where $\bar{d}$ is the sample mean difference between human signal instances and chatter mark instances, and q is the number of samples. Therefore, if $\bar{r} - \bar{d} > 0$, the human operator observes chatter before chatter marks appear on the workpiece; if $\bar{r} - \bar{d} < 0$, the human operator observes chatter after chatter marks appear on the workpiece. The chatter

recognition speed (i.e. whether the human operator recognizes chatter early or late) defined by $\bar{r} - \bar{d}$ is thus obtained.
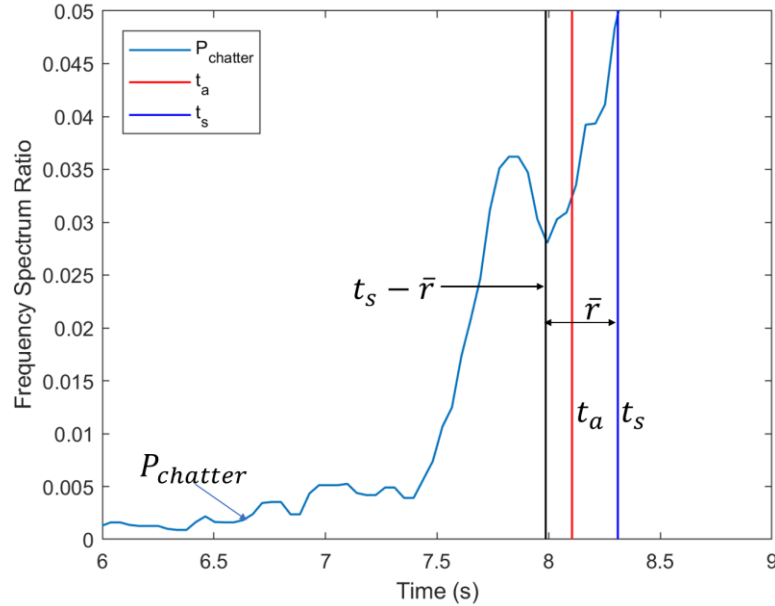
Figure 32. Example of human operator's recognition of chatter before chatter marks appear on the workpiece.

The signal time instance t_s of each milling experiment was recorded. The chatter mark was measured from the point of tool entry into the workpiece. Using the operator's mean and standard deviation of reaction times described in Chapter 3, the finite sequence a was determined for each demonstration. Recall that a is bounded by the mean reaction time minus six times the standard deviation of the reaction time and the human signal time instance. Table 4 lists the values of t_a, a_1 and a_n of nine unstable cuts and five stable cuts.

Table 4. Experimental results of human demonstrations of chatter on 4140 steel workpiece.

Operator	Human Observation	Spindle Speed (RPM)	Axial DOC (mm)	t_s (s)	t_c (s)	t_a (s)	a_1 (s)	a_n (s)
A	Chatter	750	3.5	7.591	7.285	**7.790**	**6.919**	**7.591**
A	Chatter	2000	2	3.917	3.611	3.547	3.245	3.917
A	Chatter	2000	2.5	3.785	3.479	3.110	3.113	3.785
B	Chatter	750	3.5	8.310	7.987	8.104	7.513	8.310
B	Chatter	2000	3	3.457	3.134	3.010	2.660	3.457
C	Chatter	1000	2.5	8.464	8.069	8.088	7.655	8.464
C	Chatter	1000	3	6.690	6.295	6.260	5.881	6.690
C	Chatter	1750	2	4.489	4.094	4.005	3.680	4.489
C	Chatter	1750	2.5	4.076	3.681	3.758	3.267	4.076
A	Stable	750	3	No chatter mark on the workpiece				
A	Stable	2000	1.5	No chatter mark on the workpiece				
B	Stable	750	3	No chatter mark on the workpiece				
B	Stable	2000	2	No chatter mark on the workpiece				
C	Stable	1000	2	No chatter mark on the workpiece				

Demonstrations by three human operators were collected through fourteen cutting experiments listed in Table 5. The results show that the human operators were able to accurately distinguish chatter from stable cutting in all fourteen experiments. Among the chatter experiments, eight out of nine demonstrations show that the chatter mark instance t_a falls between a_1 and a_n, with the exception of the first experiment where the operator signaled chatter even before the chatter mark appeared on the workpiece. The mean difference between the human signal instance and the chatter mark instance $\bar{d}$ is 0.407,

which is higher than all three mean reaction times of the human operators (0.306, 0.323, and 0.395). Therefore, $\bar{r} - \bar{d} < 0$, which implies that on average the operator reacted to chatter shortly after the chatter mark appeared on the workpiece. As a result, although the data confirms that the human operators could accurately distinguish between stable and unstable cutting, the mean chatter recognition speed was too slow since the chatter marks had already appeared on the workpiece. Consequently, it is important to select a probability threshold k that accounts for the late reaction the human operator to establish a chatter detection threshold that enables chatter detection before chatter marks appear on the workpiece.

4.4.2 Applying the LSP – Results and Discussion

The LSP method presented in the previous section was used to learn the chatter detection thresholds from each demonstration. Because a value of 0.5 for the probability threshold k implies that the average human chatter observation time instance is the same as the chatter mark time instance, k is set to 0.4 to account for the slower chatter recognition speed of human operators. For each chatter demonstration, Table 5 lists the chatter detection threshold P_{th} obtained through application of the LSP method, the amplitudes of the chatter frequency at the chatter mark instance $P_{chatter}(t_a)$, and the maximum signal amplitude of the chatter frequency P_{max}. For stable cutting experiments, the P_{th} and $P_{chatter}(t_a)$ are not applicable since chatter did not occur in these tests; P_{max} refers to the highest amplitude of $P_{chatter}(t)$ in a stable cutting operation.

Table 5. Chatter detection thresholds obtained from demonstrations through LSP in slot milling experiments on 4140 steel workpiece.

Operator	Human Observation	Spindle Speed (RPM)	Axial DOC (mm)	P_{th}	$P_{chatter}(t_a)$	$P_{max,chatter}$	$P_{max,stable}$
A	Chatter	750	3.5	0.020	0.038	0.143	N/A
A	Chatter	2000	2	0.038	0.056	0.092	N/A
A	Chatter	2000	2.5	0.209	0.039	0.451	N/A
B	Chatter	750	3.5	0.036	0.032	0.170	N/A
B	Chatter	2000	3	0.026	0.036	0.386	N/A
C	Chatter	1000	2.5	0.013	0.024	0.174	N/A
C	Chatter	1000	3	0.019	0.033	0.191	N/A
C	Chatter	1750	2	0.035	0.038	0.044	N/A
C	Chatter	1750	2.5	0.014	0.085	0.456	N/A
A	Stable	750	3	N/A			0.013
A	Stable	2000	1.5	N/A			0.003
B	Stable	750	3	N/A			0.021
B	Stable	2000	2	N/A			0.016
C	Stable	1000	2	N/A			0.004

To evaluate the effectiveness of the P_{th} values determined from the demonstrations, each P_{th} is first compared with $P_{chatter}(t_a)$ and $P_{max,chatter}$ for nine chatter demonstrations, and then P_{th} is compared with $P_{max,stable}$ for the stable cuts. Four possible outcomes are possible: early chatter detection, late chatter detection, false positive, and false negative.
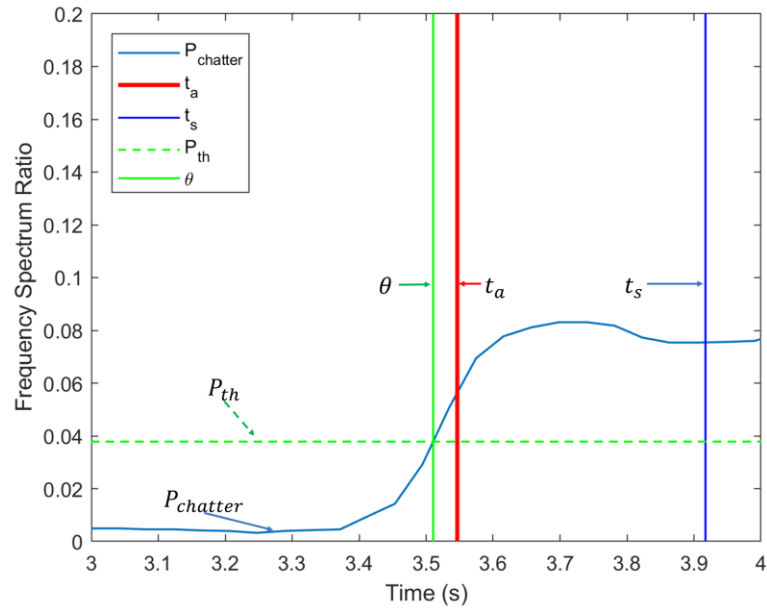
Figure 33. Early chatter detection example.

When $P_{th} < P_{chatter}(t_a)$ chatter is detected before chatter marks appear on the workpiece, which is classified as early chatter detection; when $P_{th} > P_{chatter}(t_a)$, chatter is detected after chatter marks appear on the workpiece surface, which results in late chatter detection. Figure 33 shows the time instance at the intersection of the chatter detection threshold P_{th} and the frequency spectrum ratio is earlier than the chatter mark instance t_a, whereas Figure 34 shows an example where the time instance corresponding to the chatter detection threshold occurs later than the chatter mark instance. Note that both early and late chatter detection cases indicate that chatter is detected during the milling operation.
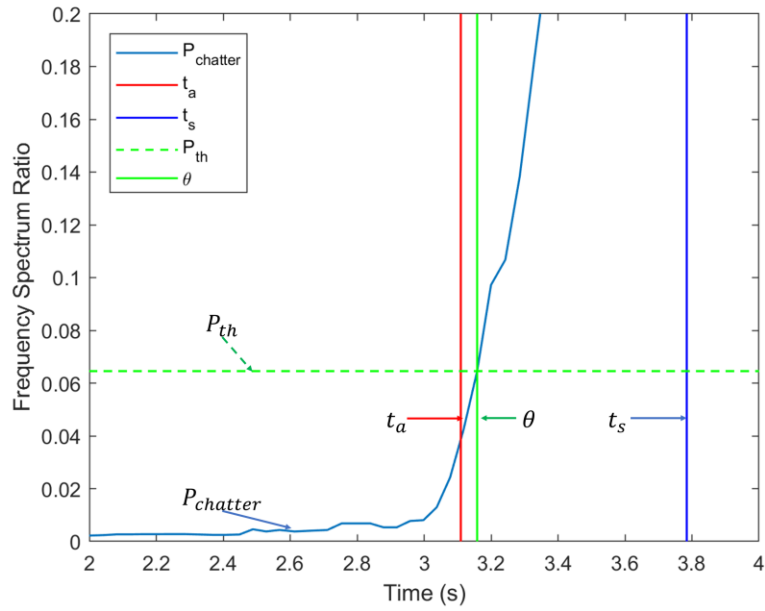
Figure 34. Late chatter detection example.
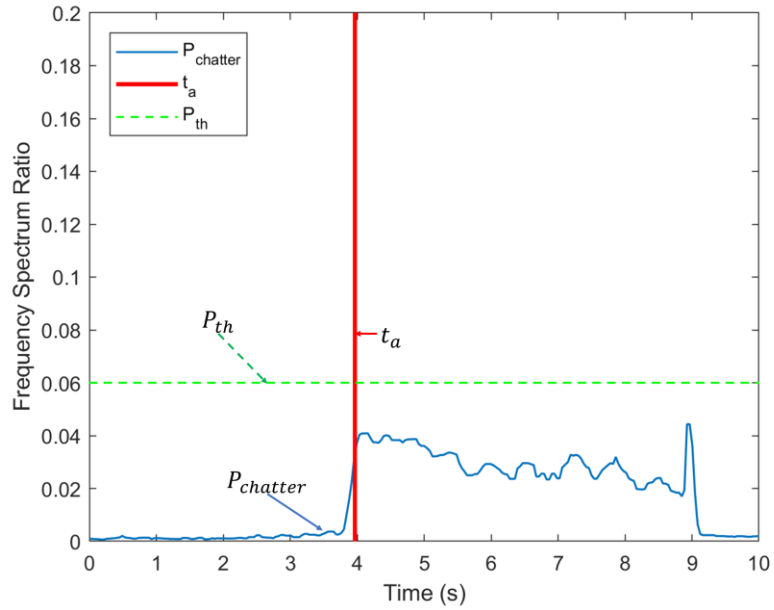
Figure 35. Example of a false negative chatter detection.

