Master Thesis

Precise Key Triggering for Piano Sample Recordings

Marcelo Chulek

Examiner: Prof. Dr. Moritz Diehl
Adviser: Prof. Dr. Leonhard Reindl

Albert-Ludwigs-University Freiburg
Faculty of Engineering
Department of Microsystems Engineering (IMTEK)
Master in Embedded Systems Engineering

September 20th, 2019

Writing period

01. 04. 2019 – 20. 09. 2019

Examiner

Prof. Dr. Moritz Diehl

Advisers

Prof. Dr. Leonhard Reindl

Declaration

I hereby declare, that I am the sole author and composer of my thesis and that no other sources or learning aids, other than those listed, have been used. Furthermore, I declare that I have acknowledged the work of others by providing detailed references of said work.
I hereby also declare, that my thesis has not been prepared for another examination or assignment, either wholly or excerpts thereof.

Place, Date	Signature

Abstract

In the music industry, sample libraries are digital collections of instrument recordings. They aim to provide the full range of sounds created by a musical instrument. Creating a piano sample library is a time consuming and repetitive process, which demands a trained pianist to play the piano keys over long periods. Human players tend to make mistakes while playing and on top of that, they can add unwanted noise to the recordings. Creating a piano sample library requires a trained pianist to play the 88 keys of the piano at as many as 20 loudness levels. Additionally, it requires playing different types of sounds produced by the piano, such as sustain resonances, sympathetic resonances, release sounds, and key-bed noises. The goal in this work is to describe, design, develop, test and discuss a state-of-the-art system to precisely trigger piano keys to create professional piano sample libraries. The system is capable of triggering the piano keys at different loudness levels as well as estimate start and end markers used for sample cropping. Moreover, the system allows achieving all types of sounds required for the creation of piano sample libraries. It focuses on pianos, but can potentially be used with other keyboard-based instruments, such as synthesizers, electric pianos, among others. The system is required to be non-invasive, i.e. no sensors are inserted into the instrument mechanics. In short, the system is composed of a linear actuator controlled by a Proportional-Integral-Derivative (PID) controller and connected to the main control software. The linear actuator is positioned vertically above the piano key with a noise-protected rack mount. The actuator's vertical slider movement imitates the behavior of a trained pianist finger, meeting constraints in position, speed, acceleration and time. Simultaneously, a microphone captures the sounds produced by the piano and several types of analyses are performed, both in the spectral and temporal domain.

Zusammenfassung

Sample Libraries sind in der Musikindustrie digitale Sammlungen von Instrumentenaufnahmen. Sie zielen darauf ab, die gesamte Bandbreite der von einem Musikinstrument erzeugten Klänge bereitzustellen. Das Erstellen einer Piano-Sample-Bibliothek ist ein zeitaufwändiger und sich wiederholender Vorgang, bei dem ein ausgebildeter Pianist über längere Zeiträume auf den Klaviertasten spielen muss. Menschliche Spieler neigen dazu, beim Spielen Fehler zu machen, und außerdem können sie den Aufnahmen unerwünschte Geräusche hinzufügen. Um eine Piano-Sample-Bibliothek zu erstellen, muss ein ausgebildeter Pianist die 88 Tasten des Pianos mit bis zu 20 Lautstärken spielen. Außerdem müssen verschiedene Arten von Klängen abgespielt werden, die vom Klavier erzeugt werden, z. B. Sustain-Resonanzen, sympathische Resonanzen, Release-Klänge und Tastenbettgeräusche. Ziel dieser Arbeit ist es, ein System nach dem neuesten Stand der Technik zu beschreiben, zu entwerfen, zu entwickeln, zu testen und zu diskutieren, um Klaviertasten präzise auszulösen und professionelle Klaviersample-Bibliotheken zu erstellen. Das System ist in der Lage, die Klaviertasten mit unterschiedlichen Lautstärken zu triggern sowie Start- und Endmarkierungen für das Beschneiden von Samples zu schätzen. Darüber hinaus können mit dem System alle Arten von Klängen erzielt werden, die für die Erstellung von Piano-Sample-Bibliotheken erforderlich sind. Es konzentriert sich auf Klaviere, kann aber möglicherweise mit anderen Instrumenten auf Tastaturbasis verwendet werden, wie z. B. Synthesizern und elektrischen Klavieren. Das System muss nicht-invasiv sein, d. H. Es sind keine Sensoren in die Instrumentenmechanik eingesetzt. Kurz gesagt, das System besteht aus einem Linearantrieb, der von einem PID-Regler gesteuert und mit der Hauptsteuersoftware verbunden wird. Der Linearaktuator ist vertikal über der Klaviertaste mit einer geräuschgeschützten Rack-Montage positioniert. Die Bewegung des vertikalen Schiebereglers des Aktuators imitiert das Verhalten eines ausgebildeten Pianistenfingers und erfüllt Einschränkungen in Bezug auf Position, Geschwindigkeit, Beschleunigung und Zeit. Gleichzeitig erfasst ein Mikrofon die vom Klavier erzeugten Klänge und es werden verschiedene Arten von Analysen sowohl im spektralen als auch im zeitlichen Bereich durchgeführt.

Contents

List of Figures

List of Tables

List of Algorithms

Symbols and Notation

A_a	Attack Amplitude
A_d	Decay Amplitude
A_r	Release Amplitude
A_s	Sustain Amplitude
t_a	Attack Time
t_d	Decay Time
t_r	Release Time
t_s	Sustain Time

1 Introduction

Due to their flexibility and sophistication, pianos are one of the most important and sought-after instruments in classical and modern music. The piano is an evolution of its ancestor's harpsichord and clavichord [1]. Both ancestors rely on a simple action mechanism composed by a lever fixed in the middle point, which is pressed by the player's finger on one side and collides with a brass or metal string on the other side. The creation of the first piano with considerable similarities to the modern versions was done by Cristofori and dates back to the beginning of the 18th century. The main difference between the piano and its ancestors lays the action mechanism, which for the piano involves is rather complex and involves some key components, such as hammer, string, damper, pedal, and key. The piano action mechanism is explained in detail in Chapter 3.

By being one of the most popular musical instruments in western music, the pianos have been extensively studied over the years. Depending on specific applications, several research topics are described in the literature, such as control systems for automatic piano playing, physical modeling of the piano sounds, among many others. Although pianos have been extensively studied, questions often arise or have not been explained satisfactorily [2]. This is mostly a result of its highly elaborated and complex action mechanism. Moreover, new applications involving pianos often arise, such as the one presented in this thesis: a precise key triggering system used for creating piano sample libraries. Such a system as not been described in the literature.

The piano's size, weight, and price lead to the development of digital counterparts, which intend to emulate the sounds of acoustic versions as closely as possible [3]. Examples include e-pianos, MIDI controllers associated with software samplers, keyboards, electric pianos, and synthesizers. The sounds created by digital counterparts are either sample-based or synthesis-based. The sample-based instruments rely on sounds files recorded in a real piano, the so-called piano samples. When played, these instruments reproduce certain sample files of a piano sample library. In a synthesis-based instrument, the sounds are produced in real-time by a physical model of the piano and do not depend on any sample files.

Piano Sample Libraries are often composed of different types of samples, such as samples of full-length notes, samples of notes with sustained resonances, samples of notes with sympathetic resonances, samples of release sounds and even samples of key-keybed collision sounds. Moreover, these samples are available at multiple loudness levels. To illustrate its dimension, an arbitrary piano sample library could contain 130 samples per key and 11440 samples per piano. These values can increase depending on the desired authenticity. In the sample-based counterparts of the piano, the samples are selected and played depending on which keys are pressed, how fast the keys are pressed, how long the keys are kept down, how fast the fingers are removed from the key, and if the sustain pedal is kept down. Recording samples is a somewhat complex process that depends on factors such as acoustics of the studio, proper instrument tuning, correct positioning of microphones, proper noise treatment, correct synchronization between numerous audio equipment, player mistakes, among many others [4]. The complete process of creating a sample library is typically performed by trained personnel and could take as long as days depending on several factors. Even trained players tend to make mistakes and recording sessions contain multiple repetitions of moments. In this context, a precise key triggering system could aid the recording of professional piano samples by increasing the quality of the samples as well as reducing the amount of time required for its creation.

This thesis describes a precise key triggering system for the creation of piano sample libraries based on a linear actuator installed onto the key's surface and a microphone used for capturing the sounds. With this basic setup, the system is capable of estimating the multiple dynamic levels of different piano keys. Moreover, the system can trigger the different types of sounds required for the creation of piano sample libraries. It also estimates important temporal events, such as the start and end of the samples, a piece of information that is used for sample cropping and that cannot be achieved with a human player. Also, the detailed system is designed to be non-invasive, i.e. it does not require the installation of any sensors inside the piano mechanics, a disadvantage that could potentially lead to changes in the characteristics of the sound. The described system is also thought to operate with different types of pianos, such as upright piano and grand piano. Furthermore, it can also be potentially used with other keyboard-based instruments depending on their construction characteristics, for example, synthesizers and electric pianos.

In a nutshell, the control software commands an external embedded hardware implementing PID control for the precise positioning of a linear actuator. The movement of the actuator slider over is described in terms of ADSR curves, which

represent a sequence of position set-points to be followed by the PID controller over time. The execution of a ADSR curve results in the movement of the piano key and, as a consequence, a sound is obtained. The control software then analyzes the sound captured by a microphone both in the temporal and spectral domains. Depending on the result of the analysis, the control software decides how to proceed.

The system has two main operation modes: calibration and normal operation. During calibration, the system estimates the loudness profile of a certain key by triggering it using different movements of the linear actuator. Next, the system fits the calibration data into a model, from where it derives the optimal movements to be executed by the linear actuator. During normal operation, the system simply executes the optimal movements estimated during the calibration.

Chapter 2 gives a general literature review from past to present. This chapter also presents a general overview of the typical research approaches used as well as a complete description of the state of scientific development in the topic of this thesis. Over the years, many technical approaches were used to describe the behavior of the piano. Most of them focus on the measurement of physical parameters, such as position, speed, acceleration, forces and contact times. This chapter also shows several applications involving the piano, such as automatic piano playing, physical modeling of piano sounds, among others.

Chapter 3 introduces the basic terminology required to comprehend the pianos. Since this work covers a musical instrument, many musical terms and parts of the piano are introduced and explained using a technical nomenclature. Terms like keyboard, action, hammer, strings, soundboards, pedal, dynamics, ADSR, velocity, escapement point, sustain resonances, sympathetic resonances, release sounds, noises, and sample libraries are carefully explained.

Chapter 4 focuses on the research approach used in this work. The design of a state-of-the-art system to precisely trigger piano keys are discussed in detail. This chapter presents the specification, architecture as well as the description of the main components. It includes the following main topics: position control using linear actuators, PID controllers, audio descriptors for loudness estimation, audio descriptors for brightness estimation, system modeling using Linear Least Squares (LLS), estimation of optimal dynamic levels, system synchronization, among other topics. Some important algorithms are described in detail, such as the algorithm for detection of the quietest sound and algorithm for estimation of the keybed bottom distance.

Chapter 5 presents several experiments, as well as a detailed discussion on their

results. An extensive analysis of the relation between the loudness and internal parameters of the ADSR curves is provided. The spectral centroid is analyzed as an audio descriptor of brightness. Estimation of the start point of a sample using the spectral disparity is assessed. Evaluation of the keybed bottom distance algorithm is performed. Different types of regression models are compared. The estimation of the escapement point is evaluated. The effects of different ADSR curves are discussed in detail. The behavior of black keys is analyzed. PID tuning is evaluated.

Chapter 6 presents the conclusions of this work. The advantages and disadvantages of the system are discussed in detail. In the end, the current challenges are presented as well as suggestions for future work on this topic.

2 Related Work

Research involving pianos dates back as early as the beginning of the 19th century. In the early research times, studies were mostly related to the physics of the piano action. Later on, research started to focus on the interaction between the player and the instrument, such as the biomechanical forces between the finger and the key, applications involving automatic piano playing, among many others. More recent research topics include physical modeling of the piano for sound emulation, as well as the topic of this thesis: automatic piano sampling. Research involving the use of a linear actuator to sample a piano has not been published as a thesis before.

Since the piano is a highly elaborated instrument containing a complex action mechanism with many components, not all aspects are fully described in the literature. In addition to that, different types of pianos have different action mechanisms and on top of that, even pianos with similar construction can have considerable changes in sound. Moreover, new research methods arise depending on new applications found for the instrument. This chapter describes the most common research approaches used for describing the piano. It also introduces typical types of technical applications involving the piano. Finally, it introduces the scarce material available in the field of this thesis.

Many books describe how to play the piano artistically correct, such as "The Technique of Piano Playing" by Gát [5] and the "Technik des Klavierspiels" by Kratzert [6]. The majority of these books debate about how to correctly press a key at different dynamic levels. One general conclusion is that the fingers of professional players tend to spend less energy while moving the piano key if compared to amateur pianists. The aforementioned debate led to many technical works, such as the one by Askenfelt described later in this chapter, among others works. These works are strongly related to the topic of this thesis since it comprises pressing the piano keys in the right manner.

Many approaches focus on remote playing the piano rather than precisely triggering the piano key, such as the one described by Risset et. al [7] [8] [9] [10]. These works do not focus on triggering a single piano key at multiple dynamic levels, but instead on

how to execute sequences of notes such as in musical pieces. Risset et al., for instance, focused mostly on musical performances with a computer-controlled piano. Remote playing the piano can be understood as the sequential triggering of one or more piano keys. Their study does not describe how to precisely trigger a key, treating it rather in a key-on, key-off approach.

Another line of work includes physically modeling the piano action mechanism. Several physical models available in the literature describe the kinematics of the piano action components, ranging from simple to complex, such as the one described by Thorin with only 1-degree of freedom [11]. The results of the proposed model are evaluated by comparison with the measurements obtained from sensors installed in the piano action mechanism. This work suggests that force-driven simulation models evaluated by position measurements do not validate dynamic models of pianos. Instead, position-driven models evaluated by force measurements are potentially better options.

Other related research approaches comprise modeling the complete piano in the time domain [12]. This is performed by modeling the piano soundboard as a bi-dimensional vibrating plate using the Mindlin–Reissner plate theory. Strings are modeled considering the longitudinal waves created by the hammer collision as well as the shear waves. The hammer-string collision forces are modeled using a non-linear function of the hammer compression. This study proposes the use of non-linear differential equations to describe the sound pressure changes fields. These equations are highly dependent on the geometry of the piano model and, as a consequence, do not generalize well since different pianos cannot be described by the same equations.

Askenfelt et al. focused on describing the contact duration between the different piano action components [13]. These include key-keybed contact duration, hammer-string contact duration, the time interval between the escapement of the hammer and hammer-string collision and the hammer-string contact duration. In this work, it is also described the motion of hammer and key for different dynamic levels. Askenfelt et al. also investigate string oscillations as well as the impact of the hammer hardness and weight in the spectrum. The results of this work are widely used for the hardware specification of the developed piano sampler system.

Similarly to Askenfelt et al., Goebl also focused on analyzing the temporal behavior of piano action components in different types of pianos. His works studied the movement of keys and hammers over the entire dynamic range, i.e. ranging from pressed touch to struck touch [14]. The results of the experiment include temporal estimation of the finger–key contact, hammer–string contact, and key-keybed bottom

contact as well as the maximum hammer speed. Goebl's also studied the delay between the finger-key contact and the key-keybed bottom, as it is explained later in Section 3.12.

Another related study showed evidence that the timbre and dynamics of different pianos are not only determined by the speed at which the hammer hits the string [15]. An experiment was performed between a group of pianists and suggested that finger-key contact noise, as well as the key-keybed contact noise, also account for the changes in the perceived timbre. In the field of piano sampling, the finger-key contact noise is not significantly loud and does not need to be sampled. Moreover, the finger-key contact noise cannot be reproduced with a MIDI keyboard, so even if the finger-key contact noise would be sampled, a piano sampler would not be able to reproduce it properly. The key-keybed bottom noise is considerably louder and can change the perception of a sound. In the framework of this thesis, the piano sampler aims to record the key-keybed bottom noise separately.

In a work related to tone changes at different dynamic levels, Susuki focused on checking spectral differences between pressed and struck touches in different keys [16]. Pressed touch can be understood as moving the key with the finger initially resting on the key surface while struck touch involves hitting the key from a certain distance above the surface. It was suggested that the spectral changes depend on the key involved and are in general small, i.e. some keys produced minimal spectral changes while others do not. One conclusion of this work is that spectral changes become evident only for certain keys.

Flückkiger et. al introduced a strain gauge-based load cell setup that can be installed in the piano key mechanism [17]. This is an invasive method and allows measuring the forces involved in the piano action up to a frequency of 1000 Hz. This study showed that different types of pianos require different setups for correct parameter estimation. The results of this study support the use of non-invasive methods to generalize well over different types of pianos.

Kinoshita focused on the bio-mechanics and finger forces [18]. Different finger force profiles in struck and pressed touches were analyzed. The obtained results relate key displacement, key speed, key acceleration and finger forces over time and for different fingers. One important finding is that pressed touch has a smooth force profile over time with only a maximum peak while the struck key touch has multiple force peaks and a noisier force profile [19]. The maximum forces involved are comparable for the struck and pressed touch. To illustrate this experiment's results, pressed touch has a force spectrum composed mostly by low-frequency components while the struck

touch contains low and high-frequency components.

In a more practical approach, the Vienna Symphonic Library offers a sample library created using an actuator system [20]. The samples were recorded with an automatic trigger system at multiple dynamic levels, although no technical details are available in the literature [20]. Also, it lacks evidence on whether or not the obtained piano library increased the quality of the samples.

Alternatively, synthesis-based softwares implement physical modeling of a piano, thus allowing a player to create piano sounds from a model in real-time. Given the complex tonal changes for different hammer speeds during a collision, as well as for different finger-key and key-keybed noises, a physical model software running in real-time would, at best, match the results of a sample-based software. An example of synthesis-based software is the Pianoteq by Modartt [21].

Other piano samplers available in the market allow triggering MIDI instruments via MIDI input port. By sending MIDI message at different velocity levels, one can record different samples of an instrument. As an example, a product entitled Sample Robot allows the recording of instruments via MIDI [22]. However, these types of solutions do not allow triggering a piano unless a proper MIDI interface is available. An example of piano with MIDI input port is the Yamaha Disklavier Enspire [23], which relies on an actuator system that is installed directly under the piano action mechanism.

3 Background

This chapter presents the fundamental theoretical background required to understand the precise key triggering system for piano sample recordings. The main concepts related to this work are explained in detail, being most of them related to the piano working principle. Important parts of the piano are discussed in detail, such as the hammer, key, strings, and soundboard. Some musical terms like dynamic level, expression, decay, sustain, release are interpreted in a technical approach. The concepts presented in this chapter are finally connected in Chapter 4 within the scope of this work.

3.1 Key Layout

The piano is a keyboard instrument, i.e. an instrument that is played using a keyboard. The keyboard is a row of levers that are pressed by the player's fingers. The piano keyboard is commonly organized in a layout composed of 88 keys located next to each other. Moreover, the piano keyboard is divided into groups of 12 keys, named octaves. Each octave is composed of 7 white keys and 5 black keys. The 7 white keys in an octave produce the notes C, D, E, F, G, A, and B. The 5 black keys either raise or lower half a tone of the adjacent key and are typically named with the term sharp or flat. The keys are normally displayed with a sub-index representing its octave, such as C_1, which means note C in the first octave. The piano keys can also be assigned with a unique identification number ranging from 1 to 88. Throughout this study, this nomenclature will be used while indicating a certain key. The typical keyboard layout is displayed in Figure 1.

Every key produces a musical note, which is a composition of a fundamental frequency and a series of harmonic frequencies. The fundamental frequency can also be named with the term pitch. Each key is tuned considering a reference pitch, typically 440 Hz, which is standardized in the ISO 16. It is important to notice that the fundamental frequencies of the keys increase exponentially over the keys and are related to the human perception of pitch. Equation (1) shows the relation between

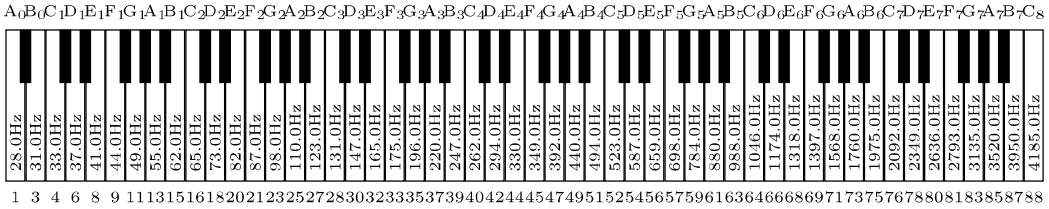


Figure 1: Standard Keyboard Layout Above the key, the note name and a sub-index representing its octave. For instance, C_1 stands for the note C in first octave. The fundamental frequency of each key is displayed on the key surface. For instance, the fundamental frequency of the note C_1 is 33 Hz. Below the key it is presented an unique identification number ranging from 1 to 88. To illustrate, the identification number of the key C_1 is 4.

the fundamental frequency of a key and the reference frequency.

$$f(n) = \left(\sqrt[12]{2} \right)^{n-49} \cdot \mathrm{f_r} \tag{1}$$

Where:

- $f(n)$: Fundamental frequency of the nth key, in Hz
- n: nth key, i.e. n = 1, 2, ... , 88
- $\mathrm{f_r}$: Reference frequency used for tuning, typically 440 Hz

Alternatively, Equation (1) can be rewritten as a function of $f(n)$ and $\mathrm{f_r}$ by isolating n. This is displayed in Equation (2).

$$n(f) = 12 \log_2 \left(\frac{f}{\mathrm{f_r}} \right) + 49 \tag{2}$$

When a certain key is pressed, the spectrum of the produced sound has harmonic behavior, i.e. the spectral components are equally spaced in frequency and are multiples of the fundamental frequency of the note played. This spectral behavior is strongly related to the vibration of strings fixed at both ends as it is explained in Section 3.3. An example of a spectrum from a sound produced by the piano is displayed in Figure 2.

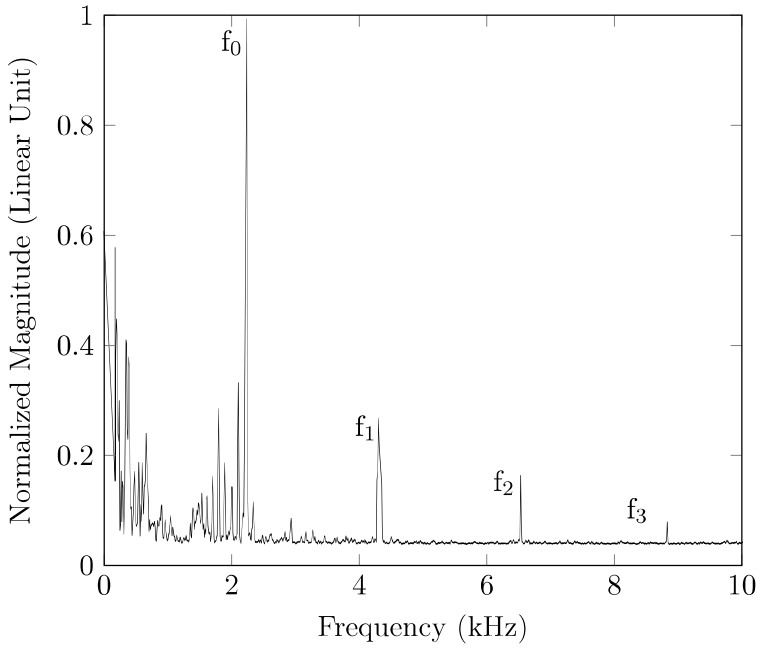


Figure 2: Spectrum of a note C_7 Spectral estimation of sustained part of a note C_7 recorded in a Yamaha Upright Piano. The fundamental frequency of the note C_7 is 2.092 kHz. The harmonics are located at frequencies multiple of the fundamental frequency, for example, 4.184 kHz, 6.276 kHz and so on. The plot is normalized in respect with the most powerful bin of the spectral estimation.

3.2 Action

The expression allows a keyboard musical instrument to achieve different sounds in response to changes in velocity, pressure or other variations related to the key. Instruments can be classified according to their expression in velocity-sensitive, after-touch sensitive and displacement sensitive. The piano is an instrument sensitive to changes in velocity, which means, that different sounds can be achieved depending on how fast a key is pressed down. Differently, some other instruments like synthesizers and MIDI controllers, under certain circumstances, allow achieving different sounds by changing the pressure between the finger and the key, a feature named after-touch. Finally, displacement-sensitive instruments achieve different sounds by different vertical displacements of the key. Examples are tracker pipe organs and electronic organs, where partial displacement are used to control airflow.

The piano action Mechanism is a mechanical system that converts vertical displacement of a key into the movement of the hammer, the damper, and several other components. Every key has an individual action mechanism and, as a consequence, a piano has typically 88 key action mechanisms. The action mechanism defines the types of pianos. The most common types of pianos are the Upright Piano and the Grand Piano.

As the name suggests, an upright piano is a piano with the hammers and strings positioned vertically. In contrast, a grand piano has the hammers and the strings positioned horizontally. The fundamental difference between both types of action lays in the use of gravitational forces. In a grand piano, the action components return to the original rest position by using the acceleration of gravity, which is uniform across all the keys actions. In an upright piano the action components return to the rest position by using springs, which might have differences between each other. Due to this dissimilarity in construction, both pianos might have differences in the acoustics [24]. Figure 3 displays the main components of a grand piano. The action mechanism of an upright piano is shown in Figure 4.