False negative means that the milling process is falsely classified as a stable process when $P_{th} > P_{max,chatter}$. Figure 35 shows an example of a false negative chatter detection result. In this case, chatter marks were found on the workpiece, but the chatter detection threshold was set too high from the demonstration. Note that the y-axis of the figure is scaled differently to display P_{th}. On the other hand, a false positive result occurs when a stable milling process is falsely classified as an unstable process, which implies that $P_{th} < P_{max,stable}$. Figure 36 illustrates a false positive result with the stable cutting signal crossing the chatter detection threshold multiple times.
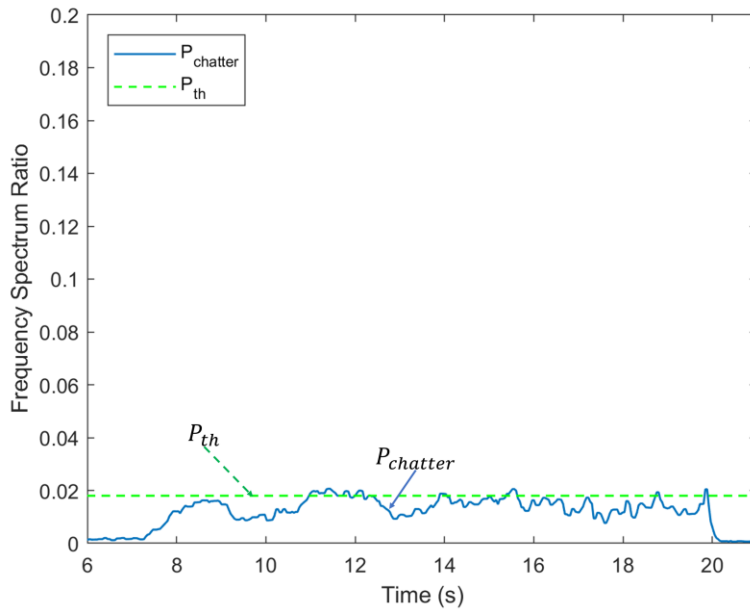
Figure 36. Example of a false positive case.

False positives and false negatives are detection inaccuracies that should ideally be eliminated. For the chatter detection approach described in this thesis, a trade-off between a false positive classification and the detection speed must be considered. A lower detection

threshold typically implies an earlier detection time but tends to yield a higher rate of false positives.

Table 6. Average performance of chatter detection thresholds.

	Occurrence	Total Tests	Percentage
Overall Accuracy	114	126	90.5%
Early Detection	60	81	74.1%
Late Detection	15	81	18.5%
False Negative	6	81	7.4%
False Positive	6	45	13.3%

Table 6 shows the overall performance of the chatter detection thresholds (in Table 5) determined from the human demonstration experiments and the LSP approach developed in this chapter. A total of 126 combinations of thresholds were evaluated. Out of 126 evaluations, 113 detected chatter or stable cutting conditions accurately without reporting false negatives or false positives, which corresponds to an overall accuracy of 89.7%. Out of 81 chatter tests, 62 were detected early, 13 were late detections, and 6 were false negatives that corresponded to actual chatter instances that were undetected by the thresholds. Finally, 7 out of 45 stable cutting tests resulted in false positives. Although the results are not perfect, considering that each chatter detection threshold was learned from a single demonstration, the results are very promising and indicate an acceptable performance of the LSP as a classifier of the cutting process stability.

It is also of interest to compare the performance of the LSP with the human's ability to detect chatter. While specific corrective actions to suppress chatter are not addressed in

this thesis, potential corrective actions can be also learned by the LSP and applied immediately after the frequency spectrum ratio $P_{chatter}(t)$ crosses the chatter detection threshold P_{th}. Therefore, because the human signal instance t_s is the earliest time instance the human operator is able to apply any corrective action to the cutting process, a comparison of P_{th} and $P_{chatter}(t_s)$ is shown in Table 7 to illustrate that LSP is capable of initiating control actions at an earlier time instance than the human operator.

Table 7. Average performance of $\boldsymbol{P_{th}}$ and $\boldsymbol{P_{chatter}(t_s)}$.

	Performance of P_{th}	Performance of $P_{chatter}(t_s)$
Overall Accuracy	90.5%	78.6%
Early Detection	74.1%	16.0%
Late Detection	18.5%	50.6%
False Negative	7.4%	33.3%
False Positive	13.3%	0.0%

Table 7 indicates that the LSP yields an improved detection speed with an early detection rate of 74.1% compared to 16.0% for $P_{chatter}(t_s)$. It also shows that the overall accuracy of $P_{chatter}(t_s)$ (78.6%) is lower, with an average false negative rate of 33.3% that is unacceptably high. Because a human operator actually recognizes chatter at an earlier time instance t_c, the accuracy of $P_{chatter}(t_s)$ is not equivalent to the human operator's accuracy in chatter recognition; however, for a quantitative comparison of the LSP with the human operator, Table 7 confirms that the LSP outperforms the human operator by eliminating the natural delay in the human's reaction time.

However, for practical applications in a production setting, the false detection rates must be further reduced. Note that the average performance does not represent the performance of any one threshold. Ideally the variance in performance should be small among different thresholds. A breakdown is shown in Table 8 to illustrate the performance of the different chatter detection thresholds.

Table 8. Breakdown of the performance of different chatter detection thresholds.

P_{th}	Early Detection Rate	Late Detection Rate	False Negative Rate	False Positive Rate	Detection Accuracy
0.020	100.0%	0.0%	0.0%	20.0%	92.9%
0.038	55.6%	44.4%	0.0%	0.0%	100.0%
0.209	0.0%	33.3%	66.7%	0.0%	**57.1%**
0.036	55.6%	44.4%	0.0%	0.0%	100.0%
0.026	88.9%	11.1%	0.0%	0.0%	100.0%
0.013	100.0%	0.0%	0.0%	40.0%	**78.6%**
0.019	100.0%	0.0%	0.0%	20.0%	92.9%
0.035	66.7%	33.3%	0.0%	0.0%	100.0%
0.014	100.0%	0.0%	0.0%	40.0%	85.7%

It can be seen in Table 8 that the highest chatter detection accuracy is 100%, and the lowest detection accuracy is 57%. Although most thresholds resulted in detection accuracies above 85%, two thresholds performed poorly with detection accuracies below 80%. Such high variance in performance must be addressed, since the ideal objective is to learn an effective chatter detection threshold from a single demonstration.

Specifically, the highest chatter detection threshold value, 0.209, is one order of magnitude greater than the other thresholds, and contributed to all instances of false

negatives. Figure 37 is plotted to investigate the cause of the high chatter detection threshold resulting from the corresponding human demonstration. Recall that on average the difference between $\bar{r}$ and $\bar{d}$ is smaller than illustrated in the figure, which means that it took much longer than the average time for the human operator to react to chatter during this milling test.
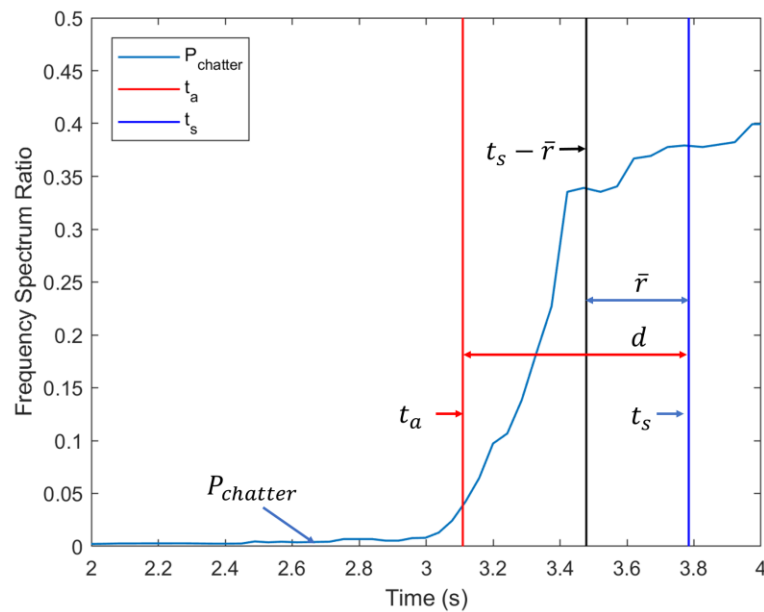
Figure 37. Chatter mark instance t_a and human signal instance t_s for a sample demonstration by Operator A; spindle speed of 2000 RPM, axial depth of cut of 2.5 mm, and 264 mm/min feed rate on a 4140 steel workpiece.

Similarly, Figure 38 indicates that the lowest chatter detection threshold was obtained from a demonstration where the operator's reaction time to chatter was faster than average. It appears that the variance in the operator's reaction time to chatter contributed to the variance in threshold values obtained from the demonstrations. In a scenario where multiple demonstrations are available, the impact of reaction time variance on the chatter

77

detection threshold can be reduced by taking the mean of the obtained thresholds. However, in a practical manufacturing application where damage to the workpiece must be minimized, a variance mitigation scheme must be developed to account for possible early or late reaction of human operators.
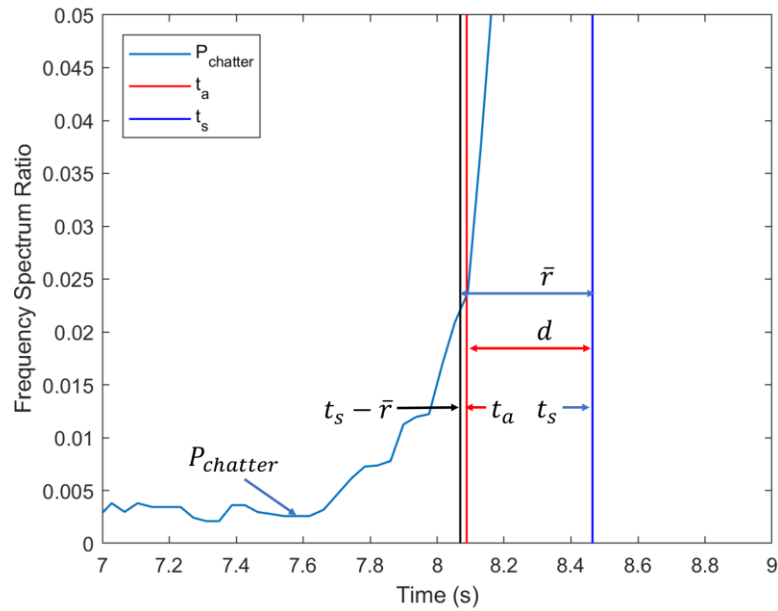
Figure 38. Chatter mark instance t_a and human signal instance t_s for a demonstration by Operator C; spindle speed of 1000 RPM, axial depth of cut of 2.5 mm, and feed rate of 132 min/mm on a 4140 steel workpiece.

4.5   Summary

In this chapter, the concept of a "learnable skill primitive" (LSP) for chatter detection was introduced. By integrating the human operator's reaction time delay with the audio signals collected from a condenser microphone, the LSP approach was shown to be capable of learning the chatter detection threshold from the operator's demonstrations with reasonable detection speed and accuracy such that in most cases chatter was detected early and accurately. Specifically, an acoustic microphone and a frequency spectrum based on-line chatter detection method were selected for developing the LSP because an analogy can be drawn between a human's hearing ability and the signals acquired from the microphone and analyzed via frequency spectrum analysis, where the frequency represents the pitch of the sound, and the amplitude is the volume. The LSP utilizes a probabilistic analysis of the human's reaction time and implements a threshold based on the preset likelihood of the occurrence of chatter. It is clear from the experimental data that human operators can clearly distinguish between an unstable (chatter) and a stable cutting process. However, the results show that the average speed of chatter recognition is too slow resulting in human operators detecting chatter after chatter marks appear on the workpiece. The LSP learns from the human demonstration and sets the detection threshold accordingly, but due to high variance in the human operator's reaction time, the chatter detection thresholds can be either too high or too low such that false negatives and false positives are obtained. Based on the experimental data, the average performance of the LSP yielded an overall chatter detection accuracy of 90.5% from a single demonstration, with most tests exhibiting early chatter detection prior to the appearance of chatter marks on the workpiece surface; however, the lowest detection accuracy was 57.1%, which is unacceptable for

manufacturing applications where only a single demonstration is possible. Therefore, in the next chapter, a human reaction variance mitigation strategy is developed to reduce the variance in the chatter detection thresholds. In addition, the robustness of the LSP to variations in the workpiece materials, cutting conditions (e.g., radial immersion), and cutting direction are evaluated.

CHAPTER 5 VARIANCE MITIGATION STRATEGY FOR THE LEARNABLE SKILL PRIMITIVE

5.1   Introduction

An in-depth introduction to the "Learnable Skill Primitive" (LSP) for chatter detection was presented in Chapter 4. While on average the chatter detection thresholds obtained from the human demonstrations via LSP showed good chatter detection accuracy and detection speed, it is clear from the data that high variance in the chatter detection thresholds can be detrimental to the general applicability of the LSP method when only one or a few demonstrations are possible. The high variance in the human operators' demonstrations must be reduced through a variance mitigation strategy such that the consistency of the chatter detection thresholds obtained from the LSP approach is improved. In addition, for the LSP approach to be a robust chatter detection method that has potential for wide acceptance in the manufacturing industry, the approach must be robust to variations in workpiece material, cutting directions, and cutting conditions (e.g., radial immersion).

In the following sections, a variance mitigation strategy for chatter detection using the LSP approach is presented, and its performance compared to the performance of the approach presented in Chapter 4. Specifically, the LSP and the variance mitigation strategy are applied to milling demonstrations, and the robustness of the resulting chatter detection thresholds is evaluated for different workpiece materials, directions of cut, and cutting conditions.

5.2   Variance Mitigation Strategy

In this section, a variance mitigation strategy for the LSP is presented to reduce the negative impact of high variance in the human operator's reaction times on the chatter detection thresholds determined using the approach described in Chapter 4. In Chapter 4, the variation in operator's reaction time to chatter, denoted by $r = t_s - t_c$, was found to significantly influence the chatter detection thresholds and in turn the chatter detection accuracy and speed of detection. To distinguish the work presented in this Chapter from the "naïve" approach presented in Chapter 4, the method presented earlier is referred to as the "naïve interpretation" approach where the variance in the reaction time of the operator during a demonstration is assumed to have negligible effect on the chatter detection threshold. The average accuracy of the chatter detection thresholds obtained using the naïve interpretation-based LSP is approximately 90.5%. However, a few thresholds showed significant variation in the chatter detection accuracy, which ranged from a high of 100% to a low of 57.1%. To minimize this variation, a heuristic rule-based variance mitigation strategy is proposed in this chapter and its chatter detection performance is evaluated.

5.2.1   Drawbacks of the Naïve Interpretation

In Chapter 3, a human operator's reaction time to chatter was characterized by a Gaussian distribution defined by a mean and standard deviation of $\bar{r}$ and $s_{\bar{r}}$, respectively. In Chapter 4, the naïve interpretation was adopted to interpret the human's input. As shown in Figure 39, assuming the true human observation instance t_c is constant and the human signal instance t_s is exactly equal to $\overline{t_s}$, the estimated chatter observation instance $\widehat{t_c}$ is equal to the true chatter observation instance t_c. If the reaction time in a demonstration is

longer than the mean reaction time for the operator, the operator reacts to the chatter late and $\hat{t}_c > t_c$; conversely, a short reaction time implies $\hat{t}_c < t_c$.
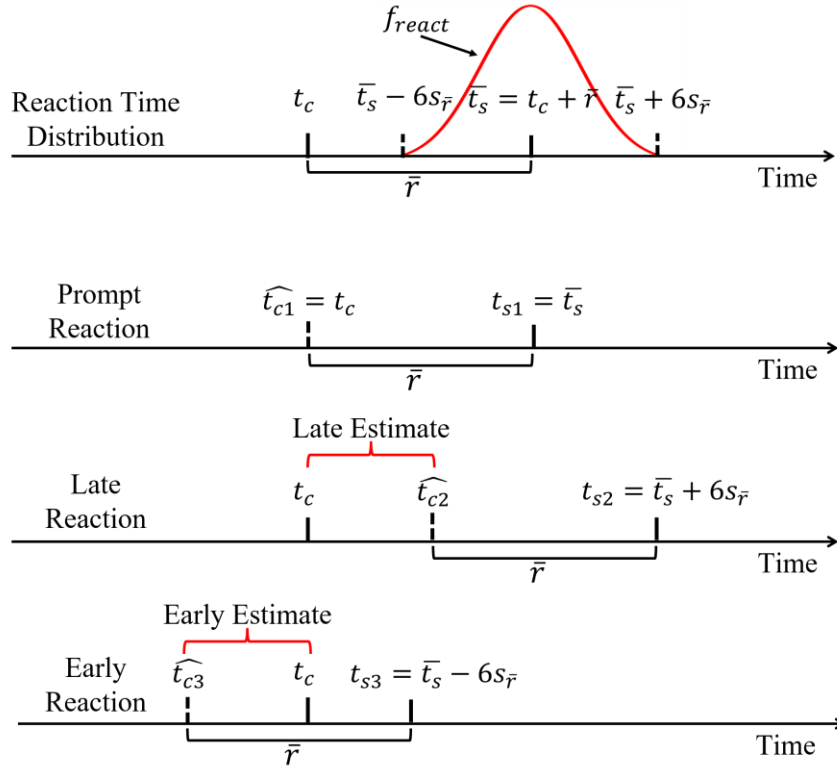
Figure 39. The effect of prompt signal instance $\boldsymbol{t_{s1}}$, late signal instance $\boldsymbol{t_{s2}}$ and early signal instance $\boldsymbol{t_{s3}}$ on $\hat{\boldsymbol{t}_c}$.

Although a large variance in $\hat{t}_c$ can result from the naïve interpretation, assuming the population follows a Gaussian distribution, the accuracy of the average $\hat{t}_c$ estimate improves with the number of demonstrations.

However, if only one or a few demonstrations are available, as would be typical in a production setting, the effectiveness of the naïve interpretation is subject to variation in the quality of demonstrations, namely, the timeliness of the operator's reaction. Therefore, a variance mitigation strategy must be incorporated to reduce the impact of potentially poor-quality demonstrations.

5.2.2   Variance Mitigation Strategy—Transient Period Rule

The heuristic rule developed in this thesis follows an essential observation of chatter in milling. It has been experimentally observed that the first chatter mark appears during the emergence of the chatter frequency in the frequency spectrum of the measured audio signal before reaching its "steady-state" amplitude corresponding to "fully-developed" chatter. Therefore, the premise of the heuristic rule is that the true t_c should fall within a transient period where $P_{chatter}(t)$ ramps up before reaching steady-state or fully-developed chatter. This heuristic is termed here as the "transient period rule."