The hammer, as the name suggests, is a hammer-shaped wooden core covered with dense felt material. When the key is pressed in a grand piano, the capstan forces a complex mechanism involving jack, whippen and other components upwards. This results in the upward movement of the hammer. Simultaneously, the damper is moved away from the string. Finally, the sound is created by the collision between the hammer and the string. After the collision with the string, the hammer returns towards its rest position until it is stopped by the back check. The damper also returns

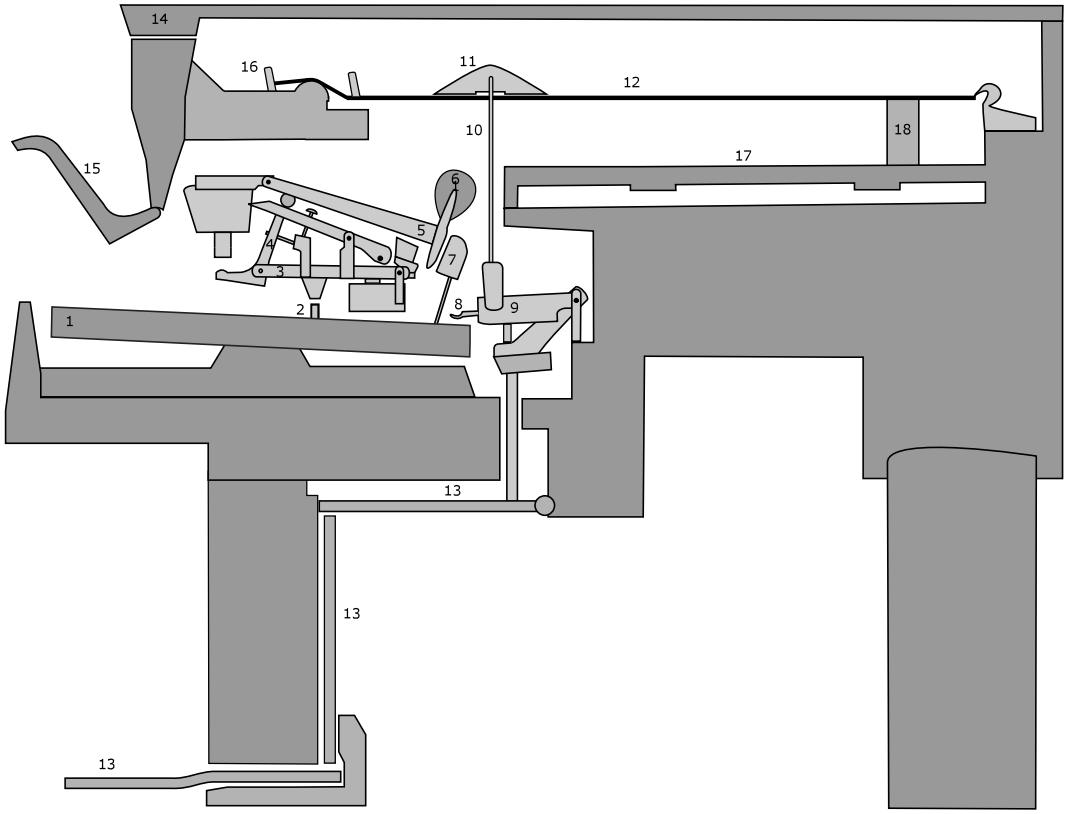


Figure 3: Grand Piano Main Components (1) Key, (2) Capstan, (3) Whippen, (4) Jack, (5) Hammer Shank, (6) Hammer, (7) Back check, (8) Spoon, (9) Damper Mechanism, (10) Damper Rod, (11) Damper, (12) String, (13) Pedal Mechanism, (14) Piano Lid, (15) Lid, (16) String Pin, (17) Soundboard, (18) Bridge.

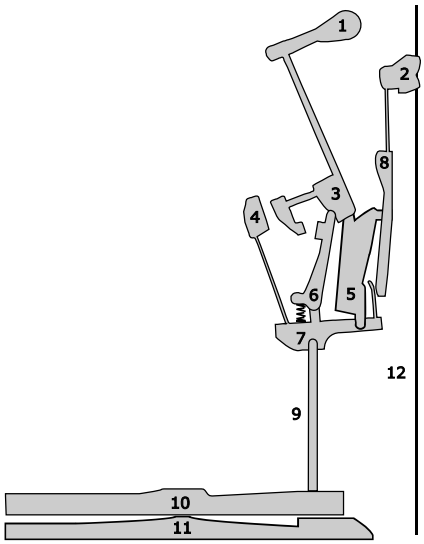


Figure 4: Key Action of an upright piano (1) Hammer, (2) Damper,(3) Hammer Tail, (4) Back check, (5) Main Action Rail, (6) Jack, (7) Action Lever, (8) Damper Lever, (9) Extension Rod, (10) Key (11) Balance Rail, (12) String.

to its rest position, i.e. touching the strings and damping the sound. Alternatively, the sustain pedal overrides the function of the dampers, i.e. it keeps all piano dampers away from the strings. A highly simplified illustration of the action mechanism in a grand piano is displayed in Figure 3.

The hammer-string collision has a considerable impact on the characteristics of the obtained sounds. The hardness or softness of a hammer changes the spectrum of the obtained sounds [25]. Soft hammers tend to have long hammer-string collision and more powerful harmonic components. Differently, hard hammers tend to have short hammer-string collision and less powerful harmonic components [25] [1]. These changes in hardness can be seen in the power of spectral components and are related to the timbre of the sound (or tone color) and are sensitive to the properties of the hammer. The effect of different hammers on the power of the harmonic components is displayed in Figure 6.

It is important to notice that the power of the spectral components is not only determined by the hardness of the hammer but also by the hammer-string collision time. The hammer-string collision time is related to the dynamics that the key is played. As a consequence, higher dynamic levels, i.e. louder sounds, have harmonic components with increases power when compared to lower dynamic levels, i.e. quieter sounds. In the field of audio engineering, this phenomenon is often described with

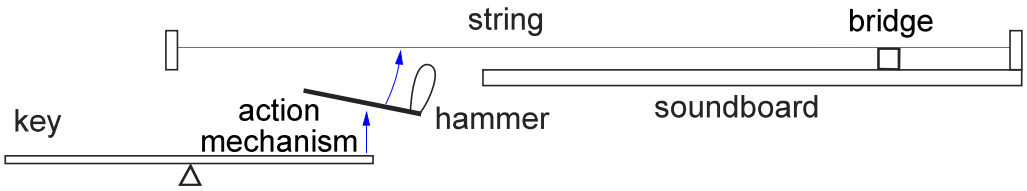


Figure 5: Simplified action mechanism in a Grand Piano [1] The key forms a lever that moves the action mechanism upwards. As a consequence, the hammer moves towards the string. Upon collision, the string vibrates and creates a sound. The sound created by the string is transferred to the soundboard by the bridge. The soundboard behaves as a vibrational plane and amplifies the sound produced by the string due to its large area. The blue arrows represent the movement of the action mechanism before the hammer-string collision.

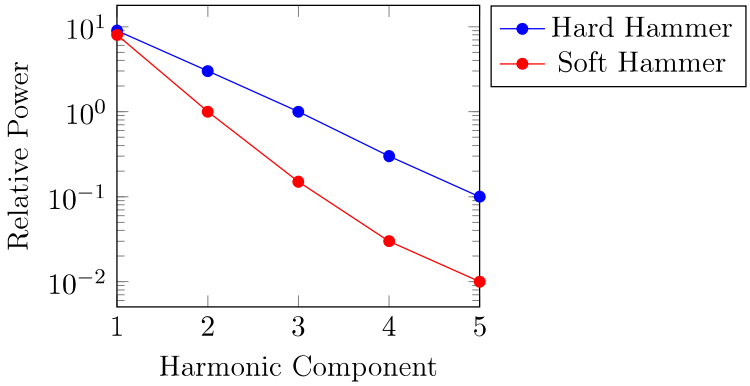


Figure 6: Relative power between harmonics for different hammer hardness [1] Effect of hammer hardness in the harmonics of a note C_4 played at mezzoforte (mf) in a Grand Piano, as described by Giordano [1]. To illustrate, for a hard hammer the third harmonic is approximately ten times less powerful than the first harmonic, i.e, the fundamental component.

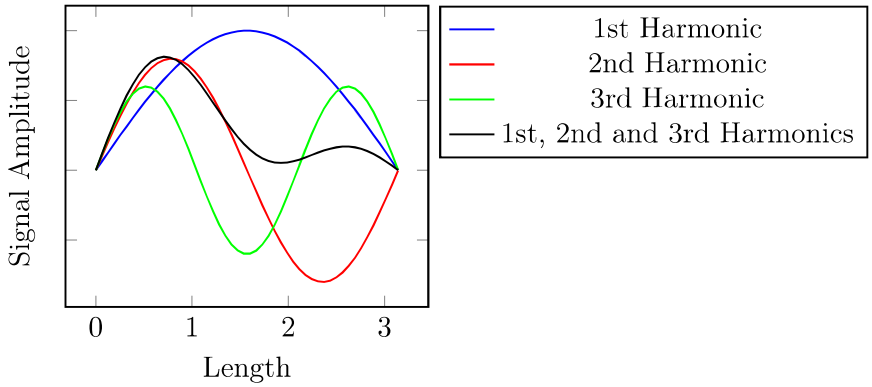


Figure 7: String vibration. Hypothetical situation of a string with length L = π fixed at both ends and composed of 3 harmonics.

the term "changes in overtones".

3.3 Strings

Pianos produce sounds based on the principle of transverse oscillations of strings fixed at both ends. When transverse oscillations are produced in a stretched string fasten at both ends, standing waves are created. Standing waves are waves that oscillate in time although their envelope profile remains fixed. Standing waves contain nodes and at the nodes, no amplitude changes are observed. After the excitation, a standing wave is formed and only inter-nodal distances multiple of half-wavelengths are allowed. If the excitation happens at nodal points, no oscillations are produced. In a piano, when a string is excited by the hammer, it will vibrate at many of its harmonic frequencies at the same time. The hypothetical situation of 3 harmonics combined is displayed in Figure 7.

The exact frequencies that will compose final string vibration depend on where the excitation takes place. Typically the excitation by the hammer happens at $\frac{1}{8}$ of string length L [1]. The equation relating the length of a string L to the wavelength lambda and the harmonic index n is shown in Equation (3).

$$L = n \frac{\lambda}{2} \quad n = 1,\ 2,\ ... \tag{3}$$

Where:

L: String Length.
n: nth Harmonic.
λ: Wavelength.

The relation between frequency, propagation speed and wavelength is shown in Equation (4).

$$f = \frac{v}{\lambda} \tag{4}$$

Where:
f: Frequency.
v: Propagation speed.

It is possible to calculate the frequency of the harmonic components by combining Equation (4) with Equation (3). The result can be seen in Equation (5).

$$f_n = n \frac{v}{2L} : n = 1,\ 2,\ ... \tag{5}$$

Where:
f_n: Frequency of the nth harmonic component.

The propagation speed v can be expressed in terms of tension force F_t and linear density μ as it is shown on Equation (6).

$$v = \sqrt{\frac{F_t}{\mu}} \tag{6}$$

Where:
F_t: Tension force
μ: Linear density of the string

By inserting Equation (6) into Equation (5), one can obtain Equation (7).

$$f_n = \frac{n}{2L}\sqrt{\frac{F_t}{\mu}}\ n = 1,\ 2,\ 3,\ ... \tag{7}$$

Equation (7) shows that the frequencies produced by a string fixed at both ends depend on the linear density μ of the string, the tension force F_t and the length L.

Note	Fundamental Frequency (Hz)	Diameter (mm)
A_3	220	1.0
C_4	262	1.0
A_4	440	0.95
C_5	523	0.95
A_5	880	0.91
C_6	1047	0.86
A_6	1760	0.86
C_7	2093	0.81
A_7	3520	0.76
C_8	4186	0.76

Table 1: String Diameter. Example of string diameters for different keys in a grand piano [1]. The column named frequency refers to the fundamental frequency of the key.

The linear density μ is related to the thickness of the string while the tension force F_t can be adjusted by the tuning pins. To illustrate, the lower keys of the piano have low tension, high linear density, and long length while the higher keys have high tension, low linear density and a short length. Some typical values for string thickness in a grand piano can be seen in Table 1.

Another important characteristic of pianos is that every hammer collides with a group of strings. For the lower keys, the hammer typically collides with one or two strings while for the higher keys the hammer collides with three strings. These groups of strings, excited by a single hammer, have similar values in length and tension, although these values are not equal. As a consequence, the sounds produced by the set of strings have harmonic components close in frequency. This small difference in frequency leads to alternation in time between constructive interference and destructive interference and, as a result, a wave with time-varying amplitude. For a listener, this is perceived as changes in loudness of the sound and is often described with the term beats. This phenomenon contributes to the richness of the obtained sounds and has to be taken into account while estimating the loudness later in Chapter 4. This phenomena is illustrated in Figure 8.

A modern Steinway Grand Piano has 243 strings and can produce frequencies ranging from 27.5 Hz (A_0) to 4186 Hz (C_8). Old pianos, such as the one created by Cristofori in 1720, could only produce frequencies ranging from 65 Hz to 1047 Hz [1]. As a reference, the hearing of young healthy adults ranges from approximately 20 Hz to 18 kHz. The piano strings have an overlapping arrangement to allow longer

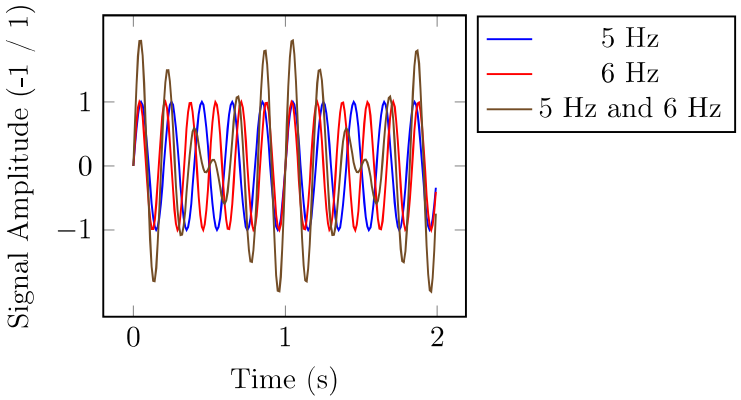


Figure 8: Beats Beats result from the superposition of two waves with identical amplitudes and slightly different frequencies. In this example, two sine waves, one with 5 Hz and another with 6 Hz, add up and form a signal with amplitude changing over time. For a listener, this is perceived as changes in the loudness of sound over time.

strings to fit into the frame. This is especially true for the bass strings, which have a considerably longer length. Moreover, the bass strings are placed towards the middle of the piano to increase the resonances in the soundboard. Figure 9 displays the typical arrangement of the piano strings in a grand piano.

The length of the piano strings ranges from approximately 5 cm to 1,6 m. To avoid the hypothetical situation where the bass strings would be too long to fit into the piano frame, piano makers take advantage of the linear density by using thicker and heavier strings for the lower notes. Figure 3 displays different string lengths for bass and treble strings in a grand piano.

3.4 Soundboard

The soundboard is a large thin wooden plate located under the piano strings with a thickness of approximately 1 cm. Since the strings do not have a large cross-section area, the amount of air pressure changes caused by them is limited. The sounds produced by the strings are couple into the soundboard by a component named bridge. Finally, the soundboard amplifies the sounds due to its large area and the amount of air it can displace. Similarly to the strings, the soundboard can achieve different vibrational modes. Some vibration modes of the soundboard are illustrated in Figure 11.

By behaving as a sound amplifier, the soundboard has a certain frequency response.

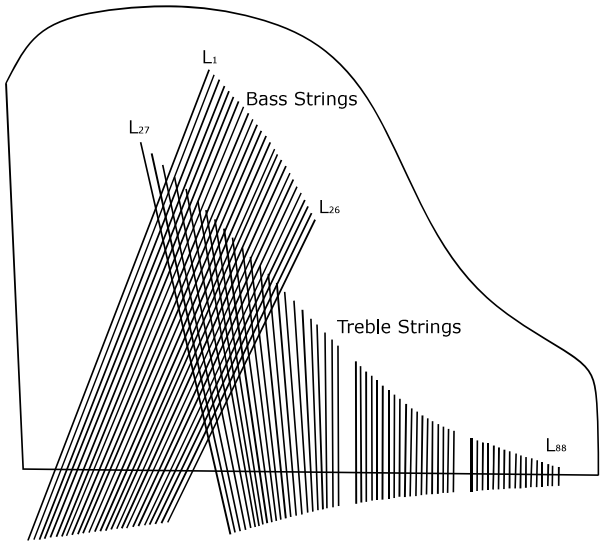


Figure 9: Typical arrangement of piano strings Thicker strings produce bass sounds, i.e. sounds with low fundamental frequencies. Thinner strings produce treble, i.e. sounds with high fundamental frequencies. For the treble strings, three strings are strung by a single key, and for bass strings, the number of strings strung by a key decreases from three, to two, and finally to one in the lowest keys.

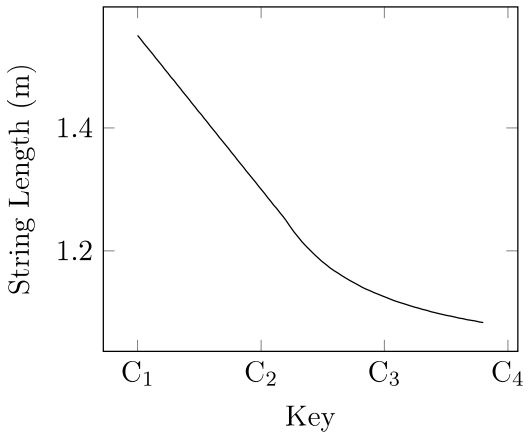


Figure 10: String length [1] Approximate string length for a grand piano Steinway S. The longest string has length of approximately 155 cm.

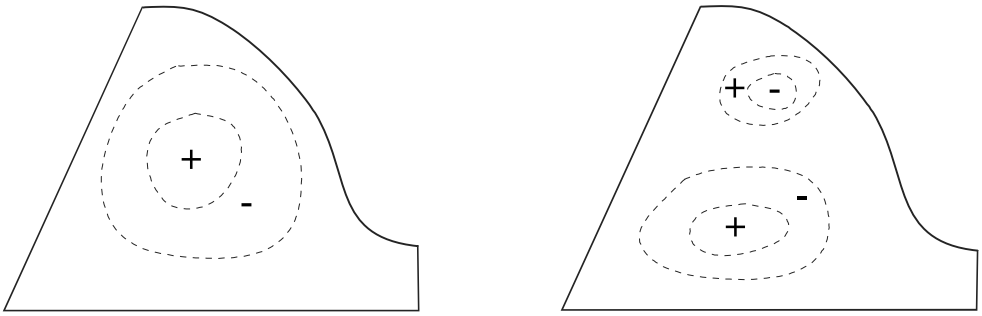
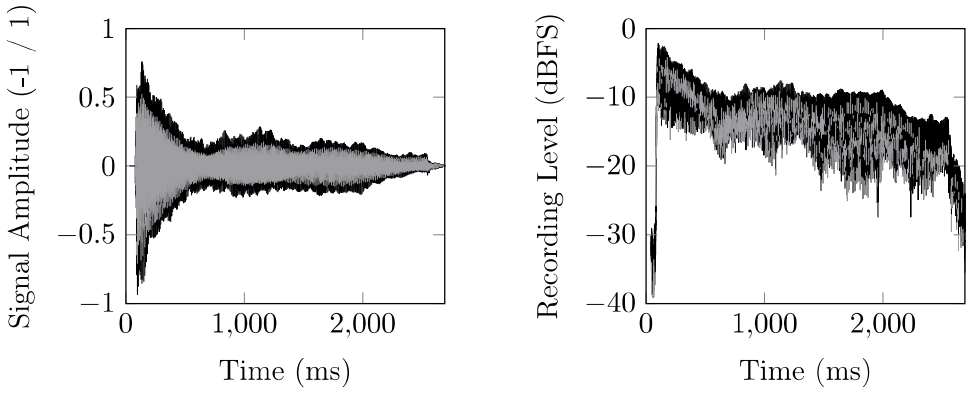


Figure 11: Sound Board Example of the first and second vibrational modes of a soundboard [1]. The + and - represent the board vibrational amplitude in different directions. The dashed lines represent constant vibrational amplitudes.

Giordano performed several experiments to characterize the frequency response of soundboards for an Upright Piano [26]. This study was performed by analyzing the relation between sound pressure changes and surface velocity. This relation gives an estimation of the efficiency of sound production by the soundboard. A general conclusion of his work is that frequencies below 500 Hz and above 5 kHz are damped by the soundboard [26].

3.5 Sound Measurement

Sound pressure is the local pressure deviation from the ambient atmospheric pressure. Sound pressure is measured in Pascals and can be captured by microphones. Microphones are transducers that convert mechanical energy into electrical energy. They are typically built with two plates, the diaphragm, and the bookplate. In condenser microphones, the vibration induced by sound pressure moves the diaphragm, which results in a change of the overall capacitance. Since the device is polarized by a fixed voltage, the changes in capacitance are translated into electrical current changes, which are later amplified by an external circuit, typically named pre-amplifier and normally implemented in the recording interface. The signals recorded by microphones are typically represented by alternating current around zero and give an estimate of the sound pressure changes in the environment. The signals captured by microphones are sampled by analog-to-digital converters implemented in professional audio equipment. After being digitalized, the signals are typically normalized in the range between -1



(a) Linear **(b)** Decibels

Figure 12: Typical Piano Sound Piano sound of a note C_3 in an upright piano. The sound produced by the piano is normalized in the range -1 / 1. In the music industry, sounds are normally analyzed in terms of recording level using the unit dBFS.

and 1. Figure 12 displays a typical signal of a piano sound.

In the scope of audio equipment, 0 dBFS is defined as the point at which signals start to distort because of clipping. As a consequence, in digital audio recording devices, signals are normally measured in negative values of dBFS and give an estimate of how far a signal is from distortion. As an example, while recording the signal with amplitudes ranging from -1 to 1, a signal amplitude of -0.7 has a recording level of 0.7 or -3 dBFS. This value means that the absolute values of a signal amplitude peaks at 70% of the maximum allowed amplitude before distortion by clipping. Amplitudes that are higher than the maximum amplitude clip the signal and are measured in positive values of dBFS. The unit dBFS is defined in the standards AES SAES17-1998,IEC 61606band ITU-T P.38x. It is important to notice that the recording level does not provide direct information about the absolute measurements of sound pressure in the environment, which would be done in Pascals. Equation (8) shows the equation for converting the level from the linear scale to dBFS.

$$l_{dBFS} = 20\ log\ \frac{l}{\mathrm{l_{max}}} \tag{8}$$

Where:

l: Level in linear scale, i.e. absolute value of the signal amplitude.

l_{dBFS}: Level, in dBFS.

$\mathrm{l_{max}}$: Maximum level, in linear scale, before clipping.

In the scope of audio recording equipment, peak measurements are not good descriptors for noise performance or estimation of loudness, being RMS measurements often more used [27]. In this thesis, loudness is estimated using RMS values and dBFS unit. The RMS value of a set of values is the square root of the arithmetic mean of the squares of the values. In the case of n levels $\{l_1, l_2, \ldots, l_n\}$, the calculation of the RMS level $\mathrm{l_{rms}}$ is displayed in Equation (9).

$$l_{\mathrm{rms}} = \sqrt{\frac{1}{n}\left(l_1^2 + l_2^2 + \cdots + l_n^2\right)} \tag{9}$$

3.6 Dynamics

The dynamics are a musical term used to describe the perception of loudness. An actuated system for piano sampling should be capable of reproducing the different dynamic levels. Table 2 displays different types of dynamics and also their typical interpretation by musicians.

Dynamics are related to changes in loudness and are perceived by an individual as a relative measurement rather than absolute. As an example, the dynamic level p alone does not represent a precise level of loudness, but it rather indicates that it is perceived as less loud than a mp and louder than a pp. Although the dynamics are highly subjective to interpretation of the player, the dynamic symbols are widely used in the technical literature for describing different loudness.

One important characteristic of the piano is that different dynamics produce changes in the power of the harmonic components. In the spectrum, these changes in power are represented by variations in the magnitudes of the spectral components representing the harmonics. These changes in the spectral components of the harmonics are often described with the terms timbre or tone color. To illustrate one's point, Figure 13 shows the changes in relative power between the spectral components of the harmonics for a note A_4.

Since dynamics have a major role in the piano sampling system, a proper mathematical interpretation has to be defined. In the scope of this work, different dynamics are described in terms of RMS recording levels using dBFS units. Chapter 4 provides the methods and equations used to describe the different types of dynamics.

Word (in Italian)	Symbol	Pianist Interpretation
Forte Fortissimo	fff	As loud as possible
Fortissimo	ff	Very loud
Forte	f	Loud
Mezzo forte	mf	Medium loud
Mezzo piano	mp	Medium quiet
Piano	p	Quiet
Pianissimo	pp	Very quiet
Piano Pianissimo	ppp	As quiet as possible

Table 2: Dynamics. In music, the dynamics represent variation in loudness. The dynamics are interpreted in a relative manner rather than absolute. For instance, a p in one piece may sound louder than a p in a another piece, although a mp shall sound louder than a p when played after each other. The dynamics go beyond loudness and include changes in timbre, i.e. changes in the power of the harmonic components.