The transient period rule classifies the chatter signal into three periods: stable period, transient period, and chatter period. The boundary between the stable period and the transient period is denoted by t_{st} while the boundary between the transient period and the chatter period is denoted by t_{tc}. The task of identifying the transient period is then reduced to identifying t_{st} and t_{tc}, where $P_{chatter}(t_{tc})$ is the updated upper bound of the chatter detection threshold, P_{th}, and $P_{chatter}(t_{st})$ is its lower bound.
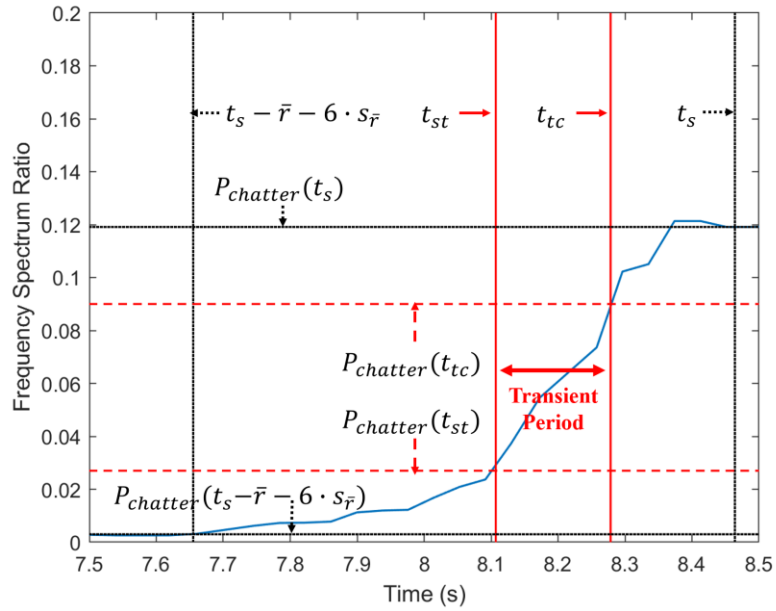
Figure 40. The transient period rule reduces the range of P_{th}.

As shown in Figure 40, the transient period rule serves to limit the possible range of P_{th} by setting the upper bound to $P_{chatter}(t_{tc})$ and the lower bound to $P_{chatter}(t_{st})$, thereby mitigating the impact of large variations in the operator's reaction time on the chatter detection thresholds i.e. poor quality demonstrations.

The algorithm for determining t_{st} and t_{tc} is as follows. Because the transient period is characterized by a rapid increase in the frequency spectrum ratio, the rate of change of the frequency spectrum ratio is first evaluated:

$$DP_{chatter} = \frac{dP_{chatter}(t)}{dt} \tag{20}$$

However, the time derivative $DP_{chatter}$ is sensitive to the feed rate in milling. Therefore, $DP_{chatter}$ is normalized by the tooth passing frequency, which is given by the spindle rotation frequency multiplied by the number of cutting edges of the end mill, as follows:

85

$$\frac{DP_{chatter}}{f_{tooth}} = \frac{dP_{chatter}}{f_{tooth}dt} \tag{21}$$

where f_{tooth} is the tooth passing frequency. This normalization results in a dimensionless quantity that enables comparison across different feed rates. Figure 41 shows the normalized rate of change of the frequency spectrum ratio $\frac{DP_{chatter}}{f_{tooth}}$ for two feed rates.
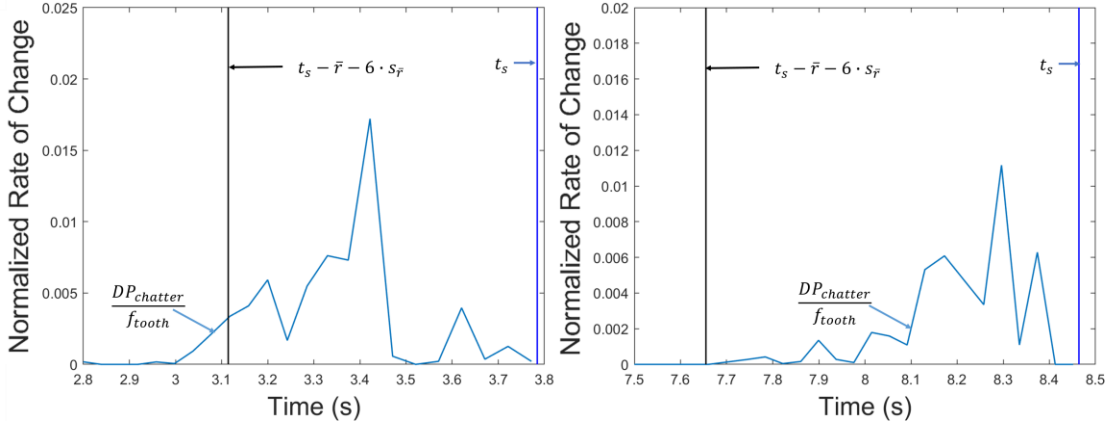
Figure 41. Normalized rate of change of frequency spectrum ratio; left: feed rate = 264 mm/min, right: feed rate = 132 mm/min.

Step 1 of the algorithm is to identify $\max(\frac{DP_{chatter}(t)}{f_{tooth}})$, namely the highest normalized rate of change of the spectrum frequency ratio. Step 2 of the algorithm is to compare $\max(\frac{DP_{chatter}(t)}{f_{tooth}})$ with a pre-determined upper limit of the normalized rate of change of frequency spectrum ratio. Based on the stable cutting region, where the likelihood of $\hat{t}_c$ is zero as shown in Figure 42, the upper limit, UL, of $\frac{DP_{chatter}(t)}{f_{tooth}}$ is set to $UL = \mu_{stable} + 6 \cdot \sigma_{stable}$, μ_{stable} is the mean value of $P_{chatter}(t)$ in the stable cutting region and σ_{stable} is the corresponding standard deviation.

86
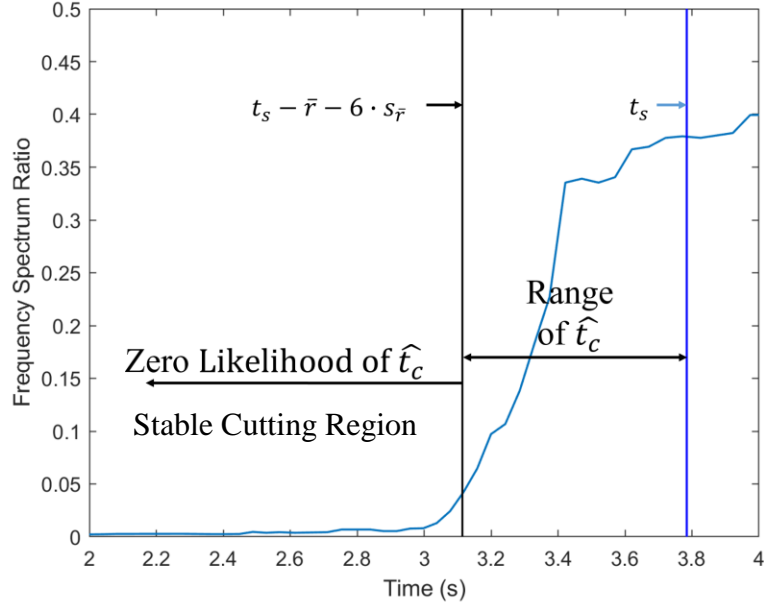
Figure 42. Stable cutting region where the likelihood of $\widehat{t_c}$ falling is zero.

Considering that the frequency spectrum ratio $P_{chatter}(t)$ is proportional to the instantaneous cutting force and vibration in a milling process, the rationale for setting $UL = \mu_{stable} + 6 \cdot \sigma_{stable}$ is that if the magnitude of increase in vibration from the previous tooth pass to the current tooth pass exceeds the average vibration plus six times its standard deviation in the stable cutting region, the milling process is certainly experiencing chatter. Based on this, t_{tc} assumes a value such that:

$$\frac{DP_{chatter}(t_{tc})}{f_{tooth}} = min\begin{cases} max\left(\dfrac{dP_{chatter}(t)}{f_{tooth}dt}\right) \\ UL \end{cases} \qquad (22)$$

If $max \frac{DP_{chatter}(t)}{f_{tooth}} > UL$:

$$t_{tc} = min\left\{t \middle| \frac{DP_{chatter}(t)}{f_{tooth}} = UL\right\} \qquad (23)$$

Otherwise:

$$t_{tc} = argmax_t(\frac{DP_{chatter}(t)}{f_{tooth}})$$ (24)

This ensures that t_{tc} is set to the earlier of the two time instances. Similarly, Step 3

of the algorithm calculates $\frac{1}{\gamma} \cdot max(\frac{DP_{chatter}(t)}{f_{tooth}})$, while Step 4 compares it with $\frac{1}{\gamma} \cdot UL$,

where γ is a parameter that can be tuned to yield the desired false positive rate. A high

value of γ would increase the false positive rate, but a low γ can result in late detection of

chatter. In this thesis, γ was set to 4. It follows that:

$$\frac{dP_{chatter}(t_{st})}{f_{tooth}dt} = max \begin{cases} \frac{1}{\gamma} \cdot max\left(\frac{dP_{chatter}(t)}{f_{tooth}dt}\right) \\ \frac{1}{\gamma} \cdot (UL) \end{cases}$$ (25)

Therefore, if $\frac{1}{\gamma} \cdot max\left(\frac{DP_{chatter}(t)}{f_{tooth}}\right) > \frac{1}{\gamma} \cdot UL$:

$$t_{st} = min\left\{t \left| \frac{DP_{chatter}(t)}{f_{tooth}} = \frac{1}{\gamma} \cdot max\left(\frac{DP_{chatter}(t)}{f_{tooth}}\right)\right.\right\}$$ (26)

Otherwise:

$$t_{st} = min\left\{t \left| \frac{DP_{chatter}(t)}{f_{tooth}} = \frac{1}{\gamma} \cdot UL\right.\right\}$$ (27)