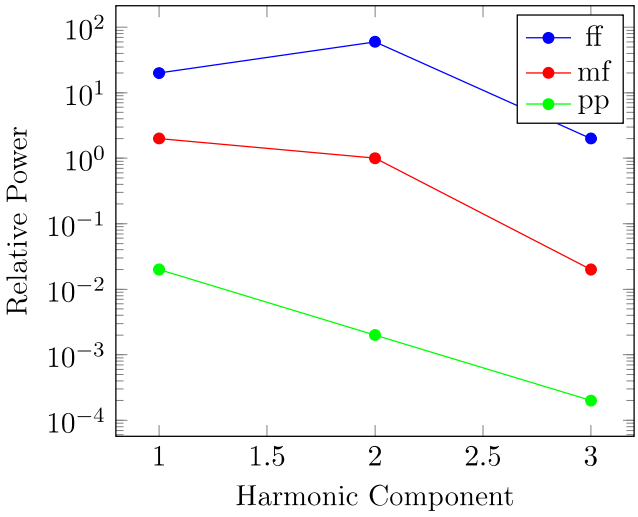


Figure 13: Relative power between harmonics for different dynamics [1]. This plot displays the changes in relative power between the spectral components of the harmonics for a note A_4 in an Upright Piano. It is noticeable that the relative power between the harmonic components change for different dynamic levels. As an example, for a pp (pianissimo) the first harmonic component is approximately 10 times less powerful than the fundamental component.

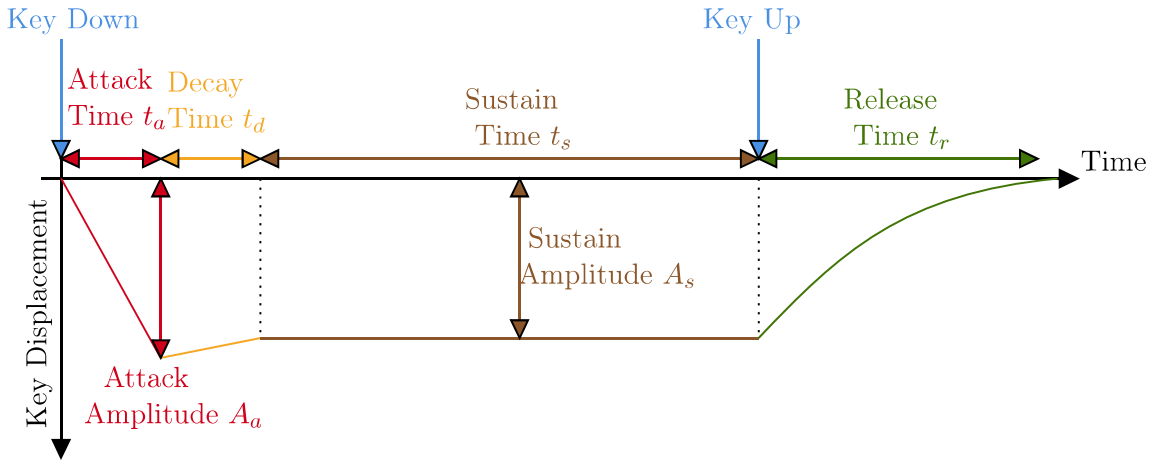


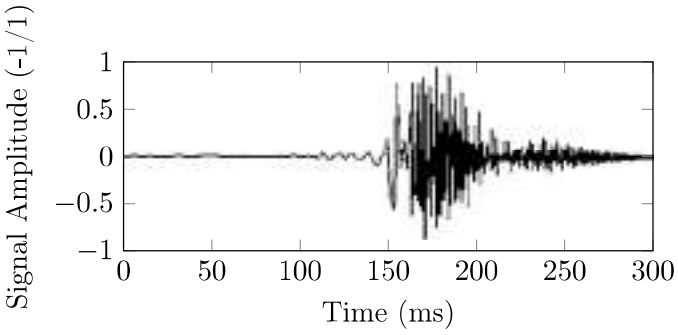
Figure 14: ADSR curve The ADSR curve represents the vertical displacement of the key over time. It is composed by four phases: Attack (A), Decay (D), Sustain (S) and Release (R). Each phase has a time duration and a maximum displacement amplitude. The curve parameters are defined as attack time t_a, decay time t_d, sustain time t_s, release time t_r, attack amplitude A_a and sustain amplitude A_s.

3.7 ADSR Curves

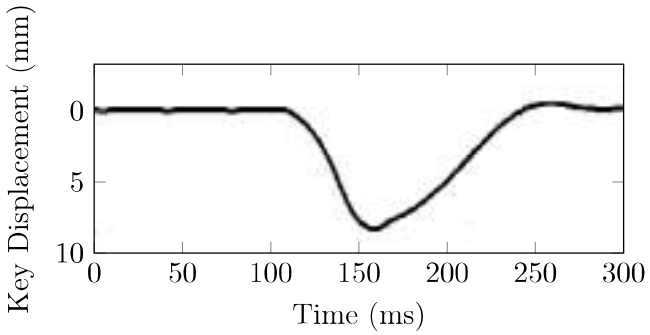
In sound literature, ADSR is a simplified model used for describing sound changes over time and is often named with the term sound envelope. This type of envelope has multiple uses but is normally used to represent the extreme amplitudes of an oscillating signal over time. As an example, when a piano key is pressed and kept down, a near-instantaneous sound arises and decreases in loudness to zero. This loudness behavior can be expressed using ADSR envelopes.

In this work, however, this term is used differently. ADSR is used to describe the movement of the piano key over time. As it is explained in Chapter 4, the piano key is displaced by the linear actuator slider which is positioned on the top of the key. The ADSR curves represent the position of the actuator slider over time and, as a consequence, of the key. Figure 14 shows a typical ADSR curve and its most important parameters.

The ADSR curves are composed of four phases: attack, decay, sustain and release. The attack phase comprises the first movement of the key and ranges from the initial rest position to its maximum displacement. The decay phase is a transition region connecting the attack phase to the sustain phase. In the decay phase, the position of the key moves back until it reaches the sustain amplitude. During the sustain phase,



(a) Recorded Signal



(b) Key Displacement

Figure 15: Signal amplitude and vertical key displacement [18]. Signal amplitude and vertical key displacement as described by Kinoshita et al. [18]. Figure 15a displays the captured signal amplitude while Figure 15b shows the position of the key over time. The initial rest position of the actuator slider is approximately 0 mm.

no movement is observed and the key remains stopped at a fixed position. Finally, the release phase brings the key back to its initial rest position. The duration and amplitude of each phase have an impact on the dynamics of the produced sound.

The modeling of the ADSR curves follows Kinoshita's results. Kinoshita investigated the key displacement over time as well as key-finger forces over time for different dynamics in an upright piano. This study revealed that simple exponential functions can be used to describe the key displacement [18]. Figure 15 shows the signals recorded by Kinoshita.

Kinoshita's experiments have a sustain phase with a duration close to zero, i.e. the key returns to the rest position immediately after reaching the maximum displacement. However, as it is explained later on, the creation of sample libraries requires the recording of full-length notes. A full-length note is achieved by playing the key and

waiting for the loudness to fall below a threshold before entering the release phase, i.e. when the key moves back to its original rest position. To illustrate, a note C_1 can sound as long as 60 seconds before falling under a threshold of -60 dBFS. Only after falling under -60 dBFS the key returns to the original rest position. This condition explains why the sustain region is included in the modeling of the ADSR curve.

3.8 Escapement Point

A detailed explanation of how to correctly press a piano key as provided by Gát [5]. Gát meticulously described the ideal key-press from a pianist's perspective. When a piano key is pressed, two things happen at the same time. The damper moves away from the string while the action mechanism triggered by the key moves the hammer towards the string. At a certain point, the hammer loses mechanical contact with the action mechanism at a point named escapement point. Parlitz suggested that trained pianist accelerate the key up to its maximum speed at the escapement point [28]. After the hammer escapes from the action mechanism, the speed of the key decreases from its maximum value. Parlitz also suggested that amateur pianists tend to spend more finger energy than professional pianists [28]. This situation takes place because amateur pianists move the key up to its maximum speed before the escapement point of the action mechanism. Figure 57 illustrates the speed of the key for professional and amateur players.

3.9 Pedals

Typically, pianos have three types of pedals: the soft pedal, the center pedal, and the damper pedal. In grand pianos the soft pedal, also named *una corda*, shifts all the action mechanisms, including the hammer, slightly to the right. In this situation, the hammers, which normally would strike three strings, will strike only one of them. In upright pianos the soft pedal behaves differently, i.e. instead of shifting right, the hammers get closer to the strings to limit the loudness.

Depending on the type of piano, the center pedal has different uses. In grand pianos, for instance, the center pedal is used as a selective sustain pedal, also known as *sostenuto* pedal. In this case, if a key is played while the center pedal is already pressed, this note will continue to sound even after lifting the finger. On the other hand, notes played after the pedal is lifted will only sound for as long as they are kept down. In certain types of upright pianos, this pedal brings a felt cloth between

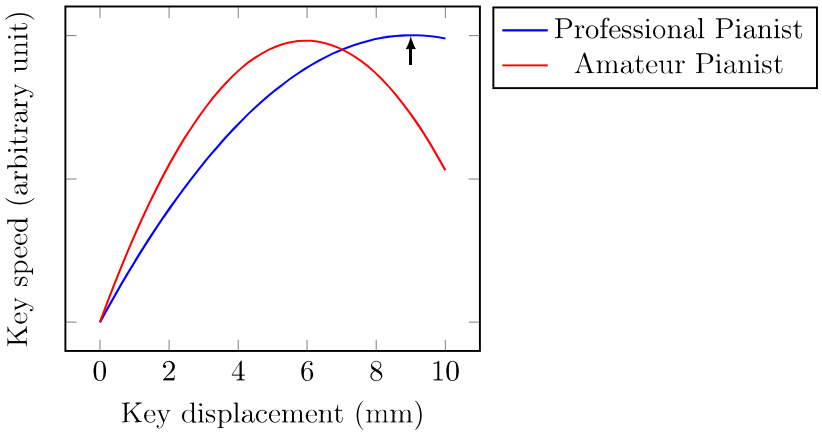


Figure 16: Escapement Point. The escapement point is the moment when the hammer loses mechanical contact with the piano action mechanism. As suggested by Gát, amateur pianists are prone to reach the maximum key speed before the escapement point of the key [5]. Professional pianists tend to reach the maximum key speed exactly at the escapement point [5]. The black arrow shows the approximate position of the escapement point at 9 mm.

the hammers and the strings to muffle the sound.

The damper pedal, as the name suggests, controls the dampers. This pedal is also often named as sustain pedal. While kept down, this pedal elevates all dampers away from the strings, thus allowing them to resonate freely. When a certain key is pressed, all other strings are free to resonate. When this pedal is lifted, all dampers return to the initial position, i.e. damping the strings. The sustain pedal is used to achieve the sustain resonances, which are explained in detail in Section 3.10. The typical arrangement of these pedals is displayed in Figure 17.

Figure 17: Piano pedals. In grand pianos, the soft pedal forces the hammer to strike only one string, the center pedal allows a sound to continue playing even after the key is lifted and the damper pedal moves all dampers away from the strings, thus allowing them to resonate freely.

3.10 Sustain and Sympathetic Resonances

Sustain resonances are obtained when the sustain pedal is held down and a key is pressed. With the sustain pedal down, all the dampers are moved away from the piano strings while the hammer of the current pressed key strikes its strings. In this situation, the harmonic components of the vibrating string excite other strings containing common harmonic components. For example, if the key A_6 (fundamental frequency 1760 Hz) is pressed while the sustain pedal is held down, then the sounds of the keys A_4 at 440 Hz and A_5 at 880 Hz will also be perceived since they contain a harmonic component at 1760 Hz. Thus, every time a key is pressed while the sustain pedal is down, the spectrum of the produced sound contains harmonic components of the string struck by the hammer as well as multiple other resonating strings.

Sympathetic resonances occur when one key is pressed while another key with common harmonic components is already kept down. It is worth mentioning that the key kept down does not produce any sound, only its damper is kept away from the string. In this scenario, the string of the already pressed key will resonate with the sound of the pressed key due to the excitation done by common harmonic components. Again, the spectrum of the produced sound contains harmonic components of the string struck by the hammer as well as from the resonating string. For example, if the key A_6 with fundamental frequency 1760 Hz is pressed while the key A_5 with fundamental frequency at 880 Hz is kept down, then the key A_5 will resonate due to common harmonic components.

A good piano sample library should contain samples of sustain and sympathetic resonances since they account for the authenticity and genuineness of the piano sounds. However, the sustain and sympathetic resonances cannot be recorded directly from the piano and require the use of audio processing techniques. A common strategy used for obtaining these resonances consists of phase matching equal sounds recorded with and without resonances, for instance with the pedal up and then down. After that, the individual sounds are subtracted from each other and the corresponding resonance is obtained.

3.11 Release Sounds

When a key is pressed and the hammer collides with the string, the produced sound attenuates over time with a certain decay rate as long as the key is kept down. When the key is released, i.e. lifted, the damper returns to its original rest position, i.e. touching the string and ceasing the sound. Slowly or rapidly releasing the key

produces different sounds. Often terms like soft release, medium release, and hard release are employed. Moreover, since the sound is still perceived in the sustain region, lifting the key at different moments also leads to different sounds.

An authentic piano sample library should contain all these release sounds. For that, it is necessary to record all different types of sounds individually while creating the piano sample library. A precise key triggering system should be able to trigger different types of releases sounds (such as soft release, medium release, and hard release) at different moments in the sustained region of a sound. Chapter 4 details a procedure to obtain the release sounds using the system described in this thesis.

3.12 Timing

A typical recording session can involve dozens of microphones and last as long as days. Audio engineers tend to use as many microphones as possible in order to decide on a later moment what type of mix will be used for the recordings. To create individual samples, this large amount of recording material needs to be cropped, i.e. the start and end markers of each sample need to be known. To set the markers for trimming the samples correctly, the most important temporal events involved in the piano action mechanism need to be properly understood. To illustrate one of the many temporal events, when the piano key is played, there is a time delay between the beginning of the key movement and the beginning of the sound onset. This delay needs to be considered by the key triggering system that estimates the start point of a sample.

Askenfelt estimated typical contact duration of action components for different keys [13]. By using an experiment setup with several switches, it was estimated that for a typical staccato note at dynamic level f, the finger-key contact lasts approximately 80 ms while the key-keybed contact lasts about 35 ms. The contact between the hammer and string is much shorter, in this experiment approximately 2 ms. The complete sound produced lasts for about 140 ms [13]. Figure 18 displays these temporal events.

The delay between the beginning of the hammer-string contact and key-keybed contact varies depending on the dynamic level that the key is pressed. The start of hammer-string contact can be interpreted as the exact moment of the sound onset while the start of key-keybed contact can be understood as the point where the key displacement reaches its maximum value, i.e. maximum attack amplitude A_a in the ADSR curve. For high dynamic levels, i.e. short attack times t_a, the key-keybed contact starts before the hammer-string contact. The opposite is also valid, low

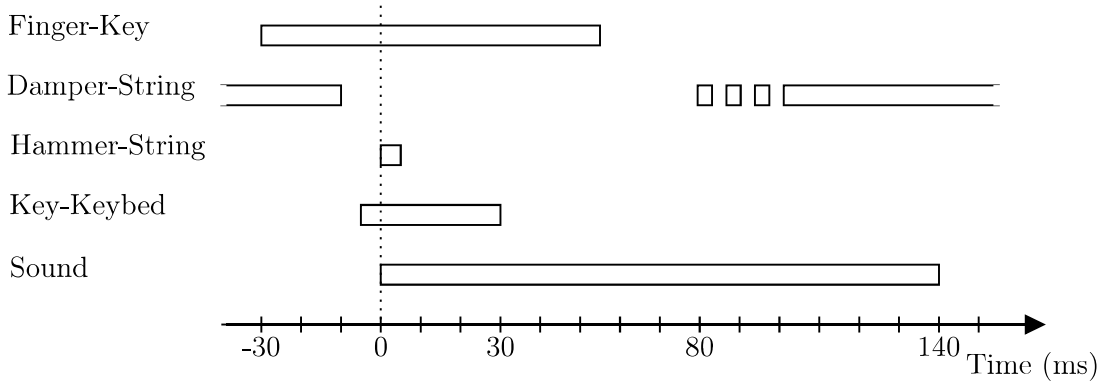


Figure 18: Contact times Askenfelt described the contact durations involved in the piano action mechanism [13]. In this figure, it is displayed the typical durations for a note C_4 played at dynamic level p [13].

dynamic levels, i.e. long attack times t_a, produce hammer-string contact before key-keybed contact. These delays can reach up to 20 ms and should be considered in the precise key triggering system. Figure 19 displays the delays between the hammer-string contact and key-keybed contact for different dynamic levels.

As described by Askenfelt, the duration of the hammer-string collision is short, typically in the range of 1 to 4 ms [13]. The duration of the hammer-string collision is mainly determined by the pressed key and the dynamic level. Higher dynamic levels have a shorter duration of the hammer-string collision while lower dynamic levels have a longer duration. Figure 20 shows the relation between the duration of the hammer-string collision for different dynamic levels as described by Fletcher [24].

As explained in Section 2, Goebl focused on analyzing the temporal behavior of piano action components in different types of pianos. The movement of keys and hammers were monitored with accelerometers over the entire dynamic range [14]. The accelerometer's data allows the temporal estimation of the finger–key contact, hammer–string contact, and key-keybed bottom contact as well as the maximum hammer speed. One of the conclusions of this study suggests that the delay between the finger-key contact and the key-keybed bottom contact changes significantly for different dynamic levels, although only slightly between different pianos. Figure 21 shows the results of Goebl's experiments.

Dijksterhuis investigate the relationship between the hammer speed immediately before the hammer-string collision and the depressing time of the key, i.e. the time required for the key to reaching the keybed bottom. In this experiment, Dijksterhuis considered the finger-key force to be constant over time. The results of the experiment

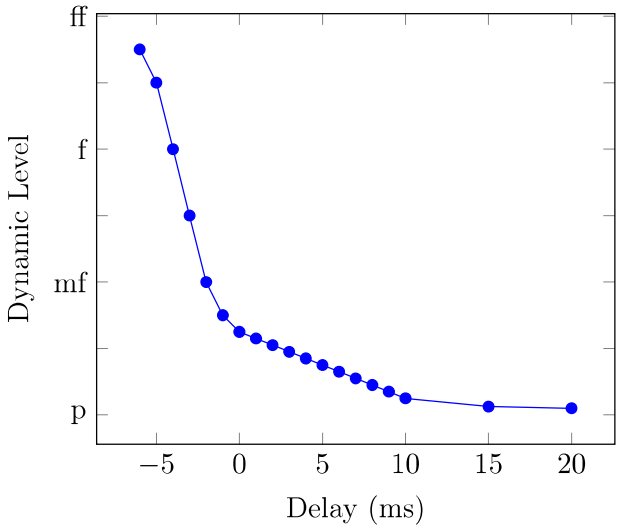


Figure 19: Delay between hammer-string collision and key-keybed contact [29]. The hammer-string collision takes place at 0 ms. Positive delay values imply in key-keybed contact taking place after hammer-string collision. Negative delay values imply in key-keybed contact taking place before the hammer-string collision.

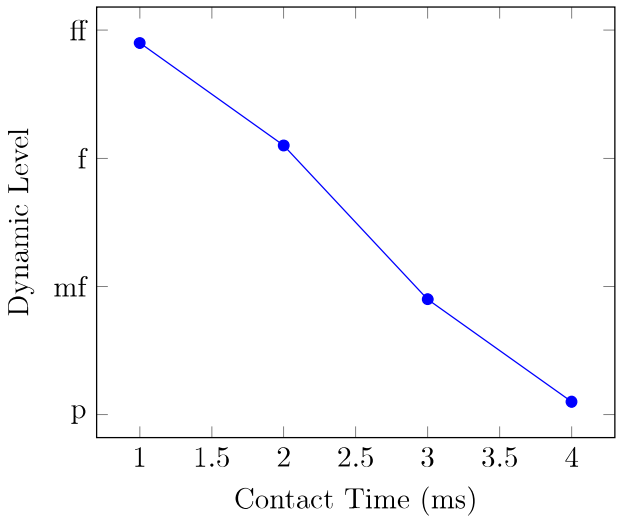
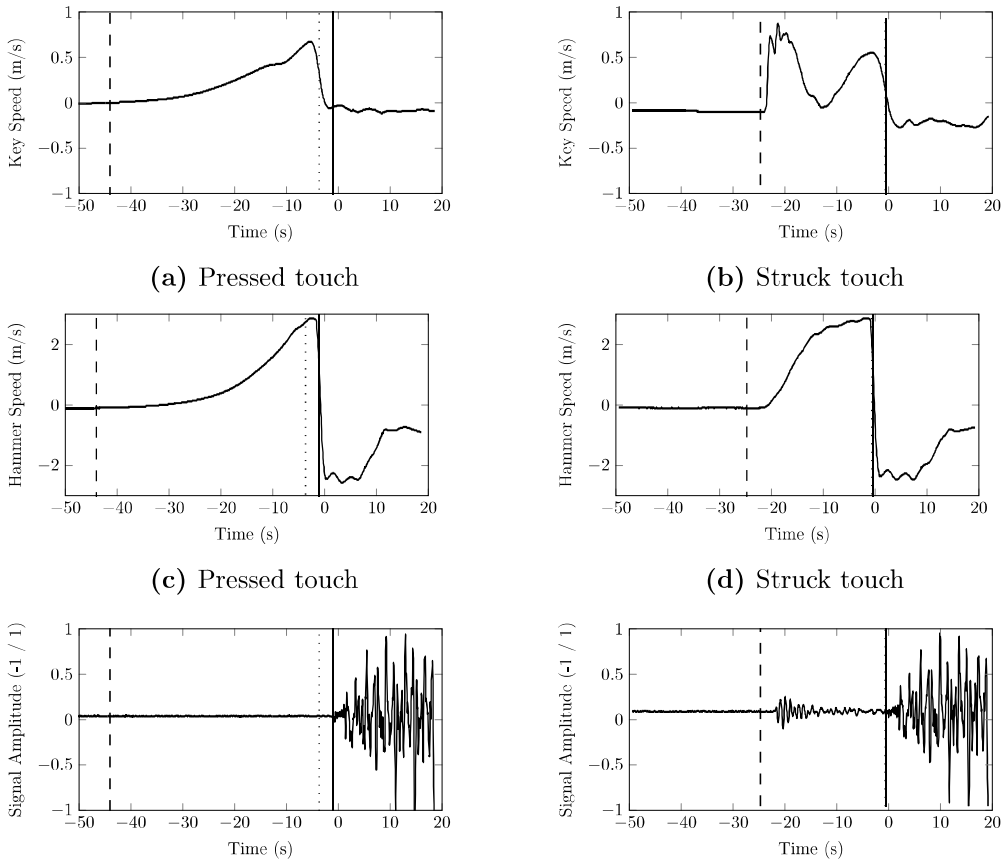


Figure 20: Duration of the hammer-string collision [24]. As described by Fletcher, the duration of the hammer-string collision can be as low as 1 ms for a dynamic level ff and as high as 4 ms for a dynamic level p [24].



(a) Pressed touch **(b)** Struck touch

(c) Pressed touch **(d)** Struck touch

(e) Pressed touch **(f)** Struck touch

Figure 21: Struck and pressed touch [14]. Results of an experiment to analyze the temporal relations in struck and pressed touch performed by Goebl et al. [14]. The long dashed line represents the finger-key contact point. The dot-dashed line represents the moment when the key reaches the keybed-bottom and the solid line represents the moment of the hammer-string collision. One general conclusion of Goebl's study is that the interval between the finger-key contact point and the hammer-string collision changes with different touches. These plots are an adaptation from Goebl's results and might contain minor differences.

are displayed in the Figure 22. Low dynamic levels are related to lower forces whereas high dynamic levels are related to higher forces.