Which effectively sets t_{st} to the later of the two time instances. Figure 43 is a flowchart

summarizing the four steps of the transient period rule.
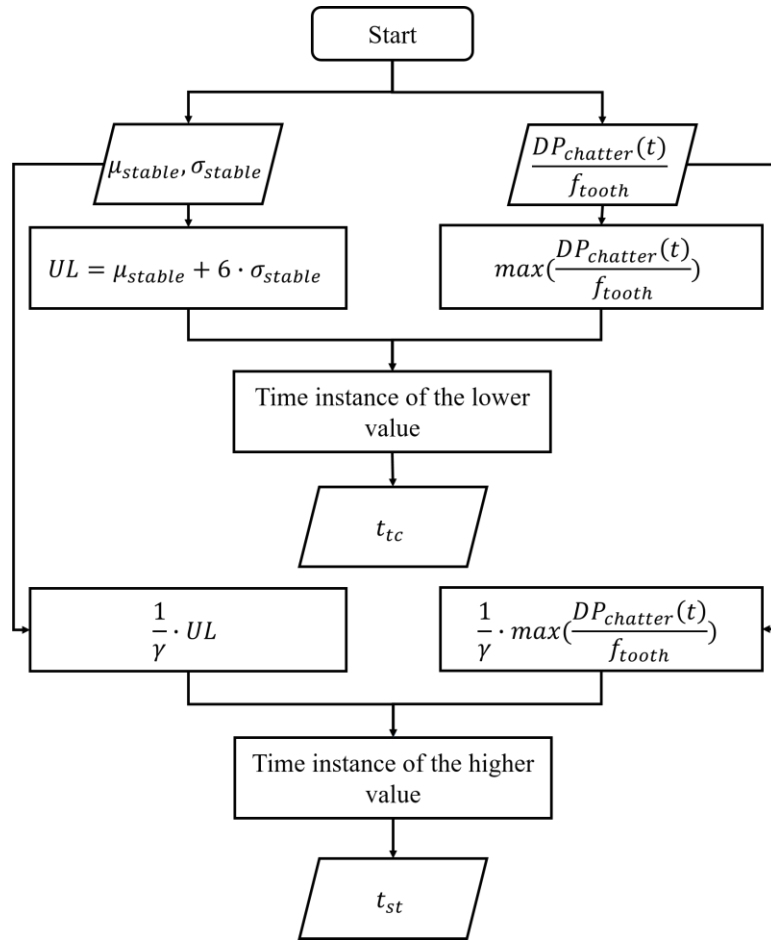
Figure 43. Flowchart of the transient period rule.
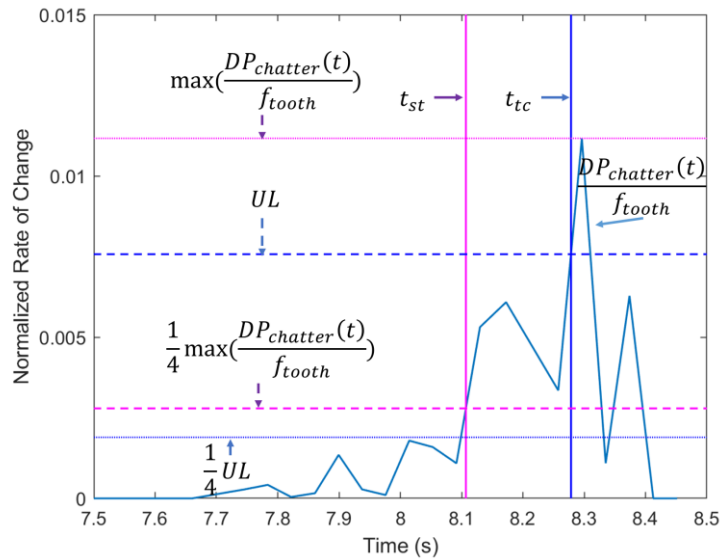
Figure 44. Transient period rule; identify t_{st} and t_{tc}.

Figure 44 is an example of the transient period rule. t_{tc} is the first time instance when the lower of the top two horizontal lines intersects with $\frac{DP_{chatter}(t)}{f_{tooth}}$, and t_{st} is the first time instance when the higher of the bottom two horizontal lines intersects with $\frac{DP_{chatter}(t)}{f_{tooth}}$. Note that t_{st} is modulated based on $t_{st} \leq t_{tc}$. In the next section, the effectiveness of the transient period rule is tested under different cutting conditions.

5.2.3 Variance Mitigation Strategy—Results and Discussion

Using the variance mitigation strategy defined by the transient period rule, the updated chatter detection thresholds, P_{th}, were computed for the demonstration experiments analyzed in Chapter 4 and are given in Table 9.

Table 9. Updated chatter detection thresholds compared with the original chatter detection thresholds; workpiece material is 4140 steel.

Operator	Spindle Speed (RPM)	Axial DOC (mm)	Original P_{th}	Updated P_{th}	$P_{chatter}(t_a)$	$P_{max,chatter}$	$P_{max,stable}$
A	750	3.5	0.020	0.020	0.038	0.143	N/A
A	2000	2	0.038	0.038	0.056	0.092	N/A
A	2000	2.5	0.209	0.065	0.039	0.451	N/A
B	750	3.5	0.036	0.036	0.032	0.170	N/A
B	2000	3	0.026	0.075	0.036	0.386	N/A
C	1000	2.5	0.013	0.037	0.024	0.174	N/A
C	1000	3	0.019	0.061	0.033	0.191	N/A
C	1750	2	0.035	0.032	0.038	0.044	N/A
C	1750	2.5	0.014	0.051	0.085	0.456	N/A
A	750	3	N/A				0.013
A	2000	1.5	N/A				0.003
B	750	3	N/A				0.021
B	2000	2	N/A				0.016
C	1000	2	N/A				0.004

While six out of nine chatter detection thresholds were updated by the variance mitigation strategy, the others remained the same as the original thresholds obtained from the naive interpretation approach used in Chapter 4. The average performance of the updated chatter detection thresholds is shown in Table 10.

Table 10. Average performance of the updated chatter detection thresholds after applying the variance mitigation strategy.

	Occurrence	Total Tests	Updated Performance	Original Performance
Overall Accuracy	121	126	96.0%	90.5%
Early Detection	37	81	45.7%	74.1%
Late Detection	40	81	49.4%	18.5%
False Negative	4	81	4.9%	7.4%
False Positive	1	45	2.2%	13.3%

It is clear from the performance data that the variance mitigation strategy yields a higher overall chatter detection accuracy. It is also evident from the table that the false positive and false negative rates of the updated thresholds are lower than those for the original thresholds. However, the updated thresholds are unable to detect chatter as early as the original thresholds, which is a trade-off with improved detection accuracy. In order to lower the false positive rate, i.e. falsely identifying a stable process as unstable, the updated threshold values must be set higher than the original thresholds, which means that chatter is detected at a later time instance. Figure 45 shows examples of the tradeoffs between detection speed and detection accuracy. In the figure on the left, the lower original threshold obtained from the human demonstration detects chatter earlier than the updated chatter detection threshold, whereas the figure on the right shows the same thresholds detect a stable cutting process in contrast to the original threshold, which produced a false

positive result. The chatter detection accuracy is therefore improved by setting a higher

chatter detection threshold at the cost of detection speed.
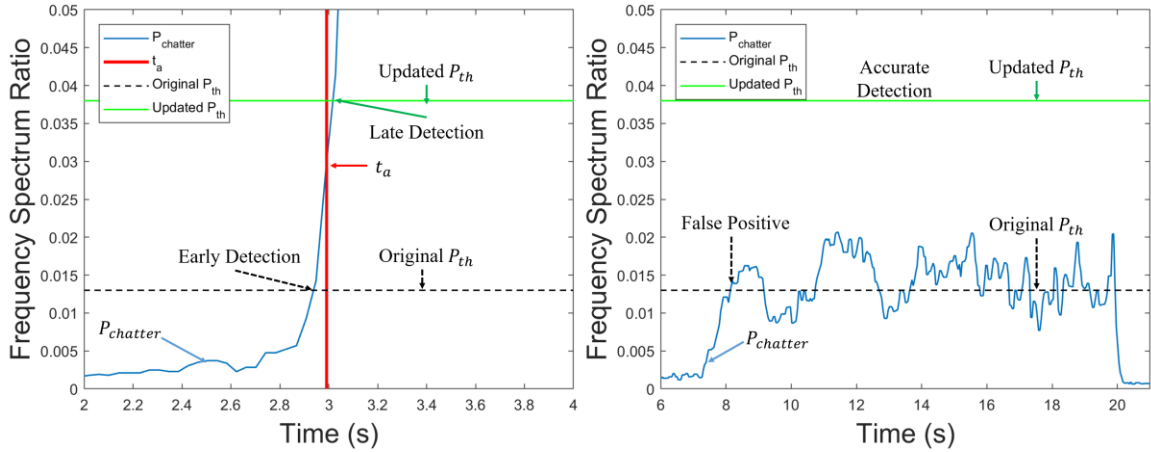
Figure 45. Trade-off between detection speed and detection accuracy. Left: comparison of detection speed; right: comparison of the detection accuracy.

While the average performance of the chatter detection thresholds is improved by

the proposed the variance mitigation strategy, the objective of the variance mitigation

strategy was to reduce the variance of the chatter detection thresholds. Unlike the naïve

interpretation approach where the variability in the human reaction time can produce

outlier thresholds that are prone to false detections, Table 11 shows consistent detection

accuracy across all chatter detection thresholds obtained using the variance mitigation

strategy. In comparison to the detection accuracy of the original chatter detection

thresholds listed in Table 8 in Chapter 4, it is clear that the variance mitigation strategy

successfully reduces the likelihood of obtaining an outlier chatter detection threshold that

is prone to false detections across different milling conditions.

Table 11. Breakdown of the updated chatter detection thresholds.

Updated P_{th}	Early Detection Rate	Late Detection Rate	False Negative Rate	False Positive Rate	Detection Accuracy
0.020	100.0%	0.0%	0.0%	20.0%	92.9%
0.038	55.6%	44.4%	0.0%	0.0%	100.0%
0.065	11.1%	77.8%	11.1%	0.0%	92.9%
0.036	55.6%	44.4%	0.0%	0.0%	100.0%
0.075	11.1%	77.8%	11.1%	0.0%	92.9%
0.037	55.6%	44.4%	0.0%	0.0%	100.0%
0.061	11.1%	77.8%	11.1%	0.0%	92.9%
0.032	88.9%	11.1%	0.0%	0.0%	100.0%
0.051	22.2%	66.7%	11.1%	0.0%	92.9%

Ultimately, based on data obtained from the milling experiments conducted on 4140 steel alloy, the variance mitigation strategy successfully improved the consistency of the chatter detection thresholds, which resulted in good detection accuracy and acceptable detection speed. The next section examines the effectiveness of the LSP and the variance mitigation strategy in a milling experiment conducted on a 7075-T651 aluminum workpiece.

5.3 Applying the LSP and Variance Mitigation Strategy to Milling of 7075-T651 Aluminum

The same experiments were conducted on a 7075-T651 aluminum workpiece to examine the robustness of the LSP and the variance mitigation strategy to variations in the workpiece material. The feed per tooth in the aluminum milling experiments was 0.0165 mm. The experimental setup was identical to that used in the previous experiments on 4140

steel alloy. Figure 46 is a picture of the slot end milling cuts made in the aluminum workpiece and Figure 47 shows an example of the resulting chatter marks in one of the experiments.

Figure 46. Slots milled in 7075-T651 aluminum workpiece.
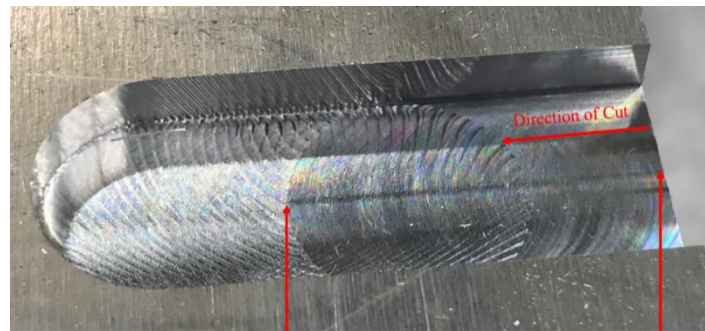
Figure 47. Representative chatter marks on the aluminum workpiece.