Another important characteristic of pianos is the decay rate of the loudness over time. When a key is pressed and is kept down, the loudness of the sound decays with different decay rates for different keys. Lower keys have lower decay rates and longer sounds. Higher keys have higher decay rates and shorter sounds. Cheng investigated

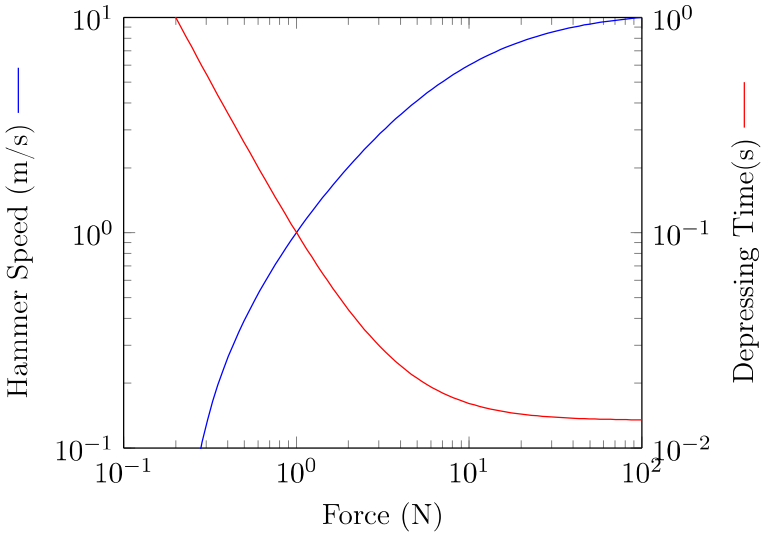


Figure 22: Relation between hammer speed, depressing time and force [24] [30]. An experiment performed by Dijksterhuis showed that the hammer speed before collision and the depressing time have an inverse relation. In this experiment the force that the key is pressed is assumed to be constant over time. Dijksterhuis also performed experiments considering other force models [24]

the decay rate for different keys at different dynamic levels. One of the conclusions is that the dynamic level at which a key is pressed has no significant effect on the decay rate [31]. However, the decay rate changes significantly between the keys. In the scope of the key triggering system, it is important to understand the behavior of the decay rate mainly to determine the end of the sample. Figure 23 shows the behavior of the decay rate for keys.

3.13 Haptics

Musical haptics concerns the sense of touch while playing musical instruments. Tactile aspects have considerable importance in music performance and perception. The touch feedback provided by the instrument plays an important role while playing a musical instrument. Papetti described the interaction between player and instrument as multi-input, multi-output systems [32]. This is displayed in Figure 24.

In the scope of this thesis, understanding the haptics behind the instrument is an essential part of the design of a system used to excite the piano. Figure 24 shows that the player receives two inputs, an acoustic signal that is captured by the ear

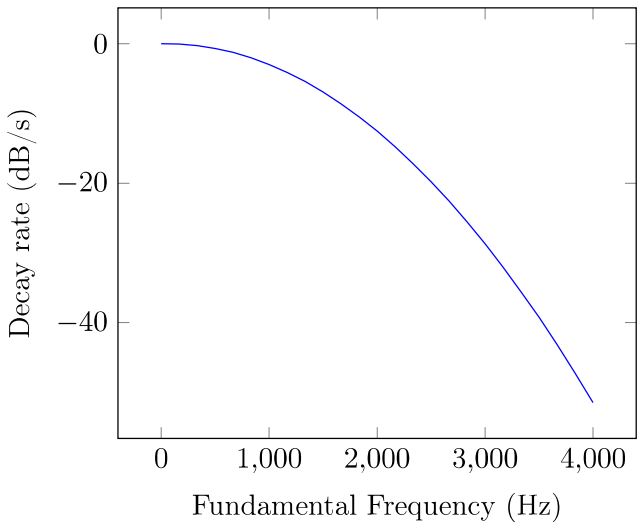


Figure 23: Decay rate. Decay rate over different note's fundamental frequencies [31]. When pressed, low keys have smaller decay rates and last longer. High keys have larger decay rates and last shorter.

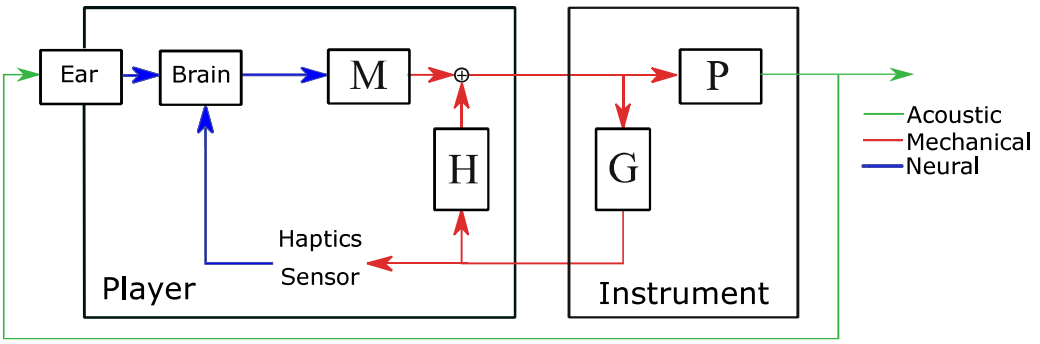


Figure 24: Haptics [32] As suggested by Papetti, player and instrument can be modeled via multi-input, multi-output systems [32]. M, H, P, and G are directly involved in the interface between player and instrument. M represents the conversion from the neural signals to the excitation of the instrument. The excitation of the instrument is transformed by P into acoustic energy, which is then captured by the ear. G converts the excitation of the instrument in the mechanical response of the instrument. The tactile concerns mostly H, which converts the mechanical response of the instrument into excitation of the instrument [32].

and a mechanical signal which is composed of the tactile feedback provided by the instrument. This information is used for the general concept of the key triggering system, which relies on a microphone and a linear actuator.

3.14 Noises

As described by Gát [5], when a piano key is pressed, three major types of noises can be perceived in the produced sounds. These noises are a result of the hammer-string collision, the key-keybed bottom collision, and the finger-key collision. When analyzing the spectrum of these different types of noises, one can notice low and high-frequency components. The low frequencies are mostly produced by the key-keybed contact [5]. The high frequencies are mostly a result of the finger-key contact [5]. These noises contribute for the authenticity of piano sample libraries and should ideally be recorded separately. However, when a key is pressed, the string vibrates and as a consequence, these noises cannot be recorded separated. In fact, the noises are much quieter than the sounds produced by the strings. To record the noises without the sound of the string, a common method consists of damping the strings with a soft material or removing the hammer from the action mechanism. As concluded by Askenfelt, the damped strings should keep the spectral characteristics of the original noise, only that in the damped situation the sound is attenuated [33]. Damping the string with soft material is not a completely effective method since the string sounds can still be perceived. Removing the hammer from the action mechanism solves this problem but has the major disadvantage of changing the mechanics of the action mechanism.

If recorded individually, the noises can be reproduced by digital implementations of pianos. In a piano sample library, the recording of these noises accounts for the authenticity of the obtained sounds as well as the realistic perception of the virtual instrument. Figure 25 shows the typical spectral profile of the different noises.

3.15 Sample Libraries

A sample library is a collection of digital audio recordings, commonly named samples. Sample libraries can be classified into loop libraries and instrument libraries. Instrument libraries use recordings of single notes performed by a real player. To reproduce the full artistic spectrum of the instrument (and the player), different velocities and articulations must be recorded. A complete piano sample library would

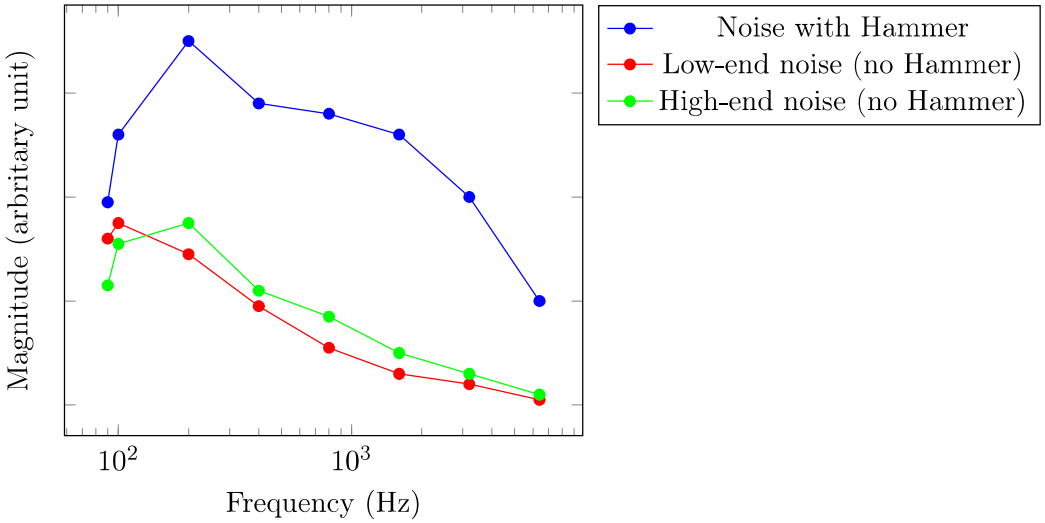


Figure 25: Noise of a note G in a 274 Steinway-Flügel [5]. This spectrum displays two types of noises, a low-end noise, mostly produced by the key-keybed bottom collision, and the high-end noise, mostly produced by the finger-key collision. [5]. The noise with hammer is a combination of the low-end noise, high-end noise and the sound produced by the string.

contain notes with the sustain pedal up, notes with the sustain pedal down, notes with sympathetic resonance, different release sounds (soft, medium, hard), and keybed noises. Moreover, these different types of sounds are recorded at different dynamic levels. Finally, the single-note samples are arranged inside in the sampler.

MIDI represents a communication protocol, digital interface, and connectors that link computers, electronic musical instruments and various types of audio devices. A MIDI connection is normally performed via MIDI cable and can transport information in sixteen different channels. Each of the channels can be linked to different devices or instruments. MIDI transports event messages, i.e. data that specifies information for music, such as note, pitch, loudness (or velocity), vibrato, stereo panning and clock signals used for tempo synchronization. A MIDI controller is a specific hardware component that converts key-presses, button presses, knob turns and slider changes into MIDI data, sending it via MIDI protocol. To illustrate, MIDI control change CC#64 is used to select the sustain pedal up and sustain pedal down. In music, a common application consists of playing a MIDI controller hardware mapped to a software sampler containing the sounds of a musical instrument, such as a piano.

To illustrate the relation between sample libraries and MIDI, samples are normally

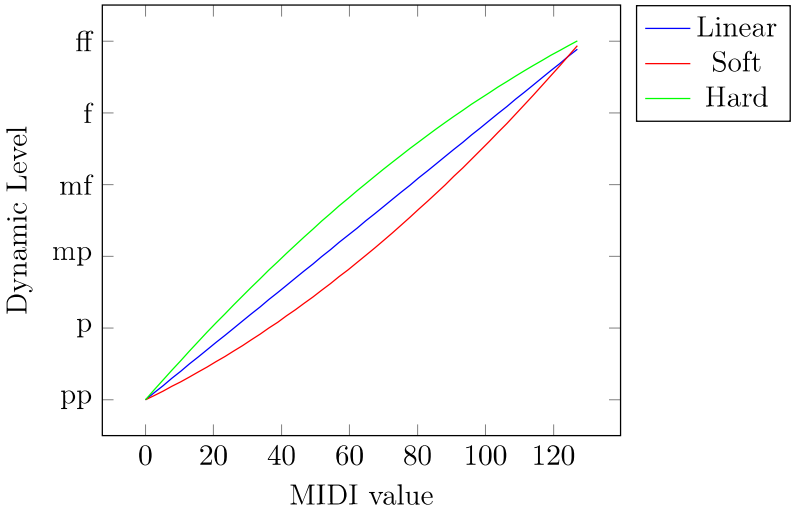


Figure 26: Velocity Curve. MIDI controllers and software sampler implement a digital velocity curve, where an input MIDI value ranging from 1 to 127 is converted into a dynamic level. Knowing the dynamic level, the corresponding sample is finally played.

loaded into a software sampler and triggered via MIDI keyboard. Software samplers are typically available as Virtual Studio Technology (VST) plug-ins and can be loaded in Digital Audio Workstation (DAW)s. Frequently, when a key is pressed in the MIDI keyboard, an internal micro-controller calculates the time delay between the closing of two on-off switches located along the horizontal axis of the key lever. This time delay gives an estimation of how fast or how slow the key was pressed. This delay is finally converted into a MIDI data between 1 and 127 and appended to a note-on message in the MIDI protocol. This MIDI note-on message is then transmitted to a software sample, where it is finally converted into a dynamic level via the velocity curve implemented inside the sampler. Figure 26 displays examples of typical velocity curves implemented by software samplers.

In the software sampler, when a single key is pressed, the MIDI value is converted into a dynamic level and the corresponding sample is reproduced. Moreover, combinations of key-presses are normally implemented via input-output matrices, where the input is set of keys pressed and the output is the sound resulting from reproducing multiple samples. As an example, consider that the key C_1 is pressed at ppp and simultaneously the key C_2 is pressed at fff. In this situation, the piano sampler would simultaneously reproduce the samples C_1 at ppp and C_2 at fff originally recorded in a piano. Table 3 displays the basic structure of a sample library.

		Key				
Dynamics	**MIDI**	A_0	B_0	**...**	B_7	C_8
ffff	115-127	Sample 1	Sample 5	...	Sample 9	Sample 13
fff	101-114	Sample 2	Sample 6	...	Sample 10	Sample 14
...	...	...	...	...	...	...
ppp	11-23	Sample 3	Sample 7	...	Sample 11	Sample 15
pppp	0-10	Sample 4	Sample 8	...	Sample 12	Sample 16

Table 3: Sample Library. Basic organization of a sample library in a piano sampler. When a key is pressed, a MIDI value ranging from 0 to 127 is converted to a dynamic level and the corresponding sample is played. For example, if the key A_0 is pressed at fff and the key B_7 is pressed at ppp, then the samples 2 and 11 are played by the software sampler.

To illustrate the dimension, an arbitrary sample library of an 88-key piano could create 11440 samples if one considers 130 samples per key. These 130 samples would account for the different dynamic levels sustain resonances, sympathetic resonances, release sounds at different sustain times, key-keybed noises, among others. These numbers can considerably change depending on numerous factors, such as the number of dynamic levels recorded, the number of microphones used in the recording session, the number of piano keys and so on. Figure **??** shows a piano instrument loaded in the software sampler Steinberg HALion.

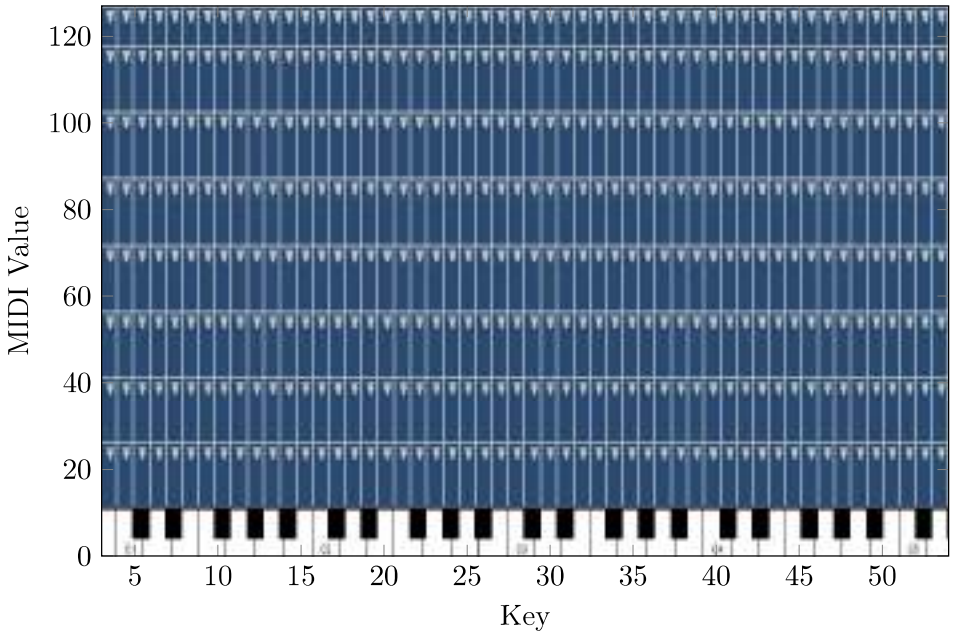


Figure 27: Software Sampler Steinberg Halion. Each rectangle represents an individual sample recorded in a real piano. When the key is pressed in a MIDI controller, the dynamic level is interpreted and the corresponding sample is played by the software sampler.

4 Approach

This chapter presents the main components of a state-of-the-art system for precisely triggering the piano keys. It all starts with the system requirements, from which the main component specifications are derived. Next, the system architecture is presented, as well as the description of the main components. The system is composed of hardware and software components, which together allow the precise triggering of the piano keys. The main hardware components include the linear actuator, a PID controller, a computer, an audio interface, a microphone as well as a mounting system. The software component is mostly composed of a central control software implementing different types of algorithms based on sound analysis.

4.1 Requirements

To achieve realistic piano sample libraries, a set of main requirements for an ideal precise key triggering system for piano sample recordings was created. These requirements theoretically allow the creation of realistic samples for piano sample libraries. In that context, the precise key triggering system should ideally:

- **Mimic the finger movement of a trained pianist:** As detailed in Section 3, different types of pianists press the key in different ways. In this thesis the piano triggering system focus on trained pianists.
- **Be non-invasive:** No sensors should be installed inside the piano mechanics. This requirement aims to avoid modifications that could lead to changes in the authenticity of the sounds.
- **Trigger notes without sustain resonances at different dynamic levels**. In simplified terms, the system should trigger the piano keys at different dynamic levels and with the sustain pedal kept up.
- **Trigger notes with sustain resonances at different dynamic levels**. The system should trigger the piano keys at different dynamic levels and with the sustain pedal kept down.

- **Trigger notes with sympathetic resonances at different dynamic levels**. The system should trigger the piano keys at different dynamic levels considering the inter-key resonances, i.e. when passive strings respond to external vibrations caused by other strings with harmonic likeness.

- **Trigger notes with different release sounds at different dynamic levels and different note lengths**. One of the most complex features of all, the system should trigger the keys at different dynamic levels and release them at different moments and with different types of releases, i.e. soft to hard.

- **Estimate the start and end of samples:** The start and end markers are used for sample cropping. The start of the sample should consider the delays involved in the piano action mechanism while the end of the sample should simply indicate when the sound completely fades away.

- **Be silent** This requirement is necessary due to the commitment with the authenticity of sounds. Post recording methods, such as noise filtering, are not ideal and should be avoided as much as possible.

4.2 System architecture

The precise key triggering system is composed of a linear actuator, a controller, a control software, an audio interface, and a microphone. The microphone is a transducer that converts changes in sound pressure into electrical current. This current is sampled by the analog-to-digital converter of the audio interface at a frequency of 44.1 kHz. The sampled signal is finally analyzed by a control software running on a computer. Depending on the specific task being executed, the controller uploads different types of motion curves (ADSR set-point curves) into a PID controller. These curves consist of sequential position set-points that must be followed by the linear actuator. The linear actuator finally moves the actuator slider to the position set-points over time, thus achieving the desired actuator slider's movement. Figure 28 displays the system architecture.

4.3 Coordinate System

In the scope of this thesis, the reference coordinate system is positioned vertically in a way that it matches the direction of the key displacement as well as the actuator slider axis. The rest position of the slider is adjusted so that it touches the surface of

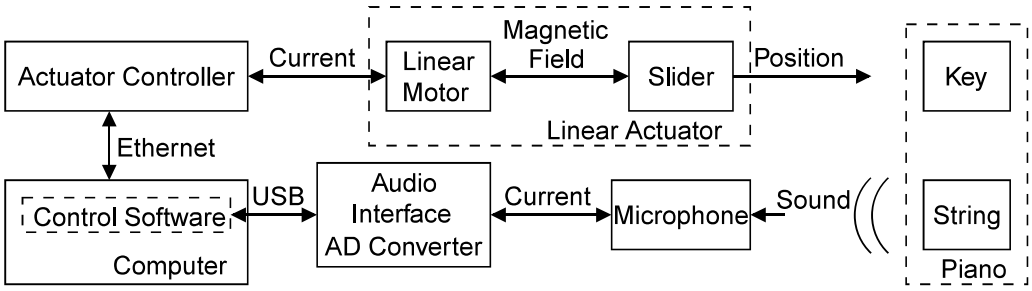


Figure 28: System architecture. The sounds produced by the piano are captured by a microphone, sampled by an audio interface and finally analyzed by a control software. The control software controls a PID controller, which in turn moves the actuator slider in the linear actuator. Finally, the actuator slider moves the piano key and the desired sound is achieved.

the piano key without pressing it down. This point of immediate contact is considered to be the zero mark of the selected coordinate system. Any vertical displacements of the key are measured using this reference. As an example, a position of 5 mm means that the actuator slider moved down by 5 mm and as a consequence, the piano key also moved down by 5 mm. In this thesis, typical values for the actuator slider position are within the range 0 to 10 mm, although in some experiments values below 0 mm and over 10 mm might appear. Figure 29 shows the coordinate system used in this work.

4.4 Linear Actuator

To reach a precision movement similar to the finger of a trained pianist, a precise linear actuator is used. A linear actuator allows the motion to be executed along a certain axis and is based on a linear induction motor, having a working principle similar to standard induction motors.

Typical linear systems are composed of a control drive and a linear actuator. The linear actuator is composed of two main components, a tubular stator, and a slider bar. The tubular stator is constructed with motor windings, bearings to support the slider motion, position sensors and a micro-processed circuit for controlling the current. The slider bar is constructed with neodymium magnets along the axis of a stainless steel bar. The position of the slider bar can be monitored at all moments, i.e. when the slider bar is stopped and also when it is moving. Whenever an error between the actual position and set-point position is detected, the control system ensures

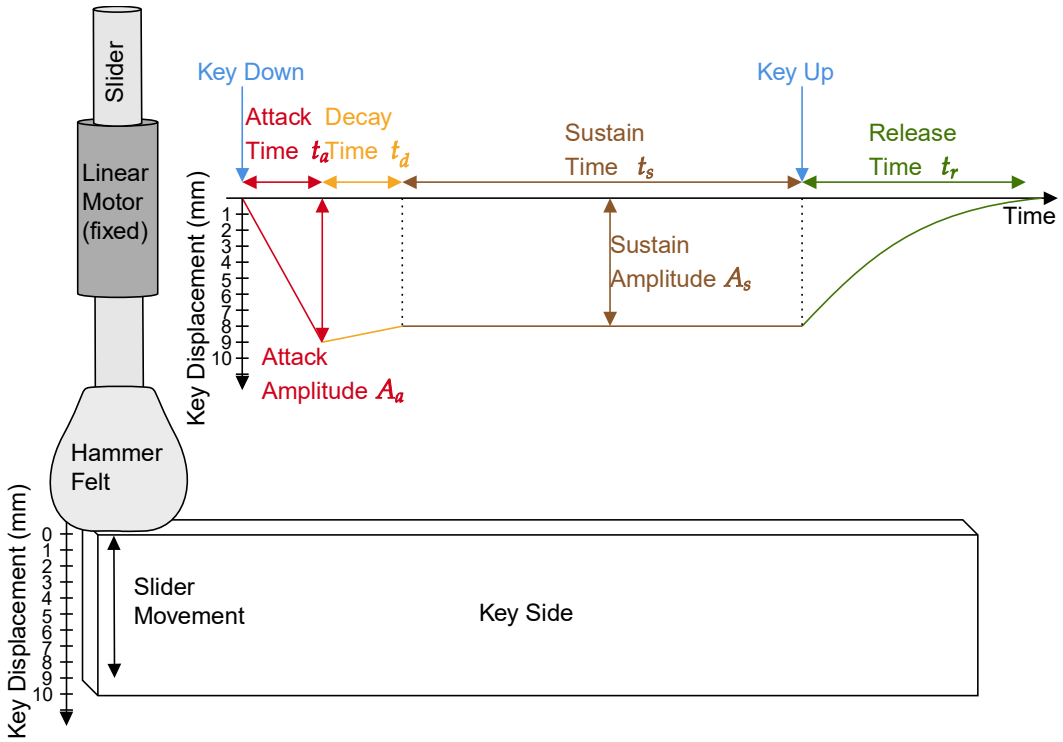


Figure 29: Coordinate System. The ADSR curve represents the position of the actuator slider over time. Before triggering the key, the lower-end of the actuator slider stands still at the position 0 mm in the coordinate system. When the key is triggered, the actuator slider moves down and its position increases. The slider's movement, as well as the vertical displacement of the key, is physically limited by the keybed bottom distance, typically located around 10 mm. After following the positions of the ADSR curve, the actuator slider returns to its original rest position at 0 mm.

(a) Linear Motor (Stator) **(b)** Slider (Rotor)

Figure 30: Linear Actuator. The combination of linear motor and slider bar forms the linear actuator. The linear motor is constructed with unrolled motor windings and behaves as a stator while the slider bar is constructed with magnets and acts as a rotor. The length of the slider bar is 170 mm and allows a maximum stroke distance of 40 mm.

fast correction. The specified linear system allows positioning over the complete stroke range, in this case, rated at 40 mm. Related variables such as maximum speed, maximum acceleration, and maximum current can also be controlled. Figure 30 displays the linear motor and the slider.

The linear actuator is based on a linear induction motor. It has a similar working principle of induction motors, except that the construction is linear. Just like an inductive motor, a linear induction motor is composed of a pair rotor-stator. In a linear induction motor, the stator is unrolled and composed of several coil windings constructed in a tubular form. The rotor is composed of several neodymium magnets oriented in opposite directions along the slider bar axis. As a result, the device produces linear forces along the axis. This differs from a classical rotor, which would produce torque (rotation) around the axis [34]. It is important to notice that these types of actuators do not use intermediary components such as gearboxes or spindles, which lead to less noises and maintenance.

The controller supplies 3-phase current to the coil windings in the stator. These currents are transformed into alternating magnetic fields. Due to the construction characteristics, a resultant force arises along the slider axis, resulting in movement. Figure 31 displays the basic working principle of a linear actuator. Important physical quantities such as current, force and magnetic fields are displayed and can be interpreted according to Lorentz law.

The linear actuator meets the motion and force constraints of a pianist's finger, such

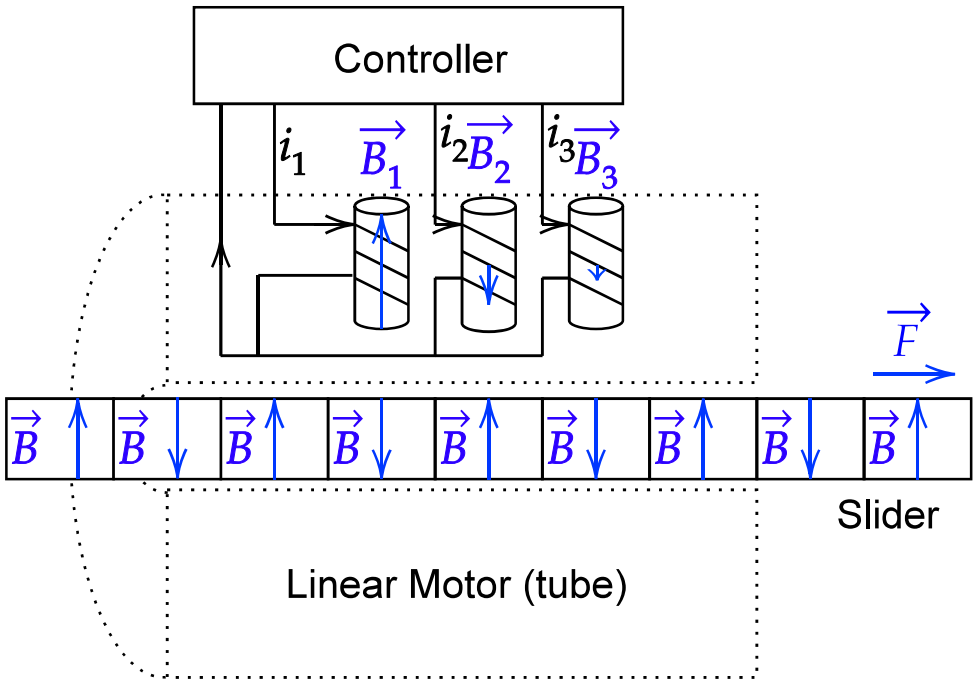


Figure 31: Linear Actuator. The linear actuator is composed of a linear motor and a slider bar. The linear motor is constructed in the format of a tube and contains coil windings, which are energized with 3-phase current supplied by the controller. Along the slider axis, several magnets are oriented in opposite directions. A resultant force arises along the axis of the actuator slider, thus resulting in its movement.

as maximum displacement, maximum speed, maximum acceleration, and maximum force. In order to specify the linear actuator, an assessment of force parameters was performed. Parlitz described typical forces for amateur and expert pianists fingers [28]. In his work, Parlitz concluded that forces differ depending on the finger, note played, and pianist experience. Likewise, Thorin determined the forces while pressing a piano key [11]. Based on these works, it is assumed that the maximum finger force does not exceed the 30 N. This value associated with a safety margin allows specifying the maximum force of the linear actuator at 67,1 N [9].

Another important parameter required while specifying the linear actuator is the maximum stroke distance. The maximum stroke distance represents the maximum allowed displacement of the slider bar. In the context of this thesis, it should be larger than the maximum vertical displacement of a piano key. Most of the pianos have a maximum displacement distance of approximately 10 mm [14]. Based on this value and considering a safety margin, the maximum stroke distance was specified at 40 mm [9]. The position precision of the linear actuator was specified at 0.05 mm [9].

Finally, it is essential to specify the maximum speed of the linear actuator. Bella studied the kinematics of pianist's finger and found average maximum speeds of approximately $0.3\frac{\mathrm{m}}{\mathrm{s}}$ [35]. Thorin also performed experiments with pianist's finger motion and estimated the maximum vertical finger speed not exceeding 0.6 m/s for dynamic level f [11]. Based on these results and considering some safety headroom, the linear actuator was specified with a maximum speed of $7.2\frac{\mathrm{m}}{\mathrm{s}}$. Table 4 lists the main specifications of the selected linear actuator manufactured by LinMot.

4.5 Controller

The controller is an embedded hardware that allows positioning the actuator slider. It operates with a position resolution of 0.1 μm and a time resolution of 1 ms. The latter constraint implies that two sequential position set-points are minimally updated every 1 ms. The controller stores up to 99 different curves containing sequential position set-points, i.e. the ADSR curves representing the set-points to be followed by the controller over time. The position set-points are uploaded to the controller via an array of integer values representing multipliers of the amplitude resolution. To illustrate, a value of 10000 represents a position of 1 mm, i.e. 10000 times 0.1 μm. These position set-point curves are uploaded into the controller's Random-Access Memory (RAM) and executed via the UDP-based protocol LinUDP, explained in detail in Section 4.5.1. Figure 32 shows the controller LinMot C-1250 used in this

Parameter Name	Value
Slider Length	170 mm
Slider Mass	130 g
Maximum Stroke Distance	40 mm
Stator Length	105 mm
Maximum Slider Speed	7,3 $\frac{m}{s}$
Maximum Force	67.1 N
Maximum Continuous Force	13 N
Position Detection Repeatability	±0,05 mm
Position Detection Linearity	± 0.3 %
Supply Voltage	72 VDC
Maximum Current (peak value)	7,4 A
Maximum Continuous Current	1,5 A

Table 4: Nominal specifications of the Linear Actuator [9] The linear motor and the slider bar meet the motion and force constraints of a pianist's finger suggested in the literature. These constraints include the values of maximum stroke distance, maximum speed, maximum force, among others.

system.

The controller's internal memory has a limitation of 16000 position set-points. If a time resolution of 1 ms is considered, this value represents 16 s. However, some piano sounds last longer than 16 s and, as a consequence, the controller's internal memory cannot store all position set-points. Just as a reference, a key C_1 pressed at ff can last over 60 seconds due to its small decay rate. To overcome this technical limitation of the controller, a strategy of dividing the ADSR curve into different set-point curves is used. By doing this, it is possible to reuse parts of the curve that are identical and optimize memory usage. This is implemented by using two separated set-point curves, one containing the attack and the decay phases, and another one containing the release phase. During the sustain phase, the position set-point remains fixed at a certain value and does not need to be updated. In practice, the sustain phase is implemented by waiting timers in the control software. To illustrate the strategy, the execution of a ADSR curve is done by triggering one attack-decay curve, waiting for a certain time interval in the sustain and finally triggering the release curve.

The controller implements a discrete PID controller. The PID controller is a closed-loop control scheme widely used in the industry. It compares the measured position z_m with the desired position set-point. The error or difference, e, is then processed to calculate a new process input, i. This input tries to adjust the measured

Figure 32: Controller. The controller is an embedded hardware responsible for positioning the linear actuator [36]. It implements PID control, i.e. a closed loop control scheme that is constantly minimizing the error between the measured position and the position set-point. The controller follows a sequence of position set-points saved internally.

position value back to the desired set-point. Figure 33 shows the block diagram of a typical PID controller.

Different from simple control algorithms such as open-loop control, the PID controller can handle inputs based on the history of a signal. This makes the PID control a stable and accurate control method [37]. The fundamental idea is that the controller reads the actual position with a sensor, calculates the position error value and then manages it in three different ways. The present is treated through the proportional term K_p, the past is treated using the integral term K_i and the future is managed by the derivative term K_d. To operate correctly, these three parameters have to be properly chosen.

The control scheme displayed in Figure 33 can be represented in s-domain by Equation (10), where T_p, T_d and T_i indicate the time constants of the proportional, derivative and integral terms.

$$\frac{i}{e}(s) = H(s) = K_p \left(1 \; + \; \frac{1}{T_i s} + \; T_d s\right) \tag{10}$$

It is possible to find the relation between $i(t)$ and $e(t)$ in the time domain, as it is displayed in Equation (11).

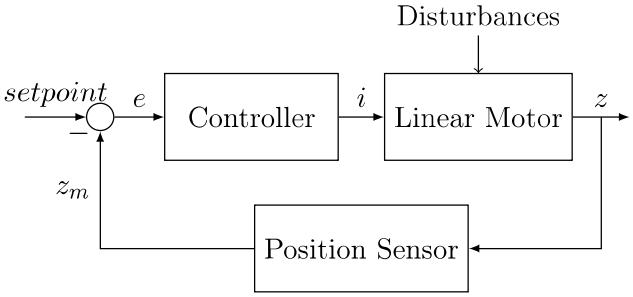


Figure 33: PID Controller Block Diagram of the PID controller. The measured position value z_m is compared with a reference position set-point. The error e is used to find a new process input i. This input tries to adjust the measured position value back to the desired set-point.

$$i(t) \;=\; K_p \left(e(t) \;+\; \frac{1}{T_i} \int_0^t e(\sigma) d\sigma \;+ T_d \frac{de(t)}{dt} \right) \tag{11}$$

Figure 11 is continuous and needs to be discretized. For that, the integral and the derivative terms can be approximated as is shown in Equation (12).

$$\int_0^t e(\sigma) d\sigma \approx \mathrm{T} \sum_{k=0}^{n} e(k) \quad \frac{de(t)}{dt} \approx \frac{e(n) \;-\; e(n-1)}{T} \qquad t = nT \tag{12}$$

Where:

n: Discrete step at time t.

T: Sampling period.

Finally, the discrete version of the PID controller can be modeled using Equation (13) [37].

$$i[n] \;=\; K_p \; e[n] \;+ K_i \sum_{k=0}^{n} e[k] \;+ K_d \; (e[n] \;-\; e[n-1]) \tag{13}$$

Where:

$$K_i = \frac{K_p \mathrm{T}}{T_i} \text{ and } K_d = \frac{K_p T_d}{\mathrm{T}}$$

After being powered on, the controller executes a procedure named homing. During homing, the actuator slider moves towards its end searching for the end position. At the end of the slider, a mechanical pin touches the linear motor, thus aligning the actuator slider with the linear motor. After this moment, the actuator slider moves

along its axis and an internal calibration procedure maps the real coordinates to its internal coordinates. The homing procedure ensures that the coordinate system of the PID controller is properly set.

The controller is powered by two different power units. The first one is a 72 VDC power supply unit and is used to control the linear motor. This is executed by a 3-phase wiring setup connected to the stator windings. The second unit is responsible for the power supply of the internal micro-processed system. This unit is rated at 24 VDC and has a maximum current of 2000 mA.

4.5.1 LinUDP

The controller supports protocols that provide determinism and real-time control such as PROFINET, EtherNet/IP and EtherCAT. However, the use of these protocols is often complex and requires the use of additional hardware, for example, a Programmable Logic Controllers (PLC) to run the logic. LinUDP allows the communication between the linear actuator and a personal computer over Ethernet and without any additional hardware. It is based on User Datagram Protocol (UDP), which means that no checks are performed for data delivery and loss. The delivery of LinUDP packages relies on system timing, which is controlled by the operating system. As a consequence, packages are delivered with jitter, often ranging from 5 ms to 10 ms.

The communication via LinUDP takes place using a Dynamic Link Library (DLL) library provided by the manufacturer LinMot. This library provides domain-specific functions used for communication with the controller. They allow updating the controller's set-point, upload set-point curves into the controller's RAM, trigger curves of set-points stored in the controller's internal memory, retrieve the actual position of the slider, retrieve the electrical current, among others. The main functions used to interface with the controller LinMot C-1250 are shown in Appendix A, Chapter 7.

4.6 Microphone

A condenser microphone consists of a thin membrane located above a solid metal plate. This membrane, or diaphragm as it is also called, has to be conductive. This is usually done with gold-sputtered mylar film. When the sound waves reach the diaphragm, a back and forth movement relative to the back-plate takes place. This movement results in a change of the distance between the two plates forming a capacitor. As a result, the capacitance of this device changes according to the changes

(a) Condenser Microphone (b) Audio Interface

Figure 34: Sampling hardware. The condenser microphone is supplied with 48 V phantom power and connected to the input of the audio interface with an AD converter sampling at 44,1 kHz.

in the sound waves. However, since the capacitance of this device is small, it needs to be processed by a buffer circuit before being connected to an audio interface.

A typical piano recording setup involves multiple microphones located at different positions. For robustness reasons, the control software uses a separated microphone to operate. The selected microphone is a condenser microphone ST-M01 manufactured by Steinberg. Condenser microphones need an external power supply to operate. Phantom power, in the scope of audio equipment, is a DC voltage transmitted through microphone cables and used to polarize condenser microphones. According to standard IEC 61938, phantom power is defined by 12, 24 and 48 V. The selected microphone operates with 48 V and is shown together with the selected audio interface in Figure 34.

4.7 Audio Interface

An audio interface is an equipment used for sound recording and playback. It provides multiple AD converters that allow sampling at different frequencies and bit depths. Also, it provides AD/DA converters with latencies as low as 1 ms.

The selected audio interface is a Steinberg UR-22 MKII. This device offers 2 AD converters with sample frequencies up to 192 kHz at 24-bit resolution. In this thesis, the precise key-triggering system samples the signals at 44.1 kHz. It is interesting to notice that this frequency is approximately two times higher than the

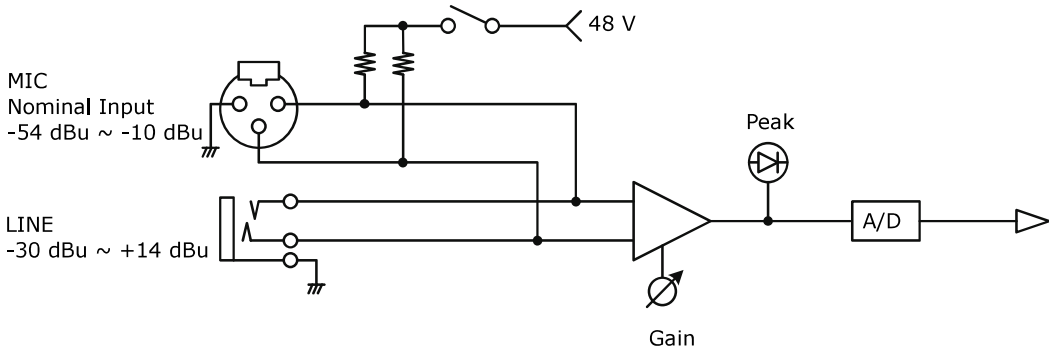


Figure 35: Audio Interface Schematic [38]. Adapted schematic of the audio interface Steinberg UR-22 [38]. The circuit's input impedance is approximately 4 kΩ. The signal captured by the microphones should be within the range +10 dBu to +54 dBu. Peaks are clipped by a peak detector. Finally the signal is sampled by a AD converter with 16 bits sampled at 44,1 kHz. The anti-aliasing filter is not displayed in this diagram.

maximum frequency in the hearing range, approximately 20 kHz. This is related to the Nyquist–Shannon sampling theorem, where the maximum frequency of a signal should not exceed half of the sampling frequency. Moreover, to avoid aliasing the signals need to be processed with a low-pass filter. Since perfect filters do not exist, an additional band of 4.1 kHz is added to the 40 kHz and provides a transition area to the filter. The schematic of the audio interface can be seen on Figure 35.

The sampled signal is sent via USB to the computer, where it is analyzed using the Waveform Audio File (WAV) file format. This format, also known as WAVEform audio format, is a file format introduced by Microsoft and IBM and is used for storing audio in computers. It became known for being the standard used in the audio Compact Disk (CD). The WAV files stores information by using linear Pulse Code Modulation (PCM) bit-streams. The bitstreams are saved in two-channels, usually recorded at sampling rate 44,100 Hz and containing 16 bits per sample, although the format supports bit-depths of 16, 24 or 32 bits and sample rates ranging from 44.1 kHz to 384 kHz. Technically, it is a format without loss and ideal for performing time-domain as well as in the frequency domain analysis, such as performed later in this thesis. In the scope of this work, the recorded signals are saved in WAV format at a sampling rate of 44.1 kHz and a bit-depth of 16 bits per sample. To clarify, this bit-depth represents 65536 quantization levels.

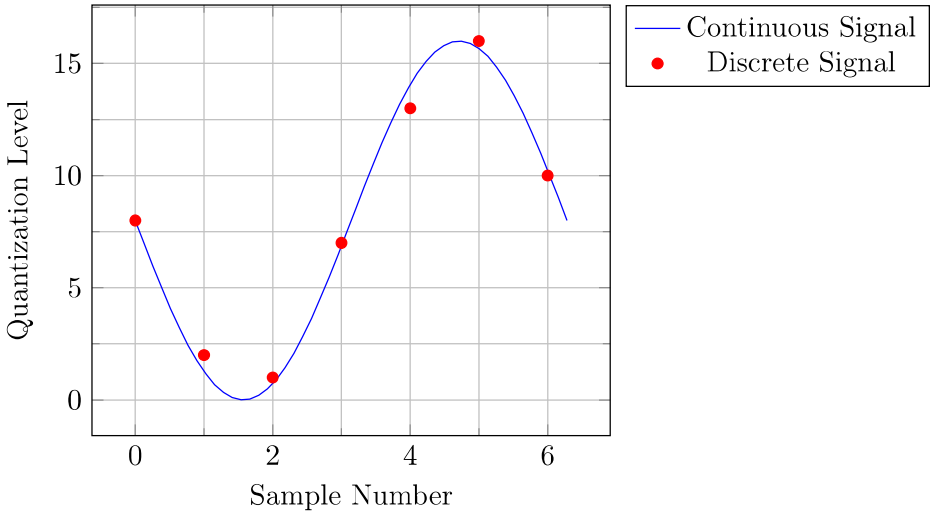


Figure 36: PCM. Example of a signal quantized at 4-bit LPCM. The input signal is divided into 16 amplitude levels. In this thesis, the signals are analyzed using WAV files quantized at 16-bit LPCM and sampled at 44.1 kHz. This bit-depth represents 65536 quantization levels.

4.8 Control Software

The control software operates similarly to a Supervisory Control and Data Acquisition (SCADA) system, where the user can supervise and control the system via a Graphical user interface (GUI). The GUI allows the user to connect to and disconnect from the controller, switch on-off the linear motor, execute the homing operation, read the actual position, read the actual current, write new set-point positions, upload sequences of set-points (ADSR curves) to be followed by the PID controller, calibrate a certain piano key, estimate the optimal ADSR curves to trigger the piano key at multiple dynamic levels, trigger different types of sounds, among many other features. The control software is developed in C#, a multi-paradigm, general-purpose, compiled programming language based on .NET Framework. The GUI of the control software can be visualized in Figure 37.

The control software communicates with two main hardware components: the controller and the audio interface. When interfacing with the audio interface, the control software monitors and records the signals captured by the microphone. This is achieved via the Naudio library [39], a library for audio recording based on the .NET Framework. After being recorded, the signal is analyzed using the MathNet library. The MathNet library offers several mathematical tools for performing signal processing analysis, such as calculation of RMS, implementation of the Fast Fourier Transform

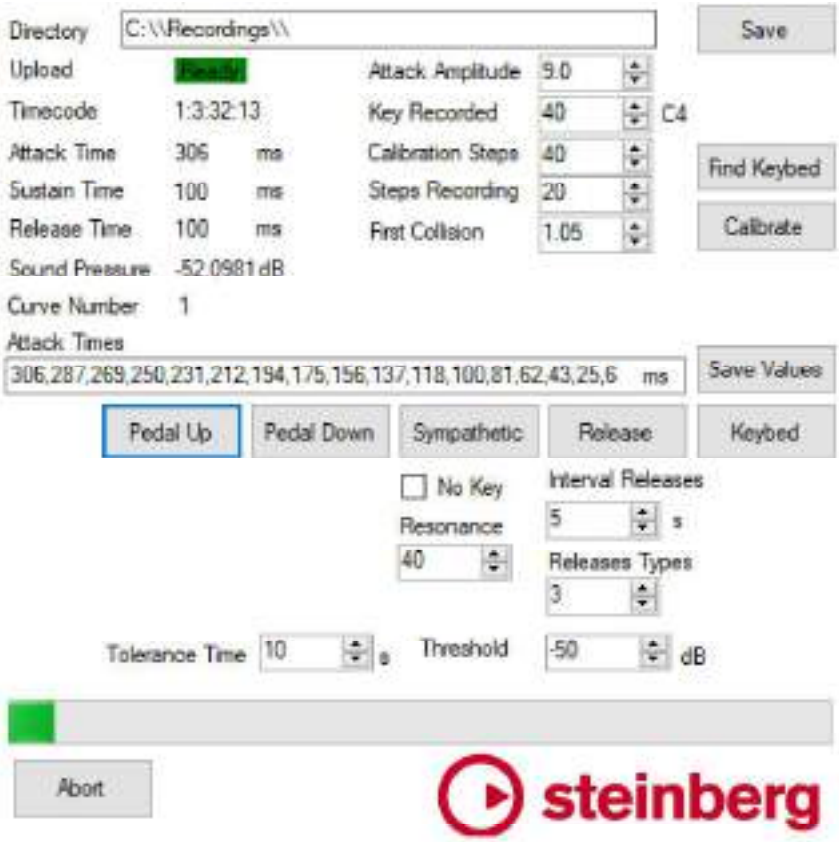


Figure 37: Control Software. The control software supervises and controls the system. Common tasks include the calibration of piano keys, estimation of the keybed bottom distance, and triggering different types of piano sounds at multiple dynamic levels.

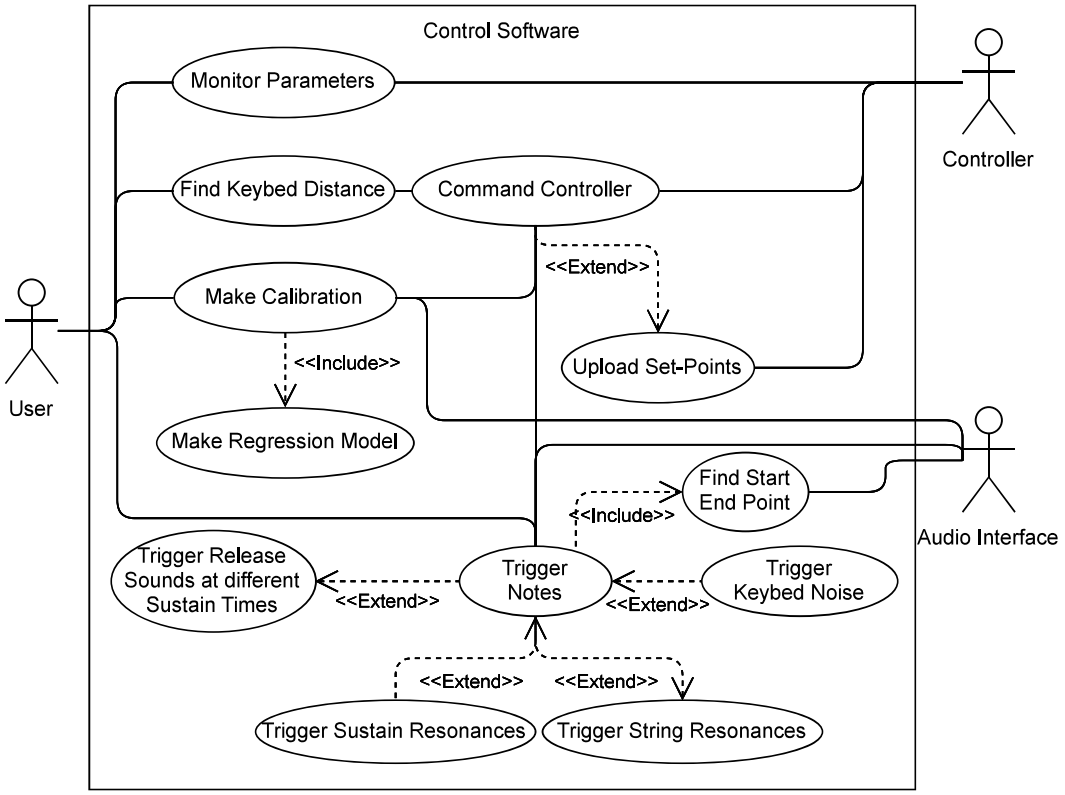


Figure 38: Control Software Use Cases. Use cases diagram showing the main interactions between the actors and the control software. Main use cases include the execution of the calibration procedure, triggering different types of notes in the piano and estimation of markers for cropping the samples.

(FFT) algorithm, different types of window functions (Hann and Hamming windows), among others. The communication between the control software and the actuator controller is executed via a DLL library provided by the actuator manufacturer, as explained in Section 4.5.1.

The Unified Modeling Language (UML) use cases diagram for the control software is shown in Figure 38.

4.9 ADSR Curve Modeling

The ADSR curves represent the actuator slider movement over time. These curves model the behavior of a trained pianist finger while playing the piano, i.e. they represent the exact vertical displacement of a trained pianist finger over time. For the

PID controller, they represent a sequence of position set-points that the controller has to follow over time. Position discontinuities are not allowed and would lead to high speed and acceleration of the actuator slider. As an example, if the set-point position changes from 0 mm to 10 mm, the PID controller tries to reach the new set-point at its maximum rated specifications. In practice, these large changes in set-points lead to an aggressive movement of the actuator slider and create loud noises, which need to be avoided for the sake of not disturbing the recordings.