5.3.1 Applying LSP using Naïve Interpretation

The naïve interpretation based LSP approach was applied to the human demonstrations of chatter during milling of 7075-T651 aluminum to obtain the corresponding chatter detection thresholds.

Table 12. Chatter detection thresholds using the naïve interpretation.

Operator	Human Observation	Spindle Speed (RPM)	Axial DOC (mm)	P_{th}	$P_{chatter}(t_a)$	$P_{max,chatter}$	$P_{max,stable}$
A	Chatter	4600	3.5	0.003	0.014	0.406	N/A
A	Chatter	9600	3	0.008	0.016	0.362	N/A
A	Chatter	8500	2.5	0.004	0.037	0.439	N/A
B	Chatter	5500	4	0.041	0.064	0.487	N/A
B	Chatter	5500	3.5	0.003	0.004	0.381	N/A
C	Chatter	6500	2.5	0.004	0.018	0.460	N/A
C	Chatter	6500	2	0.004	0.008	0.367	N/A
C	Chatter	7500	2	0.043	0.056	0.249	N/A
C	Chatter	7500	1.5	0.004	0.036	0.360	N/A
A	Stable	4600	2.5	N/A			0.005
A	Stable	9600	3	N/A			0.008
B	Stable	8500	2	N/A			0.007

Table 11 shows the original chatter detection thresholds P_{th} obtained using the naïve interpretation based LSP approach. The original chatter detection thresholds are much lower than those obtained from the 4140 steel milling experiments, which can lead to false positives. Table 13 shows the performance of each chatter detection threshold obtained in the aluminum milling experiments.

Table 13. Performance of original chatter detection thresholds.

P_{th}	Early Detection Rate	Late Detection Rate	False Negative Rate	False Positive Rate	Detection Accuracy
0.003	100.0%	0.0%	0.0%	100.0%	75.0%
0.008	88.9%	11.1%	0.0%	0.0%	100.0%
0.004	88.9%	11.1%	0.0%	100.0%	75.0%
0.041	22.2%	77.8%	0.0%	0.0%	100.0%
0.003	100.0%	0.0%	0.0%	100.0%	75.0%
0.004	88.9%	11.1%	0.0%	100.0%	75.0%
0.004	88.9%	11.1%	0.0%	100.0%	75.0%
0.043	22.2%	77.8%	0.0%	0.0%	100.0%
0.004	88.9%	11.1%	0.0%	100.0%	75.0%

It is clear from Table 13 that the original chatter detection thresholds are prone to false positives. In fact, many of the chatter detection thresholds are so low that all stable processes are falsely classified as chatter. Figure 48 illustrates the difference between a high-speed 7075-T651 aluminum alloy milling experiment at 9600 RPM and a lower speed 4140 steel milling experiment at 2000 RPM.
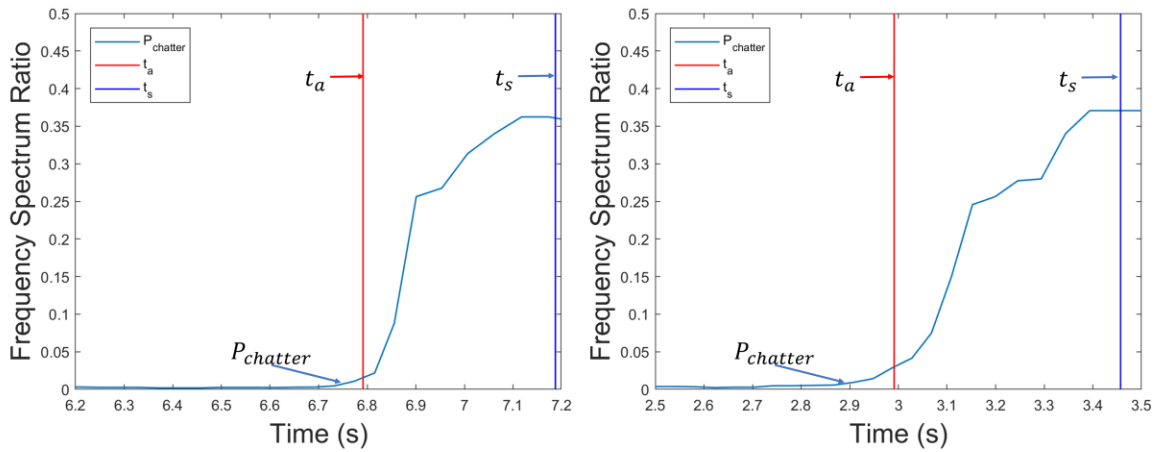
Figure 48. The signal ramp up of aluminum (left, 9600 RPM) versus steel (right, 2000 RPM.)

It is observed in Figure 48 that the frequency spectrum ratio in the aluminum experiment ramps up more quickly than in the steel experiment due to the higher material removal rate corresponding to the higher spindle speed.

Although the signal increases faster in the high-speed aluminum milling experiment, the operator's reaction time to chatter remains constant, which means that the time between chatter onset and fully developed chatter is much shorter. Even if the human operator can recognize the onset of chatter, the variance in the operator's demonstration is too large compared to the rapid growth of the chatter signal. The chatter observation instance t_c obtained from the high speed milling experiment by subtracting the average reaction time of the operator $\bar{r}$ from the human signal instance t_s is more likely to be either too early or too late, which is the primary reason why the chatter detection thresholds obtained from the human demonstration without the variance mitigation strategy are less accurate. Therefore, it is evident that direct implementation of the chatter detection thresholds P_{th} obtained from the human operator demonstrations using the naïve interpretation approach is less successful. The chatter detection thresholds obtained using the variance mitigation strategy are analyzed next.

5.3.2 Applying the Variance Mitigation Strategy

As discussed in Section 5.2, the objective of the variance mitigation strategy is to reduce the negative impact of the variance in the human's reaction time on the chatter detection thresholds. While in Chapter 4 the direct implementation of chatter detection thresholds obtained using the naïve interpretation approach was found to be effective, the

chatter detection accuracy in the 7075-T651 Aluminum experiments was drastically lower due to the higher feed rate involved. Therefore, under these conditions the accuracy of the chatter detection thresholds is more reliant on the effectiveness of the variance mitigation strategy.

Table 14. Updated chatter detection thresholds using the variance mitigation strategy.

Operator	Spindle Speed (RPM)	Axial DOC (mm)	Original P_{th}	Updated P_{th}	$P_{chatter}(t_a)$	$P_{max,chatter}$	$P_{max,stable}$
A	4600	3.5	0.003	0.062	0.014	0.406	N/A
A	9600	3	0.008	0.054	0.016	0.362	N/A
A	8500	2.5	0.004	0.049	0.037	0.439	N/A
B	5500	4	0.041	0.056	0.064	0.487	N/A
B	5500	3.5	0.003	0.087	0.004	0.381	N/A
C	6500	2.5	0.004	0.055	0.018	0.460	N/A
C	6500	2	0.004	0.056	0.008	0.367	N/A
C	7500	2	**0.043**	**0.043**	0.056	0.249	N/A
C	7500	1.5	0.004	0.058	0.036	0.360	N/A
A	4600	2.5	N/A				0.005
A	9600	3	N/A				0.008
B	8500	2	N/A				0.007

As shown in Table 14, only one out of nine chatter detection thresholds remains the same as the original chatter detection thresholds obtained using the naïve interpretation approach, while the rest are updated by the variance mitigation strategy. The performance of the updated chatter detection thresholds is shown in Table 15.

Table 15. Performance of the updated chatter detection thresholds.

P_{th}	Early Detection Rate	Late Detection Rate	False Negative Rate	False Positive Rate	Detection Accuracy
0.062	11.1%	88.9%	0.0%	0.0%	100.0%
0.054	22.2%	77.8%	0.0%	0.0%	100.0%
0.049	22.2%	77.8%	0.0%	0.0%	100.0%
0.056	11.1%	88.9%	0.0%	0.0%	100.0%
0.087	0.0%	100.0%	0.0%	0.0%	100.0%
0.055	22.2%	77.8%	0.0%	0.0%	100.0%
0.056	11.1%	88.9%	0.0%	0.0%	100.0%
0.043	22.2%	77.8%	0.0%	0.0%	100.0%
0.058	11.1%	88.9%	0.0%	0.0%	100.0%

Based on data in Table 14, the detection accuracy of the updated chatter detection thresholds obtained using the variance mitigation strategy is significantly improved over the naïve interpretation approach. In most cases, however, the updated chatter detection thresholds are unable to detect chatter prior to the appearance of chatter marks on the workpiece. The data further confirms the need for the variance mitigation strategy to improve the chatter detection accuracy in a high-speed milling scenario.

It is evident from the results presented in this section that application of the LSP in conjunction with the variance mitigation strategy in a high-speed milling operation on 7075-T651 Aluminum workpiece is effective. Although the variance in the operator's reaction times resulted in much higher inaccuracies in the performance of the chatter detection thresholds, the variance mitigation strategy successfully updated the chatter detection thresholds such that the detection accuracy was improved and the detection speed was acceptable. In the next section, the robustness of the chatter detection thresholds

obtained from the variance mitigation based LSP strategy is tested under different radial immersions and cutting directions.

5.4 Robustness of the LSP using the Variance Mitigation Strategy

In this section, milling experiments under different radial immersions and directions of cut were performed to evaluate the robustness of the chatter detection thresholds determined using the variance mitigation strategy presented earlier.

5.4.1 Radial Immersion Experiments

Radial immersion refers to the radial depth of cut, which is expressed as a percentage of the tool diameter. Experiments under different radial immersions were conducted on 7075-T651 aluminum alloy to examine the robustness of the chatter detection thresholds obtained via the LSP using the variance mitigation strategy. The experimental setup was unchanged from that used in the previous experiments. The feed per tooth and the spindle speed were kept constant at 0.0165 mm and 9600 RPM, respectively. Table 15 shows results of these milling experiments.

Table 16. Chatter and stable milling processes under different radial immersions.