As described in Chapter 3.7, the ADSR curves are composed of four stages: attack, decay, sustain and release. In this thesis, the attack phase, the decay phase, and the release phase are modeled using cosine-shaped functions while the sustain phase is modeled by a constant value. These cosine-shaped functions are modified to guarantee that transition areas do not have set-point discontinuities and speed discontinuities. The equations representing the ADSR set-point curves are achieved by changing the amplitude, frequency and offset of cosine functions and displayed in Equation (14).

$$z(t) = \begin{cases} 0.5A_a(1 + \cos\left(\left(\frac{2\pi}{2t_a}\right)(t - t_a)\right) : 0 < t < t_a \\ A_d + \frac{(A_a - A_d)}{2}(1 + \left(\cos\left(\left(\frac{2\pi}{2t_d}\right)(t - t_a)\right)\right) : t_a < t < t_a + t_d \\ A_s : t_a + t_d < t < t_a + t_d + t_s \\ 0.5A_r(1 + \cos\left(\left(\frac{2\pi}{2t_r}\right)(t - t_a - t_s - t_r)\right) : t_a + t_d + t_s < t \end{cases} \tag{14}$$

Where:

t:	Time
$z(t)$:	Position at time t
A_a:	Attack amplitude
A_d:	Decay amplitude
A_s:	Sustain amplitude
A_r:	Release amplitude
t_a:	Attack time
t_d:	Decay time
t_s:	Sustain time
t_r:	Release time

The parameter t_a, here named attack time, represents the duration between the beginning of the slider movement (from its initial rest position) to its maximum amplitude A_a. t_a is an extremely important parameter and it is linearly related to the maximum RMS level of the piano sounds. t_d represents the time required for the

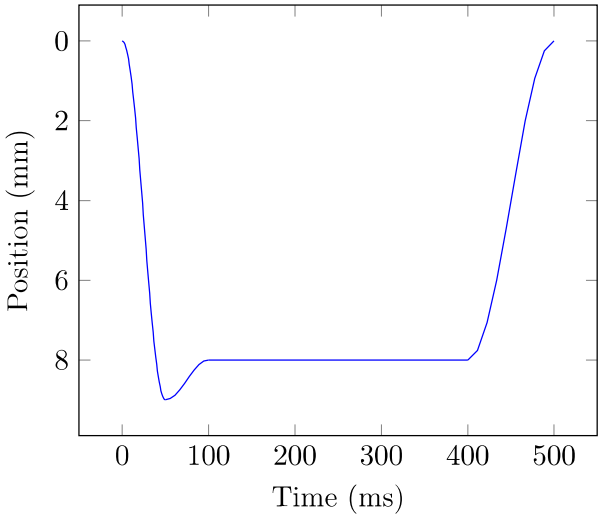


Figure 39: ADSR curves. The ADSR curves are sequence of position set-points that have to be followed by the PID controller and represent the actuator slider movement over time. In order to allow a smooth control, these curves are modeled using cosine-shaped functions.

actuator to return from A_a to A_s. During the sustain phase, the set-point remains constant and equal to A_s. The sustain phase has the longest duration of all phases and lasts t_s. Finally, in the release phase, the actuator slider moves from A_s to the rest position with a duration of t_r. The equations shown in Equation (14) can graphically visualized in Figure 39.

As explained in Section 4.5, the ADSR curves are uploaded to the controller RAM as an array of integers multiple of amplitude resolution. After being uploaded to the controller, these curves can be triggered via the control software.

4.10 Audio Descriptors

In this work, different types of audio descriptors are used for different purposes. The audio descriptors are explained in this section.

4.10.1 RMS Level

One of the goals in a key triggering system is to determine the dynamic levels of sounds produced while pressing a key. The different dynamic levels are related to the hammer speed during the hammer-string collision. During the hammer-string collision, the string oscillates and creates air pressure changes over time. These

oscillations are transmitted by the bridge to the soundboard and finally coupled into air vibrations, which are captured by a microphone. In this work, it is essential to use a good descriptor for audio loudness.

The microphone captures an electrical current which oscillates around zero. This electrical current is converted to a digital value by the AD converter present in the audio interface. The digital signal oscillates around zero and represents the behavior of the sound pressure changes in the environment. A simple mean of the digital signal would result in values close to zero and would not give a proper estimation of loudness. Due to this characteristic, RMS values are used.

In this work, the temporal estimation of loudness is done by a rectangular window $W = \begin{bmatrix} 1 & 1 & ... & 1 \end{bmatrix}$ sliding across the input signal $L_s = \begin{bmatrix} l_1 & l_2 & ... & l_N \end{bmatrix}$, where S is the size of the window and N is the size of the input signal and $S < N$. The window size S is set to be equal two times the period of the fundamental frequency of the note played in order to guarantee that all harmonics are contained within the window. For example, the fundamental frequency of the note A_4 is 440 Hz and, as a consequence, the period of the fundamental component is 2,27 ms. In this situation, the window size is set to be two times larger, i.e. 4,54 ms. Since sampling is taking place at 44,1 kHz, this implies in a window size $S = 200$. With the purpose of obtaining a smooth RMS level descriptor, the window slides at a certain overlap rate O, where $0 < O < 1$. Finally, the a RMS level $p_{s\ RMS\ k}$ is calculated as shown in Equation (15).

$$l_{s\ RMS\ k} = \sqrt{\frac{\sum_{i\ =\ 0}^{S} l_{k.S.O+i}^2}{S}} \ with\ k = 0, ..., \frac{N}{S.O} \tag{15}$$

Where:

k: k_{th} window slide

$l_{s\ RMS\ k}$: RMS level for k_{th} window slide.

After the calculation of the RMS level $p_{s\ RMS\ k}$, these values can be arranged as $L_{s\ RMS} = \begin{bmatrix} l_{s\ RMS\ 1} & l_{s\ RMS\ 2} & ... & l_{s\ RMS\ \frac{N}{S.O}} \end{bmatrix}$. A conversion to decibel unit can be executed via $l_{s\ RMSdB\ k} = 20\log(l_{s\ RMS\ k})$ and thus $L_{s\ RMS\ dB} = \begin{bmatrix} l_{s\ RMS\ dB\ 1} & l_{s\ RMS\ dB\ 2} & ... & l_{s\ RMS\ db\ \frac{N}{S.O}} \end{bmatrix}$ can be obtained. Please notice that the equations have a factor of 20 instead of 10 in their dB equation. This is a result of decibels being defined as power ratios and since these are amplitude signals, they first have to be squared in order to find the power. This leads to a factor of 20 in the equation.

Finally, after calculating a sequence of RMS $l_{s\ RMS\ k}$ values, the maximum value $l_{s\ RMS\ MAX} = max\ (p_{s\ RMS\ k})$ is selected as an estimate of loudness. Typically, the largest RMS value takes place at the attack sound and corresponds to moment of the hammer-string collision. In this thesis, this descriptor is referred to as the maximum RMS level $l_{s\ RMS\ MAX}$.

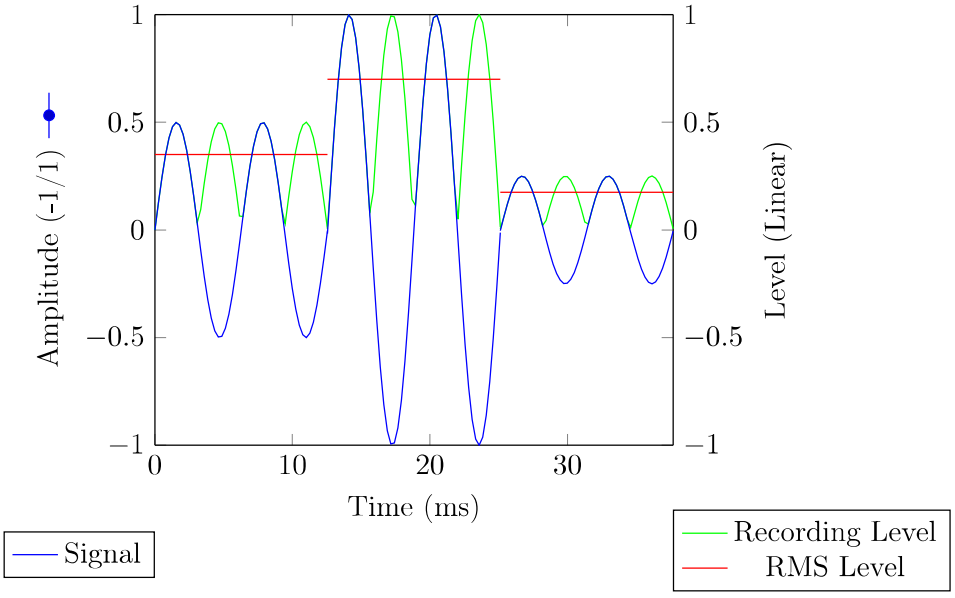
An example displaying the relation between signal, recording revel and RMS level is displayed in Figure 40. In this example the signal amplitude ranges from -1 to 1, similarly to the WAV file format. The recording level is obtained by simply taking the absolute values of the signal amplitude. Finaly, the RMS values are calculated considering a time window equals two times the period of the fundamental frequency of the signal. In this case, only 3 RMS values are calculated. The recording levels and RMS levels are also displayed in dBFS units.

4.10.2 Spectrum Estimation

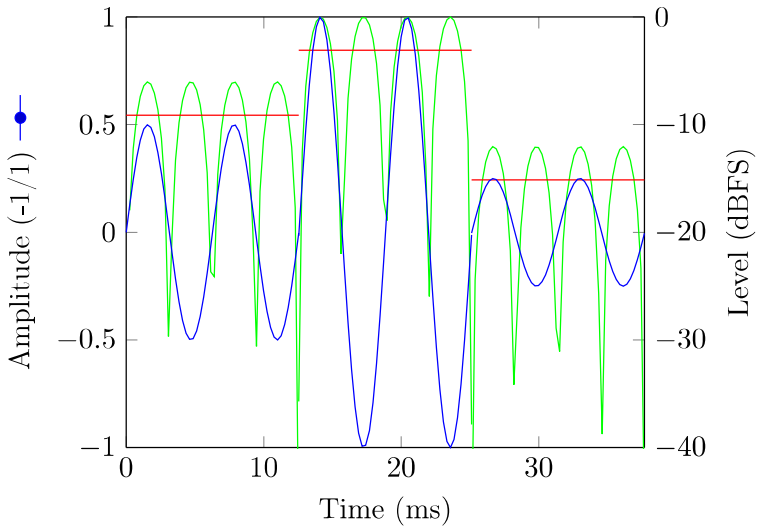
Spectrum is a representation of a signal as a function of frequency and allows to analyze aspects of the signal that are not evident in the time domain [40]. Due to sampling, copies of the spectrum appear at multiples of the sampling frequency f_s, which are then filtered by the audio interface. Spectrum estimation can be executed via parametric and non-parametric methods. Parametric methods, such as auto-regressive and moving average methods, make statistical assumptions about the signal and are computationally expensive [41]. Non-parametric methods include Discrete Fourier Transform (DFT) and Blackman-Tukey [41]. In the music industry, applications normally demand fast calculation of spectrum estimates. Since the DFT estimate allows fast computation by the FFT algorithm, it is one of the main tools used for spectral estimation. The DFT transforms a sequence of N numbers $x[0]$, $x[1]$, ..., $x[N-1]$ into another sequence of N numbers $X[0]$, $X[1]$, ..., $X[N-1]$, the so-called spectral bins. The calculation of the DFT is displayed in Equation (16).

$$X[k] \ = \ \sum_{n=0}^{n\ =\ N} x[n] e^{\frac{-j2\pi kn}{N}} \ \ k \ = \ 0,\ 1,\ ...,\ N-1 \tag{16}$$

Although the FFT is widely used in the music industry, it presents the disadvantage of having spectral leakage, i.e. main lobe and side lobes. Spectral leakage can be treated by changing the window shape [41]. As a consequence, the amplitude of the side lobes can be reduced at the cost of making the main lobe wider [41]. Applications



(a) Linear



(b) dBFS

Figure 40: Signal, Recording Level and RMS Level. This example displays an input signal composed only by the fundamental component of 159 Hz and period equals 2π ms. The amplitude of the sine wave changes from 0.5 to 1 and then finally reaches 0.25. In Figure 40a, the recording level is merely the absolute values of the amplitude. Also in In Figure 40a, the RMS level is calculated considering a sampling windows equals two times the period, i.e. 4π. Considering that the sampling frequency is 44.1 kHz, then a total of 553 samples are required to estimate the RMS level. In this example there is no window overlap. Figure 40a shows the the level values in dBFS

with audio signals typically use Hann and Hamming due to their compromise between the main lobe and suppression of side-lobes. The Hann window is displayed in Equation (17).

$$w_{hann}[n] \ = \ \frac{1}{2}(1 \ - \ cos\left(\frac{2\pi n}{\mathrm{N}-1}\right)) \ 0 \ \leq n \ \leq \mathrm{N} \ - 1 \tag{17}$$

Where:

N: Window size.

4.10.3 Spectral Disparity

The sound recorded while triggering a key can be divided into two main regions with rather similar spectral content. The first region is before triggering the key when no sound is heard and the loudness is typically varying within the noise range. Next, a second region takes place after the hammer-string collision, when the sound pressure of the vibrating strings decreases at a certain decay rate until it finally ceases below the noise threshold. Intuitively, these two regions have similar spectral content, although the second region is likely to have changes in spectral power due to the nature of the decay rate of a note.

A spectral disparity descriptor can potentially be used to indicate features in the transition region, such as the escapement point of a key or the hammer-string collision. An estimation of the spectral disparity over time can be calculated by splitting the input signal into windows, calculating the FFT of each window, and sequentially comparing spectral bins of neighboring windows. One type of spectral disparity descriptor is shown in Equation (18).

$$D_i = \sum_{k=0}^{N} \frac{||X_i[k]| - |X_{i-2}[k]||}{|X_{i-2}[k]|} \tag{18}$$

Where:

k: kth spectral bin.

i: ith disparity estimation.

D_i: Spectral disparity.

$X_i[k]$: Spectral bin at k.

N: Size of FFT estimation

4.10.4 Spectral Centroid

As explained in Section 3.6, different dynamic levels have harmonics with spectral components at different magnitude levels. These changes in the spectral components are perceived as changes in the timbre or tone color of a sound. Higher dynamic levels are expected to have an increase in the power of harmonic components if compared with lower dynamics levels.

The spectral centroid is a measurement used to characterize a spectrum. It can be interpreted as the center of mass of the spectrum and is related to the musical term brightness of a sound [42]. From a spectral point of view, the spectral centroid shifts to the right for high dynamic levels and shifts to the left for low dynamic levels. After calculating the spectrum estimation via FFT, one can estimate the spectral centroid as described in Equation (19).

$$C_S = \frac{\sum_{k=0}^{N-1} f|X[k]|}{\sum_{K=0}^{N-1} |X[k]|} \tag{19}$$

Where:

C_S:	Spectral centroid.
k:	kth spectral bin.
$\|X[k]\|$:	Magnitude of the kth spectral bin

The spectral centroid can be converted to a frequency by $f = \frac{k*f_s}{N}$.

4.11 Calibration

The calibration procedure aims to get a profile of the RMS level sensitivity of a key by exciting it using ADSR curves with different attack times. Long attack times result is a soft touch, while short attack times lead to struck touch. The calibration procedure decreases the attack time of the ADSR curves from a large value, likely not to cause a hammer-string collision, up to the smallest value admitted by the controller. This allows one to have an overview of the loudness profile of a key. To illustrate, the calibration procedure decreases the attack time from 400 ms to 10 ms in fixed steps that can be selected by the user. For each attack time, the maximum RMS level is calculated, as described in Equation (4.10.1). Another important part of the calibration procedure is determining the keybed bottom distance, which leads to the estimation of the attack amplitude in the ADSR curve.

4.11.1 Keybed bottom Distance

The maximum vertical displacement of a key, also named keybed bottom distance d_{keybed}, is used to define the attack amplitude A_a in the ADSR curves. The maximum displacement allowed is estimated by slowly moving down the actuator slider while monitoring the current consumed by the linear motor. In practice, this achieved by increasing the position set-point in small steps (such as 0.1 mm) and waiting for the linear motor current to settle before updating the next set-point value (updates every 100 ms satisfy the settling time). When the key gets in touch with the keybed bottom, the current increases exponentially. This point is detected when the current exceeds a threshold limit (typically 0.5 A). This point is considered to be the keybed bottom and represents the maximum allowed vertical displacement of a key. The A_a can be estimated as $\text{A}_\text{a} = d_{\text{keybed}} - 1$. Algorithm 1 shows the algorithm used to detect the keybed bottom distance.

Algorithm 1 : Detetion of keybed bottom distance

```
1  d_actual = 0 mm;
2  d_keybed = 0 mm;
3  while d_actual < 10mm do
4  |   moveToPosition(d_actual);
5  |   if I_controller > 0.5A then
6  |   |   d_keybed = d_actual;
7  |   |   A_a = d_actual − 1;
8  |   |   d_actual = 0 mm;
9  |   |   moveToPosition(d_actual);
10 |   |   break;
11 |   d_actual = d_actual + 0.1;
12 end
```

4.11.2 Detection of first collision

For long attack times (typically longer than 400 ms), the actuator slider presses the key so slowly that the hammer moves towards the string but no hammer-string collision takes place. In this situation, only the noise resulting from the collision between the actuator slider and the key is perceived. If the attack time continues to decrease, the actuator slider presses the key faster and faster and will, at a certain attack time, lead to a hammer-string collision and produce a sound. Determining the exact attack time needed to achieve the first hammer-string collision is an important

task of this thesis. It allows determining which data-pairs attack time and RMS level lead to changes in loudness. The data-pairs are required for fitting a regression model explained in Chapter 4.11.3.

The estimation of the first sound onset is done by decreasing the attack times while averaging prior estimations of loudness. New values of loudness are compared to the average values and, if the difference is bigger than a certain percentage, this point is set to be the first onset of sound. Algorithm 2 shows the algorithm used to detect the first sound onset.

Algorithm 2 : Detection of first sound onset

```
1  moveToPosition(0 mm);
2  averageSoundPressure = 0 dB;
3  t_a = 400ms;
4  N = 1;
5  for t_a > 10ms do
6  |  executeADSRCurve (t_a);
7  |  soundPressure = readSoundPressureMicrophone();
8  |  averageSoundPressure = (N-1)*averageSoundPressure/N + soundPressure /
   |    N;
9  |  if soundPressure > 1.05*averageSoundPressure then
10 |  |  First Onset of Sound Found;
11 |  |  break loop;
12 |  N = N + 1;
13 |  t_a = t_a − 10ms;
14 end
```

4.11.3 Regression Model

The calibration procedure produces pairs of measurements relating the maximum RMS levels and their respective attack times. Some pairs are a result of a hammer-string collision while others are not. In a piano triggering system, the data related to a no-collision situation can be ignored. By using the results of the algorithm described in Section 4.11.2, it is possible to select only the pairs that lead to a hammer-string collision.

The final pairs obtained in the calibration procedure contain information about the maximum RMS level for the whole range of attack times. Short attack times are expected to have a high maximum RMS level while long attack times are expected to have low maximum RMS level. The behavior of this curve is rather linear and

investigated in detail later in the Section 5.

Later in Section 4.12 it needed to have a mathematical function relating RMS level to the attack time. This function is used for predicting new values of attack times and is explained in detail in Section 4.12. This function is obtained by using the LLS technique.

In a nutshell, the measurements $t_1, ..., t_N$ are the independent variable and stored in a vector ϕ while the measurements $y(1), ..., y(N)$ are the dependent variable and stored in a matrix y_N. ϕ and y_N are displayed in Equation (20).

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \quad \phi = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_N \end{bmatrix} \tag{20}$$

The idea behind the LLS is to find θ that minimizes the sum of the squares of the prediction errors. The cost function for the LLS problem is given by $f(\theta) = \frac{1}{2}||y_N - \phi_N\theta||_2^2$ and is a convex function [43]. For the key triggering system, given the maximum RMS levels obtained during the calibration procedure $\{t_i\}_{i=1}^N$ with the corresponding attack times $\{y_i\}_{i=1}^N$, then it is possible to find a parameter vector $\theta = (\theta_1, \theta_2)$ that satisfies the prediction function $p(t;\theta) = \theta_1 + \theta_2 t$ which gives a prediction y for attack time t [43]. The optimization problem is written in Equation (21).

$$\underset{\theta \in \Re^2}{min} \frac{1}{2} \sum_{i=1}^{N} (y_i - p(t_i;\theta))^2 = \underset{\theta \in \Re^2}{min} \frac{1}{2}||y - \phi \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}||_2^2 \tag{21}$$

Since the cost function is convex, the solution is found by setting the gradient to zero and isolating the minimizer θ^*. It must be noted that the $\left(\phi^T\phi\right)^{-1}\phi^T$ represents the pseudo inverse of the matrix ϕ only under the condition that $\phi^T\phi > 0$. The solution for the LLS problem it is displayed in Equation (24).

$$\nabla f(\theta^*) = 0 \tag{22}$$

$$\phi^T\phi\theta^* - \phi^T y = 0 \tag{23}$$

$$\theta^* = \left(\phi^T\phi\right)^{-1}\phi^T y \tag{24}$$

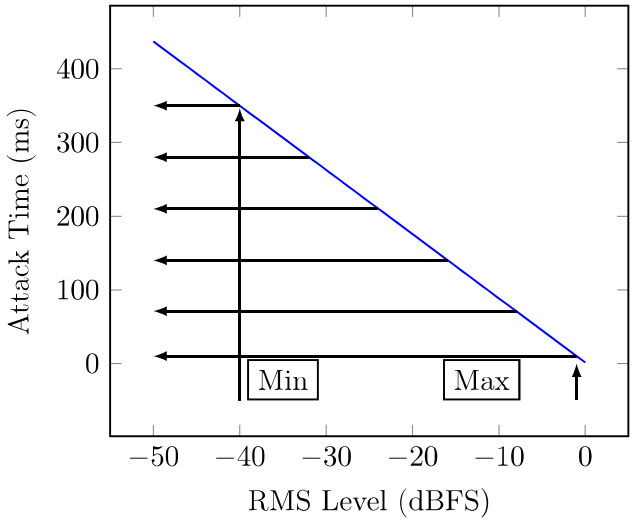


Figure 41: Calculation of Attack Times. Example displaying calculation of six attack times used to achieve 6 dynamic levels of a key. In this example the maximum RMS level is -40 dBFS while the minimum RMS level, calculated using the first sound onset algorithm, is -40 dBFS. By knowing the minimum and the maximum RMS levels and the linear function obtained through the LLS model, it is possible to calculate values of attack times equally separated in loudness, in this case, 10 ms, 71 ms, 140 ms, 210 ms, 280 ms, and 350 ms. These attack times allow achieving six dynamic levels of a piano key.

4.12 Triggering notes

After calibrating the key, obtaining data-pairs relating RMS level and attack time and solving LLS problem, a linear curve is obtained and the values of attack times required to obtain the different dynamics can be derived from the RMS levels. Figure 41 shows and example of calculation of the attack times used for triggering the linear actuator. After obtaining different values of attack times, the system can sequentially trigger the key at different dynamic levels.

4.12.1 Notes without resonances

Notes without sustain resonances, also known as notes with the pedal up, are triggered by simply executing ADSR curves with the attack times calculated using the LLS model. Start markers used for sample cropping are suggested at 50% of the attack time, the point where the speed reaches its maximum value. End markers are set when the loudness falls below the noise threshold. Only after this point, the

system can trigger the next ADSR curve with the next dynamic level.

4.12.2 Notes with resonances

As mentioned in Section 3.10, to create realistic piano samples one needs to record the sustain and sympathetic resonances of a key. Sustain resonances are recorded by keeping the sustain pedal down and then pressing the key with the actuator at optical attack times suggested by the LLS model. Sympathetic resonances are recorded by keeping certain piano keys down and then triggering the actuator slider at the optimal attack times proposed by the LLS model.

It should be noted that resonances cannot be recorded separately from the attack sound. A common strategy used to obtain separated resonance sounds is to record the sounds with and without the resonances and subtracting one signal from the another. This method can be used if the initial phase of both signals is matched. The problem of using this method is obtaining sounds with similar phase and amplitude behavior over time. Normally, human players are unlikely to press the piano key with the same movement. The precise piano key trigger system is expected to reach better results than human pianists due to the precision and accuracy of the linear actuator.

4.12.3 Notes with release sounds

In Section 4.12.1 and Section 4.12.2, the key is pressed using different ADSR curve and then the RMS level is monitored until it falls below a threshold. The recording of release sounds differs in the manner that the actuator slider is removed from the key while the sound is being produced. In this situation, the actuator slider moves up the key and different types of release sounds are produced, namely soft, medium and hard releases. In practice, this is achieved by simply changing the release time in the ADSR curves. Furthermore, these release sounds need to be recorded at different time-stamps. To illustrate, consider that the key C_1 is pressed by the actuator slider with the attack time of 10 ms and produces sound for approximately 60 s until it finally ceases. In this example, the release sounds (soft, medium, hard) would have to be performed three times at each of these times: 5 s, 10 s, ... 60 s. This amounts to the execution of 36 different ADSR curves.

Type	Timecode In	Timecode Out	Note	Length	ID
SimpleNote	00:00:03:31	00:00:10:22	C_1	7	1
SustainedNote	00:00:19:06	00:00:23:46	A_4	5	2
ReleaseSound	00:00:31:33	00:00:43:14	B_5	12	3

Table 5: CSV file. Example of CSV file created by the key triggering system. It shows the note, how the note is triggered, start and end point of the note based on the MIDI time-code received from the DAW.

4.13 Synchronization

The actual recording of musical instruments is typically done in a DAW. They allow advanced editing and batch processing of samples and are the standard tool for processing samples in the music industry. During a recording session of a piano, typically 12 to 16 microphones are continuously recorded in a DAW. Just to illustrate, a typical recording session can take as much as 12 hours, resulting in up to 192 hours of continuous recordings. These recordings need to be cropped at the beginning and end of each note.

Correct cropping of sound events in time is essential for the creation of realistic piano samples. Ideally, the start point of a sound should consider the delay present in the piano action mechanism, often in the milliseconds range. In practice, this is executed by trimming the sound at the hammer-string collision point and after adding the time delta between the escapement point and hammer-string collision via scripting language, such as Lua. The endpoint is more flexible and takes place when the RMS level falls below a certain noise threshold, typically -50 dB.

In the scope of this work, the sample recording and the key triggering system are running in independent computers. This architecture allows higher robustness, i.e. the ability to cope with errors during a recording session. The piano trigger system saves the information needed for cropping of the start and the end of the samples in a CSV file, which is then loaded into the DAW during the post-recording phase thus allowing the cropping of the samples. A typical file created by the trigger system is shown in Table 5.

During a recording session, the complete setup is composed of the recording computer, the key triggering system computer and the linear actuator controller, all of them with different clock sources that need to be synchronized. To ensure the correct cropping of the recording files, a common clock source is used: the MIDI time-code in the DAW. The MIDI timecode is sent from the recording computer to

Piece	Data Byte	Significance
0	0000 ffff	Frame number lsbits
1	0001 000f	Frame number msbit
2	0010 ssss	Second lsbits
3	0011 00ss	Second msbits
4	0100 mmmm	Minute lsbits
5	0101 00mm	Minute msbits
6	0110 hhhh	Hour lsbits
7	0111 0rrh	Rate and hour msbit

Table 6: MIDI time code quarter messages 8 quarter messages provide the full timecode stamp from the DAW in a format hh:mm:ss:ff.

the control software computer via a physical MIDI cable connected to the MIDI ports in the audio interface. The time-code is composed of MIDI quarter-frame messages, where a set of eight quarter-frame messages contains information about the hour, minute, second and frame number. This information is typically represented in the format hh:mm:ss:ff. Each quarter-frame message is delivered with a fixed time interval depending on the computer internal clock, typically 10 ms. As a result, the complete time-code message is only formed after approximately 80 ms. The control software ensures that ADSR curve is executed only when a complete time-code message is received. Examples of time-code quarter messages are shown in Table 6.

4.14 Installation

The correct installation of the linear actuator is essential for the system to function properly. The linear actuator has precision in the micrometer, which becomes meaningless if the actuator is not well fixed while pressing a key. Also important, the noises produced by the linear actuator need to be treated to achieve a satisfactory sample recording.

4.14.1 Mechanical Structure

To fix the linear actuator on the top of the piano key, a mechanical structure composed of two tripods and transverse bars is used. They avoid undesired vertical displacements of the actuator during the with the key. The complete structure weighs about 40 Kg and allows precise positioning of actuator slider on the top of the piano key by the adjustment of nut screws. It should be mentioned that the triggering

system achieves the precision of the linear actuator due to the use of Algorithm 1. Algorithm 1 estimates the attack amplitude A_a at a certain current value defined by the user. As a consequence, small vertical misalignment of the actuator slider should not impact the obtained sounds. 42 shows the construction of the mechanical structure.

4.14.2 Noise Treatment

Noises play a critical role in the professional sample recording. In a studio recording environment, there are plenty of sources of noises that need to be treated. Incorrect noise treatment can destroy the final samples and should be avoided. Post-recording noise treatment methods such as filtering or noise subtraction are considerably less efficient than treatment before the recording session. Pre-recording noise treatment methods include damping the noise sounds by using soundproof materials such as acoustic foam and acoustic cases.

In the scope of the piano trigger system, the major sources of noises are the transformer in the power supply, the windings of the linear actuator, the friction between the actuator slider and the linear motor, the slider-key collision, and the key-keybed bottom collision. To damp the noise resulting from the slider-key collision, a hammer felt is attached to the end of the actuator slider as displayed in Figure 43. The most problematic noise is the one produced by the windings in the linear motor. It is important to notice that the hum of the transformer in the power supply and the noise from the windings in the linear motor are classified as electromagnetically excited acoustic noises. These noises take place due to the excitation of materials by electromagnetic forces. To treat the hum in the transformer, the power supply is kept in a room separated from the recording setup by using an 8 meter supply cable. The noises produced by the windings of the linear motor are in frequencies over 1,2 kHz. In the scope of this thesis, these noises are treated by an insulation box built with wood and internally covered with polyurethane foam. Polyurethane foam provides good absorption rates in high frequencies [44].

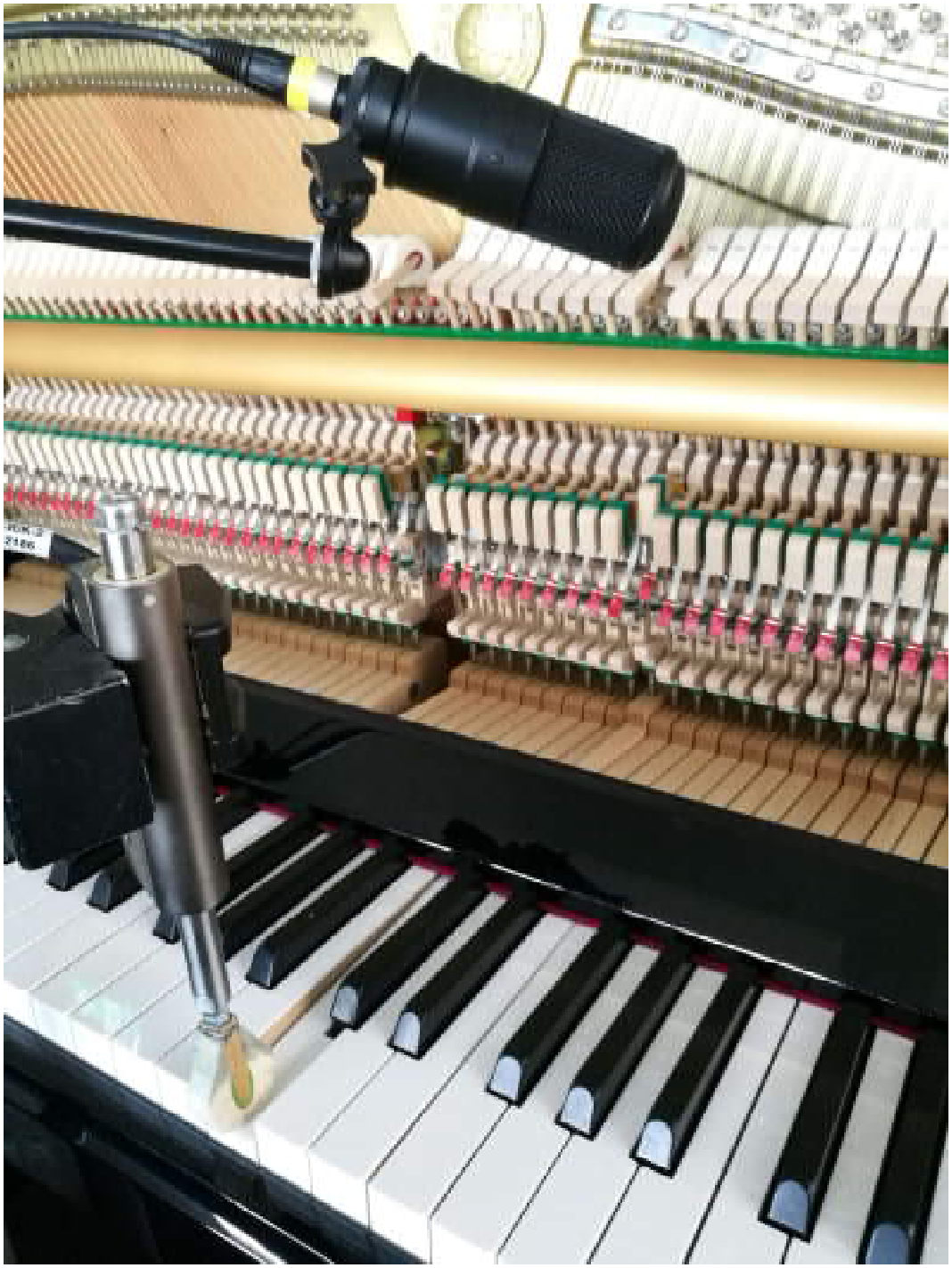


Figure 42: System Setup. The mechanical structure allows correct positioning of the actuator on the top of piano key. The microphone is installed at close to the hammer-string contact point.

Figure 43: Hammer Felt. In order to reduce the noise produced by the contact between the actuator slider and the key, a piano hammer felt was attached to the tip of the actuator slider.

5 Experiments

This chapter presents the key results of the research involving the precise key triggering system for piano sample recordings. A comprehensive analysis of results is provided for different key topics, such as the behavior of ADSR curves for different curve parameters, evaluation of used audio descriptors, evaluation of suggested algorithms, behavior of black keys, estimation of escapement point, evaluation of PID tuning, among other topics.

While planning, designing, developing and testing, the precise key triggering system was engineered to work with any type of piano. Nevertheless, the majority of experiments are executed with a Yamaha Upright Piano Model B3 [45] unless otherwise specified.

5.1 Evaluation of ADSR Curves

The system is triggered using ADSR curves representing a sequence of position set-points to be followed by the PID controller. As explained in Section 4.9, these curves are modeled using cosine-shaped functions. However, other types of ADSR curve shapes could also potentially be used. In this section it is mostly discussed the differences between the linear and cosine-shaped ADSR curves and why the use of the latter one achieves better results.

5.1.1 Linear-shaped Curves

ADSR curves represent the actuator position over time. In mathematical terms, speed is the first derivative of the position with respect to time while the acceleration is the second derivative over time. Despite not having discontinuities, linear-shaped ADSR curves present sharp edges in the transition between phases. A consequence of these sharp edges is that the first and second derivatives present discontinuities, causing the PID controller to reach high speeds and accelerations.

These high speeds and high accelerations often reach the nominal specification of the linear actuator and, as a consequence, collisions between the actuator slider and

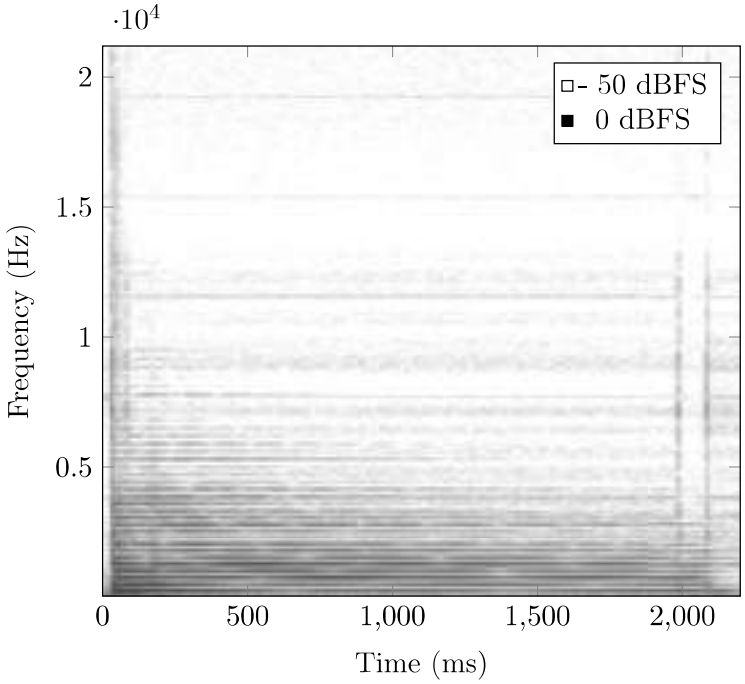


Figure 44: Spectrogram of actuator noise. Histogram displaying the actuator noise while triggering a key C_4 with a linear-shaped ADSR curve. The actuator noise is noticeable in the five edges of the linear-shaped ADSR curve. In this experiment, the curve parameters are $t_a = 10$ ms, $A_a = 9$ mm, and $A_s = 8$ mm. In this histogram the FFT is calculated with 1024 samples, overlap of 50%, Hann window, sampling at 44.1 kHz. The information is stored in 16 bit PCM WAV files.

the linear motor take place. In practice, these collisions produce loud noises and are unwanted since they destroy the sample recordings. Figure 44 depicts these noises in the transition regions of the ADSR curve.

5.1.2 Cosine-shaped Curves

Cosine-shaped functions have the advantage of not having sharp edges in their transition regions. These smooth input curves allow the PID controller to be less aggressive while handling new set-points. In practice, the small changes in the position set-point are perceived as a continuous movement of the actuator slider over time. This characteristic is strongly related to the inertial movement of the actuator slider. Figure 45 shows an example of a ADSR curve uploaded to the controller.

The proposed model with cosine-shaped ADSR curves presents good noise reduc-

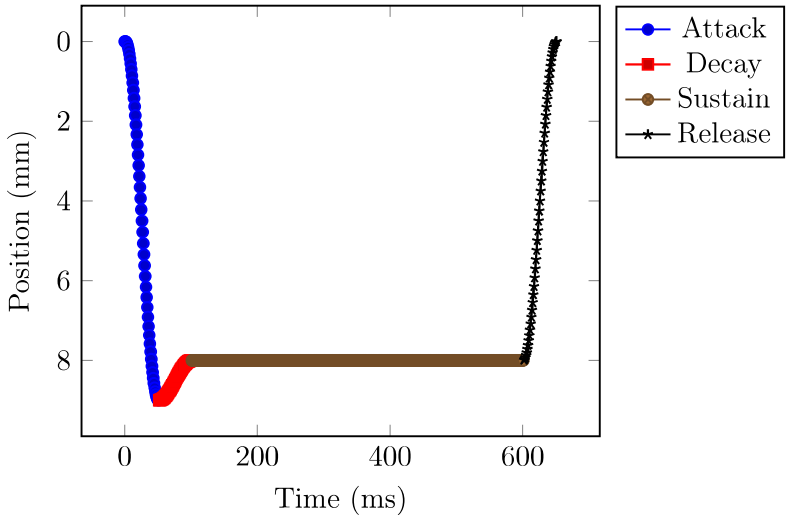


Figure 45: Real ADSR Curve. Example of a typical ADSR curve used to trigger the piano key. It consists of a sequence of position set-points executed every 1 ms by the system controller.

tion since neither speed nor acceleration reaches values comparable to the nominal specification. In practice, the PID controller is much quieter while executing smooth input signals. The first and the second derivative of the position set-points present maximum values compatible with the literature, such as the results obtained by Askenfelt [13]. Figure 46 displays the first and second derivative of the position set-points.

5.2 Evaluation of RMS level

In Section 4.10.1, the maximum RMS level is suggested as a descriptor for loudness. An experiment was conducted in order to determine the relation between the maximum RMS level, attack time and attack amplitude for different keys. The results can be seen in Figure 47.

The analysis of Figure 47 allows one to make several conclusions. The first conclusion is that lower attack times produce the first sound onset at higher attack amplitudes. In practical terms, that means that for low attack times, the actuator slider has to go deeper to reach the first onset of sound. While higher attack times require smaller attack amplitudes for the first sound onset, i.e, the actuator does not need to move so deeply to obtain the first onset of sound.

Another conclusion is that, for a certain fixed attack amplitude, shorter attack

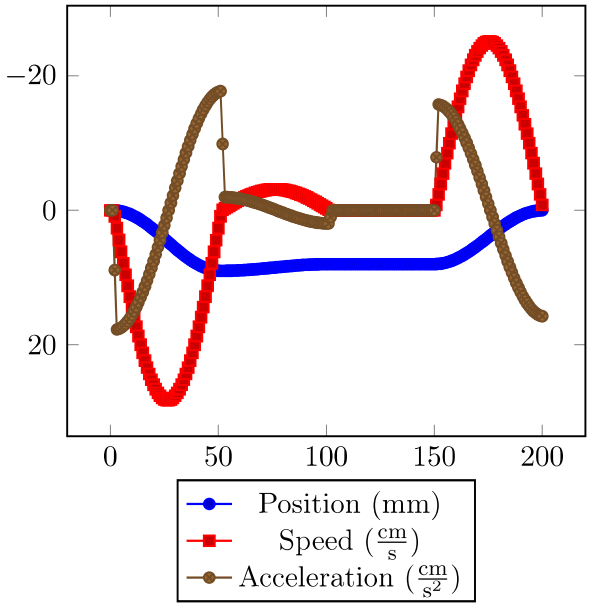


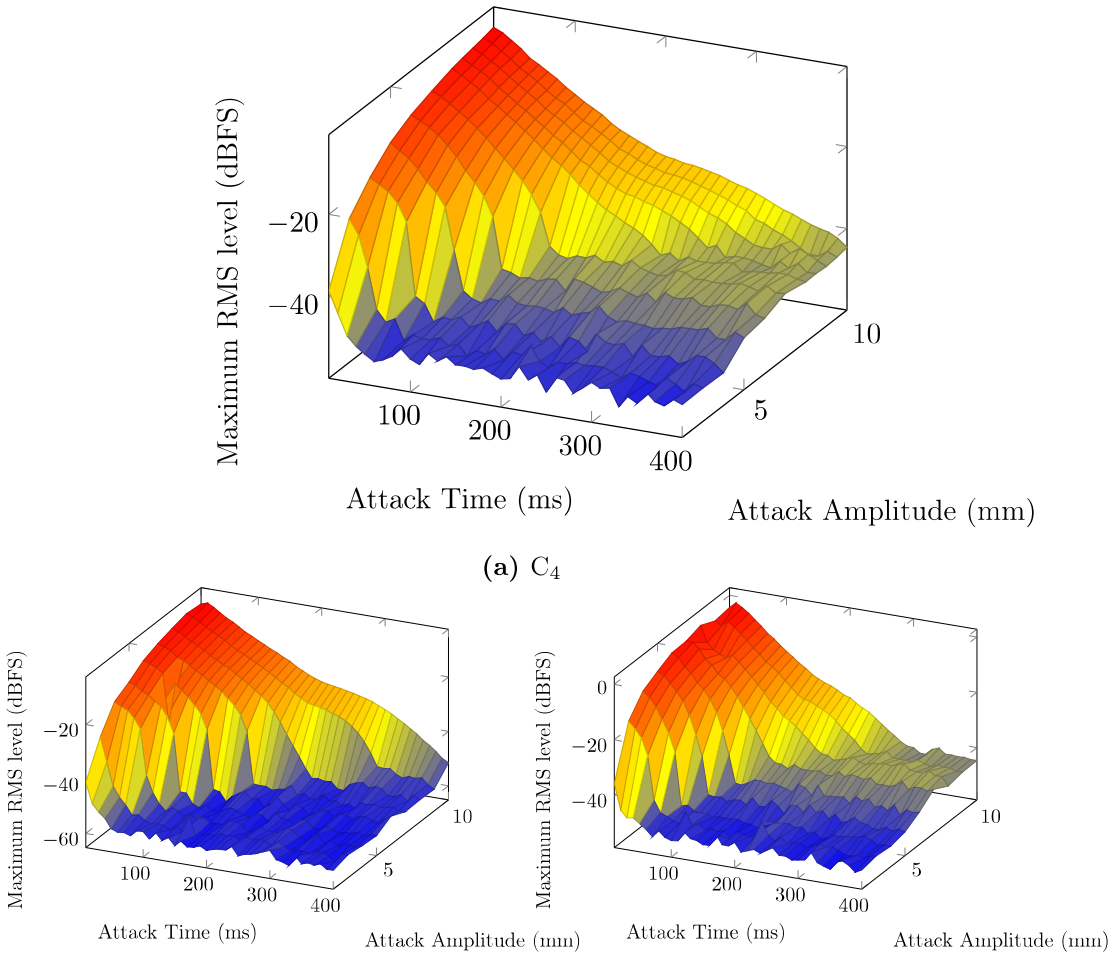
Figure 46: Position, Speed and Acceleration. Position set-points over time and its first and second derivative. The first derivative represents the speed and the second derivative represents the acceleration of the actuator slider over time.

times produce higher maximum RMS levels while longer attack times produce lower maximum RMS levels. Furthermore, for a fixed attack amplitude, the maximum RMS level has a considerable linear behavior over the attack time only at high attack amplitudes. For low attack amplitudes, the maximum RMS level does not show a linear behavior, but rather exponential. This conclusion implies that the piano trigger system shall work ideally with high attack amplitudes.

Another conclusion is that onsets of sound take place at all attack times only for high attack amplitudes, i.e. 8 mm or more. Low attack amplitudes have attack times that do not necessary produced sound, i.e. no hammer-collision takes place. Since sounds are produced for all attack times only at high attack amplitudes, a deeper investigation was performed in the range of attack amplitudes between 8 mm and 10 mm. The results are shown on Figure 48.

5.2.1 Repeatability

In a recording session, it is, in some cases, necessary to play the piano key the same way but at different moments. This could happen, for instance, if a sample has noise or if one is trying to record sounds with and without resonances in order to



(a) C_4

(b) C_1

(c) C_8

Figure 47: Maximum RMS level for Different Attack Times and Attack Amplitudes. Maximum RMS level calculated with adaptive window size equals two times the size of the fundamental frequency for the keys C_1, C_4 and C_8. The attack times ranged from 400 ms to 10 ms in steps of 10 ms. The attack amplitude ranged from 2 mm to 10 mm. The experiment was performed in an Yamaha Upright Piano. The microphone was located approximately at 10 cm from the hammer-string contact point. The experiment was performed using ADSR curves with $t_d = 50ms$, A_s equals 90% of A_a, $t_s = 1000ms$ and $t_r = 100ms$.

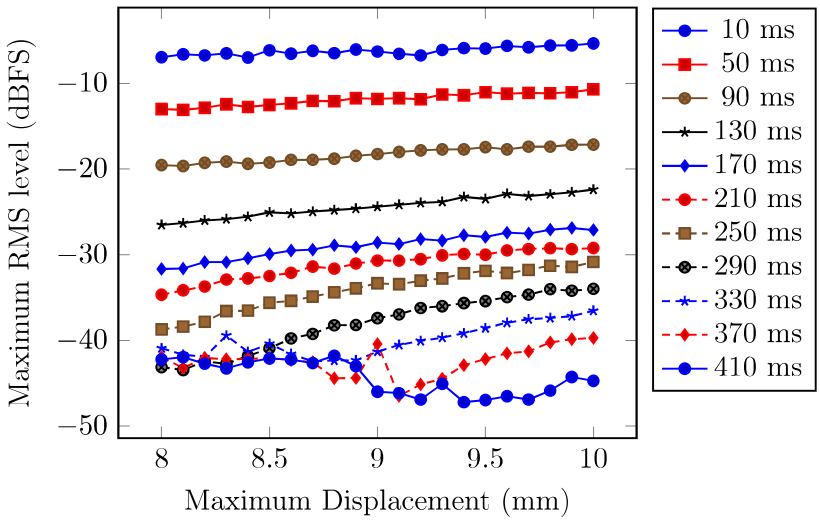


Figure 48: Maximum RMS level in the range 8 mm to 10 mm. This experiment shows the maximum RMS level for attack amplitudes in the range between 8 mm and 10 mm with steps of 0.1 mm. The experiment was performed for different attack times in a key C_4 of a Yamaha Upright Piano. The microphone was located at approximately 10 cm from the hammer-string contact point. t_a ranged from 410 ms to 10 ms in steps of 40 ms. The experiment was performed using ADSR curves with $t_d = 50ms$, A_s equals 90% of A_a, $t_s = 1000ms$ and $t_r = 100ms$.

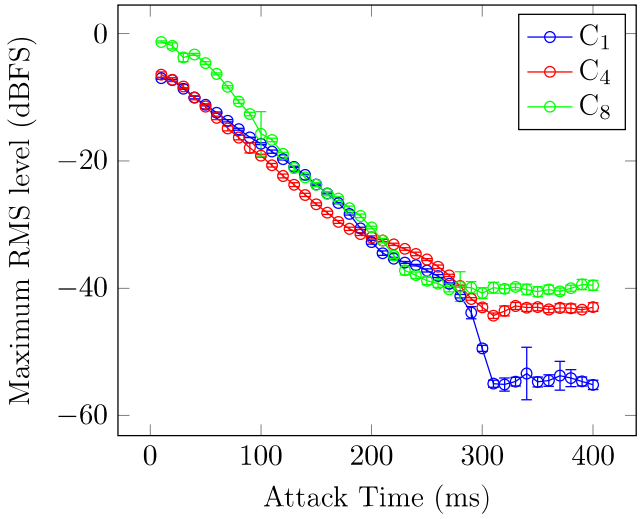


Figure 49: Maximum RMS level mean and standard deviation. Maximum RMS level over different attack times. The maximum RMS level is estimated with adaptive window size equals two times the size of period of the fundamental frequency of the key. The microphone was located approximately 10 cm from the hammer-string contact point. The attack times 10 ms to 400 ms in steps of 10 ms. 10 measurements were performed for at each attack time. The experiment was executed using ADSR curves with A_a = 9 mm, t_d = 50 ms, A_d = 80 mm, t_s = 1000 ms and t_r = 100 ms. The experiment was performed in a Yamaha Upright Piano.

subtract those sounds afterward. Due to situations like this, a reliable piano trigger system should have good repeatability. Repeatability is related to the closeness of successive measurements executed under the same conditions of measurement. While evaluating repeatability one should consider that the experiment is executed in the same studio, in the same piano, with the same observer, using the piano trigger system and following the same procedure repeated over a short period.