Workpiece	Radial Immersion	Spindle Speed (RPM)	Axial DOC (mm)	$P_{chatter}(t_a)$	$P_{max,chatter}$	$P_{max,stable}$
7075 Al	75%	9600	3.5	0.055	0.4424	N/A
7075 Al	25%	9600	6	Stable		0.011
7075 Al	25%	9600	7	Stable		0.013
7075 Al	25%	9600	8	Stable		0.012
7075 Al	50%	9600	3	Stable		0.011
7075 Al	75%	9600	1.5	Stable		0.010
7075 Al	75%	9600	2	Stable		0.013
7075 Al	75%	9600	2.5	Stable		0.017

The allowable axial depth of cut for a stable cutting process increases with decrease in radial immersion. Much higher axial DOC was required to produce chatter. Therefore, the only chatter data collected was at 75% radial immersion for a 3.5 mm axial depth of cut. The rest of the experiments resulted in stable cutting.

The performance of the chatter detection thresholds reported in Table 14 of Section 5.3.2 were evaluated under different radial immersions. Table 17 shows the average accuracies and false detection rates for all the detection thresholds.

Table 17. Average performance of chatter detection thresholds under different radial immersions

	Occurrences	Total Tests	Percentage
Overall Accuracy	72	72	100%
Early Detection	3	9	33.3%
Late Detection	6	9	66.7%
False Negative	0	9	0%
False Positive	0	63	0%

Based on the data in Table 16, the chatter detection thresholds obtained using the LSP utilizing the variance mitigation strategy are able to detect chatter accurately under different radial immersions, and the detection speeds are satisfactory based on the percentages of the early and late detections given in Table 17.

5.4.2 Cutting Direction

During a milling operation, the cutting tool can be subject to complex workpiece geometries or complex tool paths. Therefore, the chatter detection thresholds should work for multiple directions of cut. Any two-dimensional cutting path can be considered as a linear combination of two orthogonal vectors in the x- and y-directions. Previously, all experimental data were collected from linear slot end milling cuts along the y-axis of the machine tool. In this section, slot end milling cuts were made along the x-axis of the machine tool.

The radial immersion was maintained at 100% and the feed per tooth was kept constant at 0.0165 mm. Table 18 shows the experimental data collected from these experiments.

Table 18. Milling experiments in x-axis.

Workpiece	Spindle Speed (RPM)	Axial DOC (mm)	$P_{chatter}(t_a)$	$P_{max,chatter}$	$P_{max,stable}$
7075 Al	5500	3.5	0.033	0.370	N/A
7075 Al	9600	2	0.077	0.403	N/A
7075 Al	5500	1.5	Stable		0.006
7075 Al	9600	1.5	Stable		0.010

For each spindle speed, data for one chatter and one stable cutting tests were collected to evaluate the detection accuracy and speed of the chatter detection thresholds previously determined from the variable mitigation strategy based LSP applied to the slot end milling tests performed in the y-direction. Table 19 shows the average performance of the chatter detection thresholds determined from the y-direction milling experiments in detecting chatter produced in the x-direction milling tests.

Table 19. Average performance of chatter detection thresholds for milling tests along the x-axis.

	Occurrences	Total Tests	Percentage
Overall Accuracy	36	36	100%
Early Detection	8	18	44.4%
Late Detection	10	18	55.6%
False Negative	0	18	0%
False Positive	0	18	0%

Based on the above data, the detection speeds are nearly half early and half late, which means that the detection thresholds on average detect chatter right when chatter marks begin to appear on the workpiece; the detection accuracy of the chatter detection thresholds obtained via the LSP using the variance mitigation strategy remains high when applied to test data obtained from milling experiments performed along a different direction of cut, which further proves the robustness of the proposed methodology.

5.5   Summary

In summary, Chapter 5 introduced a variance mitigation strategy for the "Learnable Skill Primitive" (LSP), which significantly improved the consistency of chatter detection

accuracy and speed by minimizing the variance in the chatter detection thresholds arising from the variance in operator reaction times during demonstrations. The effectiveness of the variance mitigation strategy was further validated through high-speed milling experiments performed on a 7075-T651 aluminum workpiece, where the performance of the naïve interpretation of the human operator's demonstration used to establish the chatter detection thresholds was found to perform worse than the variance mitigation strategy presented in this chapter. Finally, the robustness of the LSP using the variance mitigation strategy was further validated through milling experiments under different radial immersions and directions of cutting.

CHAPTER 6 CONCLUSIONS AND RECOMMENDATIONS

This chapter summarizes the original contributions and main conclusions of this thesis and suggests possible areas for future studies.

6.1　Original Contributions

An innovative methodology and associated algorithms for chatter-specific knowledge acquisition through human-machine interaction were presented in this thesis. The modes of interaction between human operators and a "digital apprentice" in a production setting via perception and reasoning mapping was evaluated in detail. The "learnable skill primitive" method for chatter detection was developed to interpret human demonstrations and obtain a chatter detection threshold with acceptable detection accuracy and speed. A human reaction time variance mitigation strategy was proposed to reduce the uncertainty in human operator demonstrations. In addition, robustness of the "learnable skill primitive" and the "variance mitigation strategy" were tested under different cutting conditions, workpiece materials, and cutting paths. The originality of this research lies in the methodology for incorporating human operator demonstrations in a milling process for automated detection of chatter. The chatter detection thresholds were "learned" from the human operators, resulting in good detection accuracy and detection speed. The proposed methods in this thesis provide a novel solution to the process monitoring problem by leveraging the presence of a human operator who is experienced and knowledgeable about the manufacturing process. The contributions discussed in this thesis are applicable to other manufacturing applications where a human's experience and perception can be leveraged for better quality decision making.

6.2 Main Conclusions

The main conclusions of this thesis are summarized below.

- Experimental data demonstrated that human operators are able to accurately distinguish between an unstable (chatter) and a stable cutting process. However, human operators are unable to signal chatter without delay because of their reaction time.

- The reaction times associated with human recognition of chatter were modelled as a probability distribution. The mean estimated human observation instance of chatter $\hat{t}_c$ obtained from milling experiments suggested that human operators recognize chatter close to when chatter marks appear on the workpiece.

- Experimental data collected from milling experiments indicated that the "Learnable Skill Primitive" (LSP) for chatter detection developed using the naïve interpretation yielded good overall chatter detection accuracy and speed from a single demonstration; however, it was also found that a few outliers in performance were caused by the high variance in the operators' demonstrations.

- The proposed heuristics-based variance mitigation strategy is capable of mitigating the influence of the variance in human demonstrations, resulting in chatter detection thresholds that yield significantly improved chatter detection accuracy and speed.

- The LSP and the variance mitigation strategy were found to be robust to variations in the workpiece material, cutting condition (radial immersion), and cutting direction.

6.3 Future Work and Recommendations

The introduction of the learnable skill primitive for chatter detection marks the beginning of a series of potential future developments centering on effective human-

machine interaction in a production setting. The immediate next step is to test the effectiveness of the LSP method in other machining process such as boring and single point turning. I addition, there is interest in learning the human operator's ability to control chatter by learning a control policy consisting of a set of control actions demonstrated by human operators, which is a natural next step of the LSP method.

Another area of interest for future work is the integration of multiple LSPs for the overall improvement of a manufacturing process. By incorporating a set of LSPs that learn from human operators to monitor different aspects of the manufacturing process such as chatter and tool wear, the collective decision making performance of the human-machine interaction can potentially improve the overall productivity of a complex manufacturing process.

APPENDIX

Appendix I. Experience of human operators participated in this research.

	Position	Machining Experience	Experience with Chatter
Operator A	Graduate Research Assistant	2 years	0 year
Operator B	Graduate Research Assistant	6 years	6 years
Operator C	Mechanical Engineer II	3 years overall, 2 years of professional experience	2 years

REFERENCES

[1]   G. Quintana and J. Ciurana, "Chatter in machining processes: A review," *International Journal of Machine Tools and Manufacture,* vol. 51, no. 5, pp. 363-376, 2011.

[2]   J.-V. Le Lan, A. Marty, and J.-F. Debongnie, "A stability diagram computation method for milling adapted to automotive industry," presented at the Second CIRP International Conference on High performance cutting, Vancouver, Canada, 2006.

[3]   S. Tobias, "Machine tool vibration research," *International Journal of Machine Tool Design and Research,* vol. 1, no. 1-2, pp. 1-14, 1961.

[4]   S. Smith and J. Tlusty, "An overview of modeling and simulation of the milling process," *Journal of manufacturing science and engineering,* vol. 113, no. 2, 1991.

[5]   Y. Altintaş and E. Budak, "Analytical prediction of stability lobes in milling," *CIRP annals,* vol. 44, no. 1, pp. 357-362, 1995.

[6]   Y. Altintas and A. Ber, "Manufacturing automation: metal cutting mechanics, machine tool vibrations, and CNC design," *Appl. Mech. Rev.,* vol. 54, no. 5, pp. B84-B84, 2001.

[7]   G. Duncan, T. L. Schmitz, and M. H. Kurdi, *Uncertainty propagation for selected analytical milling stability limit analyses*. Society of Manufacturing Engineers, 2000.

[8]   L. Ma, S. N. Melkote, and J. B. Castle, "A model-based computationally efficient method for on-line detection of chatter in milling," *Journal of manufacturing science and engineering,* vol. 135, no. 3, 2013.

[9]   R. Faassen, E. Doppenberg, N. van de Wouw, J. Oosterling, and H. Nijmeijer, "Online detection of the onset and occurrence of machine tool chatter in the milling process," in *CIRP 2nd International Conference on High Performance Cutting*, 2006, pp. paper-no. 23.

[10]   P. K. Wright and D. A. Bourne, *Manufacturing intelligence*. Addison-Wesley Longman Publishing Co., Inc., 1988.

[11]   S. Chernova and A. L. Thomaz, "Robot learning from human teachers," *Synthesis Lectures on Artificial Intelligence and Machine Learning,* vol. 8, no. 3, pp. 1-121, 2014.

[12]   S. Schaal, "Learning from demonstration," in *Advances in neural information processing systems*, 1997, pp. 1040-1046.

[13]   B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and autonomous systems,* vol. 57, no. 5, pp. 469-483, 2009.

[14]   C. L. Nehaniv and K. Dautenhahn, "The correspondence problem," *Imitation in animals and artifacts,* vol. 41, 2002.

[15]   W. B. Knox and P. Stone, "Tamer: Training an agent manually via evaluative reinforcement," in *2008 7th IEEE International Conference on Development and Learning*, 2008, pp. 292-297: IEEE.

[16]   W. B. Knox and P. Stone, "Interactively shaping agents via human reinforcement: The TAMER framework," in *Proceedings of the fifth international conference on Knowledge capture*, 2009, pp. 9-16.

[17]   T. Delio, J. Tlusty, and S. Smith, "Use of audio signals for chatter detection and control," *Journal of Manufacturing Science and Engineering,* vol. 114, no. 2, 1992.

[18]   S. Smith and J. Tlusty, "Efficient simulation programs for chatter in milling," *CIRP annals,* vol. 42, no. 1, pp. 463-466, 1993.

[19]   V. Nguyen, S. Melkote, A. Deshamudre, M. Khanna, and D. Walker, "Comparison of on-line chatter detection algorithms in turning," in *2016 International Symposium on Flexible Automation (ISFA)*, 2016, pp. 87-94: IEEE.

[20]   A. Ganguli, A. Deraemaeker, and A. Preumont, "Regenerative chatter reduction by active damping control," *Journal of sound and vibration,* vol. 300, no. 3-5, pp. 847-862, 2007.

[21]   A. R. Yusoff and N. D. Sims, "Optimisation of variable helix tool geometry for regenerative chatter mitigation," *International Journal of Machine Tools and Manufacture,* vol. 51, no. 2, pp. 133-141, 2011.

[22]   E. Al-Regib, J. Ni, and S.-H. Lee, "Programming spindle speed variation for machine tool chatter suppression," *International Journal of Machine Tools and Manufacture,* vol. 43, no. 12, pp. 1229-1240, 2003.

[23]   S. Tobias and W. Fishwick, "Theory of regenerative machine tool chatter," *The engineer,* vol. 205, no. 7, pp. 199-203, 1958.

[24]   J. Tlusty and M. Polacek, "Experience with analysing stability of machine tool against chatter," in *9th MTDR Conference*, 1968, pp. 521-570.

[25]   I. Minis, R. Yanushevsky, A. Tembo, and R. Hocken, "Analysis of linear and nonlinear chatter in milling," *CIRP annals,* vol. 39, no. 1, pp. 459-462, 1990.

[26]   E. Budak and Y. Altintas, "Analytical prediction of chatter stability in milling—part I: general formulation," *Journal of Dynamic System Measurement and Control,* vol. 120, no. 1, 1998.

[27]   B. C. Rao and Y. C. Shin, "A comprehensive dynamic cutting force model for chatter prediction in turning," *International Journal of Machine Tools and Manufacture,* vol. 39, no. 10, pp. 1631-1654, 1999.

[28]   K. Jemielniak and A. Widota, "Numerical simulation of non-linear chatter vibration in turning," *International Journal of Machine Tools and Manufacture,* vol. 29, no. 2, pp. 239-247, 1989.

[29]   S. Smith and J. Tlusty, "Update on high-speed milling dynamics," *Journal of manufacturing science and engineering,* vol. 112, no. 2, 1990.

[30]   H. Li, X. Li, and X. Chen, "A novel chatter stability criterion for the modelling and simulation of the dynamic milling process in the time domain," *The International Journal of Advanced Manufacturing Technology,* vol. 22, no. 9-10, pp. 619-625, 2003.

[31]   J. Karandikar, M. Traverso, A. Abbas, and T. Schmitz, "Bayesian inference for milling stability using a random walk approach," *Journal of Manufacturing Science and Engineering,* vol. 136, no. 3, 2014.

[32]   M. G. Traverso, R. l. Zapata, T. L. Schmitz, and A. E. Abbas, "Optimal experimentation for selecting stable milling parameters: a Bayesian approach," in *International Manufacturing Science and Engineering Conference*, 2009, vol. 43611, pp. 277-282.

[33]   J. Karandikar, A. Honeycutt, T. Schmitz, and S. Smith, "Stability boundary and optimal operating parameter identification in milling using Bayesian learning," *Journal of Manufacturing Processes,* vol. 56, 2020.

[34]   L. Ma, S. N. Melkote, and J. B. Castle, "PVDF sensor-based monitoring of milling torque," *The International Journal of Advanced Manufacturing Technology,* vol. 70, no. 9-12, pp. 1603-1614, 2014.

[35]   L. Vela-Martínez, J. C. Jáuregui-Correa, and J. Álvarez-Ramírez, "Characterization of machining chattering dynamics: An R/S scaling analysis approach," *International Journal of Machine Tools and Manufacture,* vol. 49, no. 11, pp. 832-842, 2009.

[36]   J. Tlusty and G. Andrews, "A critical review of sensors for unmanned machining," *CIRP annals,* vol. 32, no. 2, pp. 563-572, 1983.

[37]   S. Y. Liang, R. L. Hecker, and R. G. Landers, "Machining process monitoring and control: the state-of-the-art," *J. Manuf. Sci. Eng.,* vol. 126, no. 2, pp. 297-310, 2004.

[38]   R. Teti, K. Jemielniak, G. O'Donnell, and D. Dornfeld, "Advanced monitoring of machining operations," *CIRP annals,* vol. 59, no. 2, pp. 717-739, 2010.

[39]   Y. Altintas and P. K. Chan, "In-process detection and suppression of chatter in milling," *International Journal of Machine Tools and Manufacture,* vol. 32, no. 3, pp. 329-347, 1992.

[40]   Y. Liao and Y. Young, "A new on-line spindle speed regulation strategy for chatter control," *International Journal of Machine Tools and Manufacture,* vol. 36, no. 5, pp. 651-660, 1996.

[41]   S. Smith and T. Delio, "Sensor-based chatter detection and avoidance by spindle speed selection," *Journal of Dynamic System Measurement and Control,* vol. 114, no. 3, 1992.

[42]   E. Kuljanic, M. Sortino, and G. Totis, "Multisensor approaches for chatter detection in milling," *Journal of Sound and Vibration,* vol. 312, no. 4-5, pp. 672-693, 2008.

[43]   Z. Yao, D. Mei, and Z. Chen, "On-line chatter detection and identification based on wavelet and support vector machine," *Journal of Materials Processing Technology,* vol. 210, no. 5, pp. 713-719, 2010.

[44]   C. L. Zhang, X. Yue, Y. T. Jiang, and W. Zheng, "A hybrid approach of ANN and HMM for cutting chatter monitoring," in *Advanced Materials Research*, 2010, vol. 97, pp. 3225-3232: Trans Tech Publ.

[45]   N. Vuković, M. Mitić, and Z. Miljković, "Trajectory learning and reproduction for differential drive mobile robots based on GMM/HMM and dynamic time warping using learning from demonstration framework," *Engineering Applications of Artificial Intelligence,* vol. 45, pp. 388-404, 2015.

[46]   B. Akgun, M. Cakmak, K. Jiang, and A. L. Thomaz, "Keyframe-based learning from demonstration," *International Journal of Social Robotics,* vol. 4, no. 4, pp. 343-355, 2012.

[47]   C. G. Atkeson and S. Schaal, "Robot learning from demonstration," in *ICML*, 1997, vol. 97, pp. 12-20: Citeseer.

[48]   C. G. Atkeson and S. Schaal, "Learning tasks from a single demonstration," in *Proceedings of International Conference on Robotics and Automation*, 1997, vol. 2, pp. 1706-1712: IEEE.

[49]   A. J. Ijspeert, J. Nakanishi, and S. Schaal, "Learning rhythmic movements by demonstration using nonlinear oscillators," in *Proceedings of the ieee/rsj int. conference on intelligent robots and systems (iros2002)*, 2002, no. CONF, pp. 958-963.

[50]   D. Kulić, W. Takano, and Y. Nakamura, "Incremental Learning, Clustering and Hierarchy Formation of Whole Body Motion Patterns using Adaptive Hidden Markov Chains," *The International Journal of Robotics Research,* vol. 27, no. 7, pp. 761-784, 2008.

[51]   D. Lee and Y. Nakamura, "Mimesis Model from Partial Observations for a Humanoid Robot," *The International Journal of Robotics Research,* vol. 29, no. 1, pp. 60-80, 2010.

[52]   S. Schaal, J. Peters, J. Nakanishi, and A. Ijspeert, "Control, planning, learning, and imitation with dynamic movement primitives," in *Workshop on Bilateral Paradigms on Humans and Humanoids: IEEE International Conference on Intelligent Robots and Systems (IROS 2003)*, 2003, pp. 1-21.

[53]   M. L. Littman, A. R. Cassandra, and L. P. Kaelbling, "Learning policies for partially observable environments: Scaling up," in *Machine Learning Proceedings 1995*: Elsevier, 1995, pp. 362-370.

[54]   C. L. Isbell Jr and C. R. Shelton, "Cobot: A social reinforcement learning agent," in *Advances in neural information processing systems*, 2002, pp. 1393-1400.

[55]   W. B. Knox, P. Stone, and C. Breazeal, "Training a robot via human feedback: A case study," in *International Conference on Social Robotics*, 2013, pp. 460-470: Springer.

[56]   A. L. Thomaz and C. Breazeal, "Teachable robots: Understanding human teaching behavior to build more effective robot learners," *Artificial Intelligence,* vol. 172, no. 6-7, pp. 716-737, 2008.

[57]   A. L. Thomaz and C. Breazeal, "Reinforcement learning with human teachers: Evidence of feedback and guidance with implications for learning performance," in *Aaai*, 2006, vol. 6, pp. 1000-1005: Boston, MA.

[58]   W. D. Smart and L. P. Kaelbling, "Effective reinforcement learning for mobile robots," in *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292)*, 2002, vol. 4, pp. 3404-3410: IEEE.

[59]   S. Schaal, "Dynamic movement primitives-a framework for motor control in humans and humanoid robotics," in *Adaptive motion of animals and machines*: Springer, 2006, pp. 261-280.

[60]   S. Mann, "Wearable computing: Toward humanistic intelligence," *IEEE Intelligent Systems,* vol. 16, no. 3, pp. 10-15, 2001.

[61]   M. Minsky, R. Kurzweil, and S. Mann, "The society of intelligent veillance," in *2013 IEEE International Symposium on Technology and Society (ISTAS): Social Implications of Wearable Computing and Augmediated Reality in Everyday Life*, 2013, pp. 13-17: IEEE.

[62]   E. M. Palmer, T. S. Horowitz, A. Torralba, and J. M. Wolfe, "What are the shapes of response time distributions in visual search?," *Journal of experimental psychology: human perception and performance,* vol. 37, no. 1, p. 58, 2011.

[63]   W. E. Hockley, "Analysis of response time distributions in the study of cognitive processes," *Journal of Experimental Psychology: Learning, Memory, and Cognition,* vol. 10, no. 4, p. 598, 1984.

[64]   *Georgia Institute of Technology Institutional Review Board Protocol H20340*, 2020.

[65]   bluedesigns.com, "<Blue_Yeti_QuickStartGuide-EN.pdf>."

[66]   H. Pham. (2006). *PyAudio 0.2.11* [Web]. Available: https://pypi.org/project/PyAudio/

[67]   J. M. Kates, *Digital hearing aids*. Plural publishing, 2008.

[68]   F.-L. Luo, B. Edwards, J. Yang, and N. Michael, "FFT-based technique for adaptive directionality of dual microphones," ed: Google Patents, 2003.