An experiment was executed to estimate the standard deviation and variance of the maximum RMS level for different attack times. The experiment was performed in the keys C_1, C_4 and C_8 of a Yamaha Upright Piano. As a result of this experiment, the average standard deviation considering all attack times is 0.47 dBFS, a small value in terms of loudness changes. The experiment results are displayed in Figure 50.

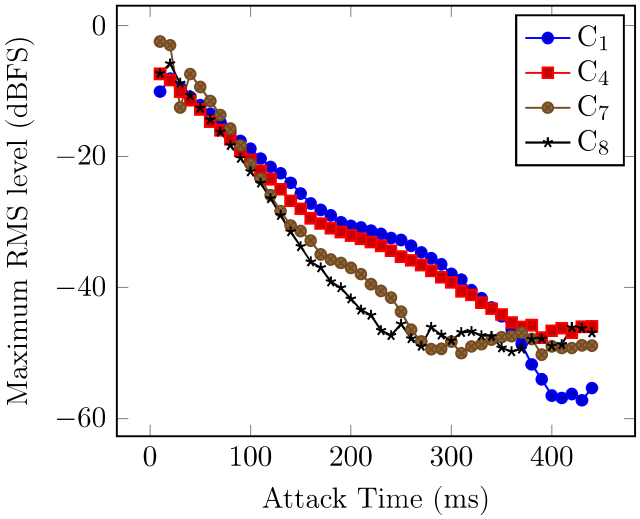


Figure 50: Maximum RMS level vs attack time. Maximum RMS level for different attack times. The experiment was performed for the keys C_1, C_4, C_7 and C_8. The maximum RMS level was calculated with adaptive window size equals two times the size of the period of the fundamental frequency. The experiment was performed in an Yamaha Upright Piano. The microphone was located at approximately 10 cm from the hammer-string contact point. The attack times ranged from 440 ms to 10 ms in steps of 10 ms. The experiment was executed using ADSR curves with $A_a = 9$ mm, $t_d = 50$ ms, $A_s = 80$ mm, $t_s = 1000$ ms and $t_r = 100$ ms.

5.3 Influence of Attack Time

An experiment was performed to investigate the influence of the attack time in the maximum RMS level for different piano keys. Understanding the influence of the attack time is a key point for triggering the piano keys at dynamic levels. As indicated in Section 5.2, short attack times have high maximum RMS levels while long attack times have low RMS levels. This experiment aims to understand the relation between the two variables. The results are shown in Figure 50.

The analysis of Figure 50 allows one to conclude that the relationship between maximum RMS level and the attack time is fairly linear when hammer-string collisions take place. It should be pointed out that for high attack times no hammer-collision takes place and therefore the maximum RMS level stays within the noise range.

The linear behavior of this experiment justifies the use of a LLS model in the system. Again, the system obtains several pairs of attack times and RMS levels during the calibration, which are later fitted into a LLS model used to obtain new

values of attack times equally separated in terms of RMS levels. As an example, if the user wants to trigger the piano key at 20 different velocities, then the LLS model provides 20 values.

5.4 Evaluation of Spectral Centroid

In music, the brightness of a sound is tightly related to the perception between different dynamics of a sound and is an indicator of the amount of high-frequency content in a sound. The brightness of a sound is commonly referred to as tone color and can be estimated via the spectral centroid. In Section 4.10.4, the spectral centroid is suggested as an audio descriptor for loudness due to the increases in overtones. An experiment was performed to analyze the relation between the spectral centroid and the attack time of the ADSR curve. The results are shown in Figure 51.

Figure 51 shows that for low attack times, i.e. when no hammer-string collision takes place, the spectral centroid is higher than the fundamental frequency of the note, 440 Hz. In this region, the spectral centroid represents merely the centroid of the noise. For intermediary attack times, i.e. close to the first hammer-string collision, the spectral centroid approaches the fundamental frequency, i.e. 440 Hz, although it never reaches this value. For short attack times, i.e. when the hammer-string collision produces loud sounds, the spectral centroid increases rapidly due to the increases in the overtones, thus moving away from 440 Hz.

By analyzing Figure 51, one can conclude that the spectral centroid is not a good descriptor for dynamic levels at long attack times. At long attack times, the spectral centroid represents merely the centroid of the noise and has a rather random behavior. Another characteristic is that the spectral centroid tends to oscillate over the sound. This situation can be visualized by analyzing the power of the harmonic components in different parts of a signal. For that reason, the use of the spectral centroid requires association with stabilization methods, such as averaging. Due to this increased complexity, the piano trigger system does not use the spectral centroid estimator in the calibration procedure, although further investigation is still required.

5.5 Evaluation of Spectral Disparity

An example displaying the spectral disparity for a piano sound is shown in Figure 52.

The spectral disparity estimation does not give information about loudness. To illustrate, a pressed touch can show a higher spectral disparity peak than a struck

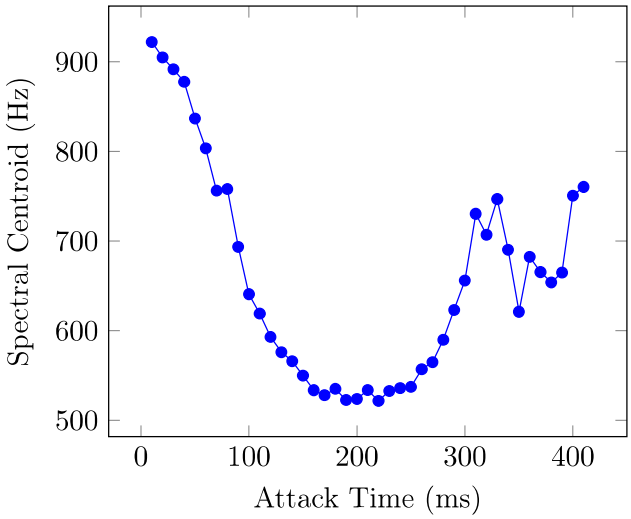


Figure 51: Spectral Centroid. The spectral estimation for a C_4 is executed with the FFT algorithm considering 1024 samples of the recorded sound taken immediately after the maximum RMS level. Hann window is used. Finally, the spectral centroid is calculated considering the first 128 frequency bins of the FFT. These comprise the frequency range between 0 Hz to 5512 Hz.

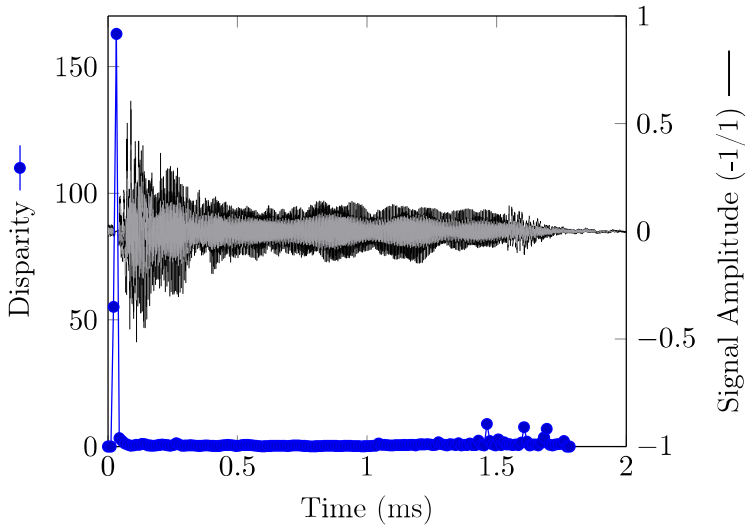


Figure 52: Spectral Disparity. This figure shows the spectral disparity for a C_4 note in a Upright Yamaha Piano. The Spectral Disparity peaks in the beginning of the attack phase.

touch. However, the spectral disparity does provide good estimations of the beginning of the attack phase.

5.6 Estimation of Keybed Bottom Distance

An important part of the precise key trigger system is to determine the maximum allowed displacement of the key, the so-called keybed bottom distance. This distance could be estimated by monitoring the force in the actuator slider for different key displacements. However, since the linear actuator does not offer a force sensor, this estimation can only be executed indirectly. Since the linear actuator controller supplies current to the actuator motor, this current can be used as an estimator for the actuator force. An experiment was performed in a Yamaha Upright Piano to evaluate the relation between current and key displacement. The results are shown in Figure 53.

5.7 Evaluation of first collision detection

As described in Section 4.11.2, the first hammer-string collision is detected by decreasing the attack time while comparing new obtained maximum RMS levels to average values captured previously. In this case, if a new value is larger than the average by a certain noise margin, then this value is considered to be the first collision. The goal of this section is to estimate reasonable values for the required margin between new values and average values of loudness. Also, the error of the first collision detection algorithm is estimated.

An experiment was performed to estimate these values. The experiment is executed by reducing the attack time until the first collision is detected by the algorithm. Besides, an observer notes the real attack that leads to the first hammer-string collision. The attack times suggested by the observer and by the algorithm are compared. The results are shown in Table 7.

5.8 Comparison of Regression Models

As described in Section 4.11 and Section 4.11.3, the calibration procedure acquires data consisting of pairs attack time - maximum RMS level, which are then fitted into a LLS model. The LLS model creates a mathematical function that can be used to calculate the optimal attack times required for achieving different dynamic levels. As

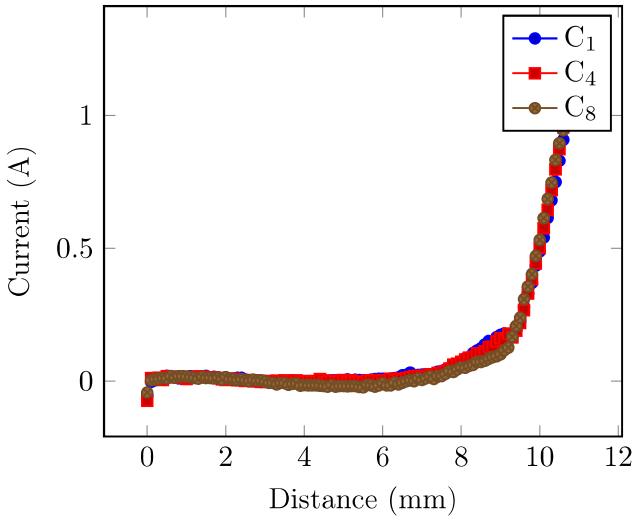


Figure 53: Current over key displacement. The current is measured for different key displacements using different piano keys. The actuator position ranges from 0 mm to 11mm in steps of 0.1 mm considering a settling time of 5 s. The experiment was performed using the keys C_1, C_4 and C_8 in a Yamaha Upright Piano.

Key	Noise Margin	t_a by algorithm (ms)	t_a by observer (ms)	Error
C_1	5%	400	370	8.1%
C_1	10%	380	370	1%
C_1	20%	340	370	8.1%
C_4	5%	360	330	8.3%
C_4	10%	340	330	3%
C_4	20%	300	330	10%
C_8	5%	280	240	16%
C_8	10%	250	240	4.1%
C_8	20%	220	240	9%

Table 7: Evaluation first collision algorithm. The experiment is executed by decreasing the attack time from 400 ms to 10 ms in steps of 10 ms. Long attack times do not lead to hammer-string collision and therefore only noise is perceived. While decreasing the the attack times, the maximum RMS level is compared to the average of the previous values. The first collision is detected when a new value exceeds the average of previous values by a certain noise margin. This experiment showed that a noise margin of 10% tends to obtain better estimates of first collision when compared with margins of 5% and 15%.

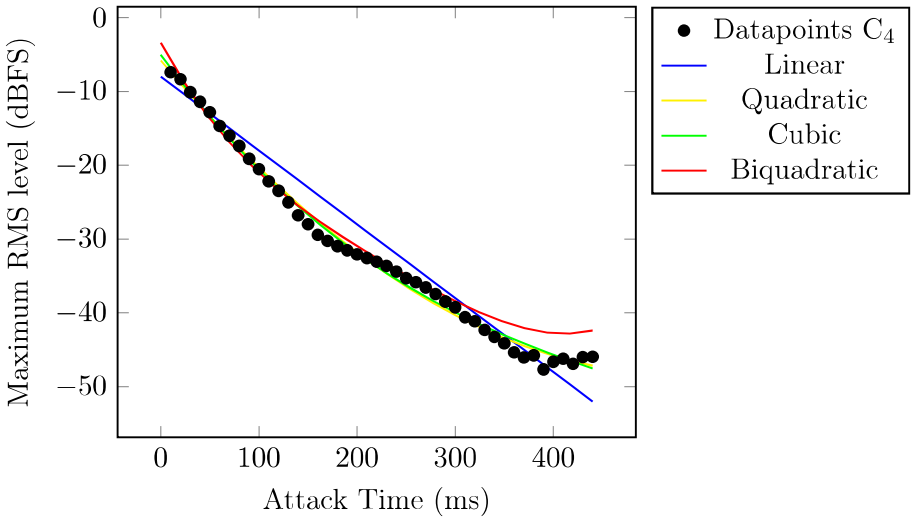


Figure 54: Comparison of Least Square Models. The Linear Model only consider data-points leading to hammer-string collision. The models with higher order consider data-points leading to hammer-string collision and not leading to hammer-string collision.

an example, if the user needs to record samples at 20 different dynamic levels, then 20 values for attack time are calculated with the mathematical function.

Different types of Least Squares estimators were compared to see how well they fit into the data acquired. The linear case only considers the data-points associated with hammer-string collisions while the models with higher consider all acquired data-points, i.e. data-points leading to hammer-string collision and also data-points not leading to the hammer-string collision.

5.9 Behavior of Black Keys

As described in Section 3.1, the keyboard is composed of black and white keys. The action mechanism is a lever connected to ground by a pivot. Due to the keyboard layout, black keys have shorter distances between the point where the force is applied (by the actuator) and the pivot. This difference in the distance could impact how the system triggers black keys and was investigated. An experiment was executed to investigate the loudness profile of black and white keys using the same ADSR curves. The results can be seen on Figure 55. A direct conclusion from this experiment is that long attack times show a slightly higher maximum RMS levels.

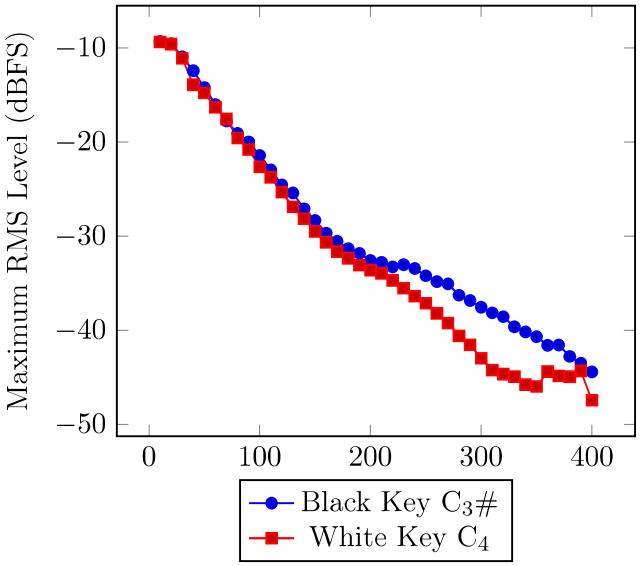


Figure 55: Black Key. The experiment was executed by ranging the attack time from 400ms to 10 ms in steps of 10 ms. The microphone was located at approximately 10 cm from the hammer-string contact point. The experiment was executed using ADSR curves considering $A_a = 9$ mm, $\mathrm{A_d} = 90\%\mathrm{A_a}$, $t_d = 50$ ms, $r_t = 100$ ms.

5.10 Estimation of Escapement Point

One important step while creating sample libraries is determining the beginning of samples, i.e. their start markers. Theoretically, the start of a sample should ideally match the moment when the hammer loses contact with the action mechanism, named escapement point. Since the piano key triggering system is designed to be non-invasive, i.e. no sensors are installed inside the action mechanism, the exact moment of the escapement point cannot be determined directly. However, it is possible to estimate the escapement point in terms of external features, such as the beginning of the ADSR curve. As shown in Section 3.8, the literature suggests that the escapement point matches the point of maximum hammer speed. In this system, the maximum speed takes place at 50% of the attack time due to the use of cosine-shaped curves and displayed in Figure 46. However, the results of this method can change drastically depending on several factors, such as pressed key and dynamic level.

An experiment was carried out in order to estimate the escapement point of the keys C_2, C_4 and C_6 in a Yamaha Upright Piano. The actuator was triggered using ADSR curves with attack times ranging from 200 ms to 10 ms in steps of 10 ms.

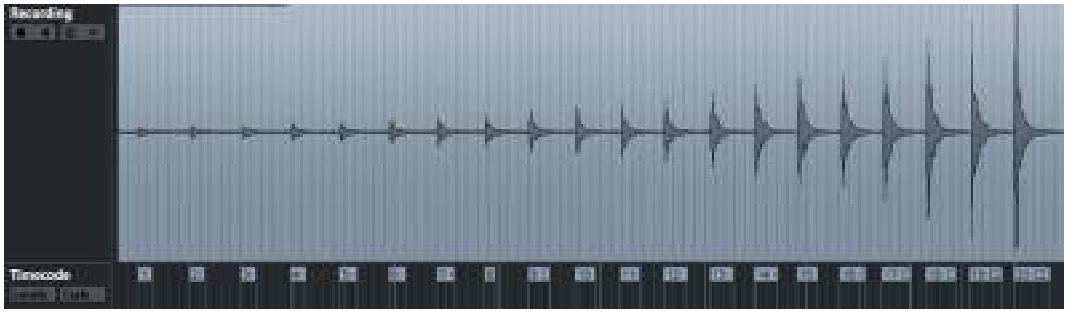


Figure 56: Markers Exported to Steinberg Nuendo. Start and end markers exported to Steinberg Nuendo.

The complete recording takes place at a DAW, in this case, Steinberg Nuendo. The control software saves start-markers at 50% of the attack time of the ADSR curves. These start-markers are then imported in Steinberg Nuendo and the samples are cropped. The results are displayed in Figure 56.

These start-markers used for sample cropping are compensated by a fixed microphone input delay and by the average delivery time of the UDP packages between the control software and the PID controller. The average delivery time of UDP packages can be estimated via the LinUDP protocol. The results of the experiment are shown in Figure 57. They show that the estimated delay values oscillate and have a certain random behavior. This happens due to the unpredictable delivery times of UDP packages, which are controlled by the operating system.

5.11 Evaluation of triggered notes

This section evaluates the obtained notes for the different cases: notes without resonance, notes with resonance, release sounds and keybed noises.

5.11.1 Notes without Resonances

An experiment was executed and a key was recorded with the sustain pedal up at different dynamic levels. Figure 58 shows the recorded signal of a C_2 for 20 different dynamic levels using the attack times calculated using the calibration algorithm.

5.11.2 Resonances

To evaluate the recording of sympathetic and sustain resonances using the piano key trigger system, an experiment was executed in a Grand Piano.

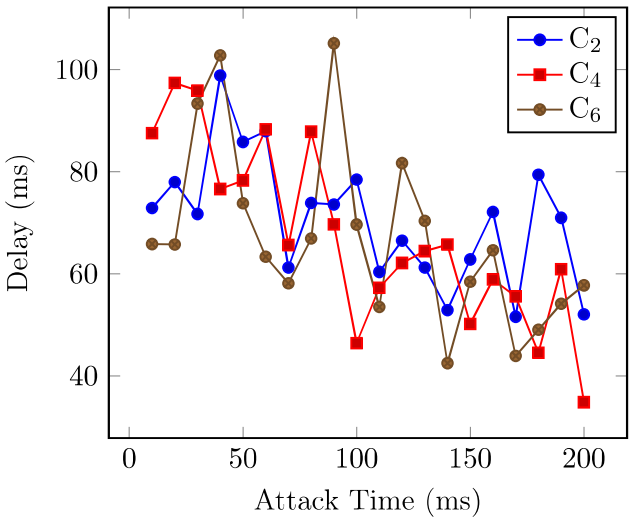


Figure 57: Escapement point - Hammer-string delay. Delay between the escapement point and the Hammer-string contact point. The escapement point is set to take place at 50% of the attack time.

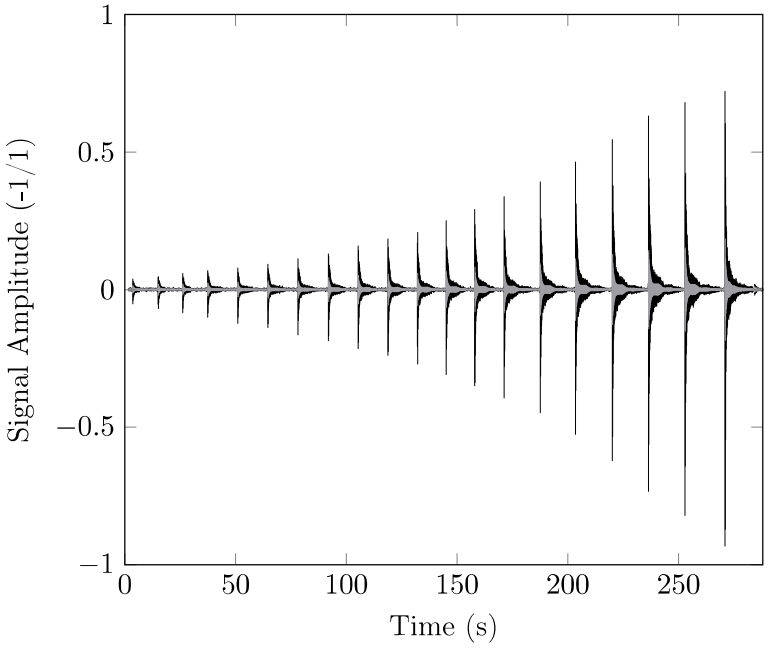


Figure 58: Recording of a key C_2. The key is recorded with the pedal up at 20 different dynamic levels with attack times calculated by the calibration algorithm.

For the sustain resonances case, an ADSR curve with attack time 20 ms triggered the key C_4 with the sustain pedal up. In the sequence, the same ADSR curve was used to trigger the key with the sustain pedal kept down. Finally, the phases of both sounds were manually matched and the samples were subtracted. The obtained sound does not have an abrupt increase in the signal amplitude in the beginning of the sound and represents to a good degree the resonance of the instrument.

Similarly, an experiment was executed to evaluate the sympathetic resonances. The system was used to trigger the key C_2 two times in a row, one with C_4 already kept down and a second time with C_4 not pressed. Next, the two obtained sounds were phase-matched and subtracted afterwards. Again, the obtained sound does not show an abrupt increase at the beginning of the sample and represents the to a good degree the sympathetic resonance of the key.

5.11.3 Release Sounds

To evaluate the release sounds, an experiment was performed. The recording of release sounds should include soft and hard release sounds. While executing the release sounds, the system should not add noises that destroy the final recording. The experiment consists of triggering ADSR curves at different release times ranging from 350 ms to 1 ms. It is important to notice that in this experiment the linear actuator is not installed in any piano key, i.e. no key is pressed. This allows one to evaluate the maximum RMS levels produced for different release times. The experiment results are displayed in Figure 59.

In the scope of the piano trigger, hard releases can be obtained by using short release times, although too short times (smaller than 50 ms) tend to produce high noises and spoil the recording. In the triggering system, hard releases are achieved with the value of 100 ms. Soft releases can be achieved by using any considerable long release time. Again, in the triggering system, the soft releases are achieved by using release times equal to 300 ms.

5.11.4 Keybed Noise

In order to evaluate the keybed noises, an experiment was executed. The strings of the key C_1 were damped with soft material and the maximum RMS level was evaluated for different attack times. The goal of damping the strings is to analyze solely the sounds produced by the key-keybed collision. Also, it is important to notice that it is not possible to completely damp the string sounds. The results of

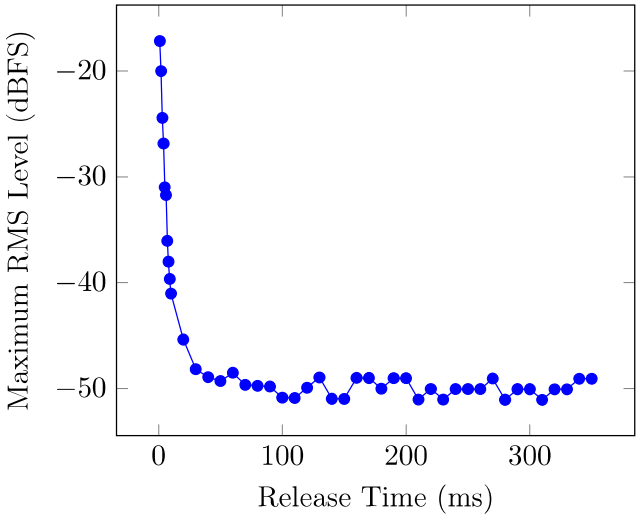


Figure 59: Release Sounds. The actuator slider is triggered without any piano key at different release times. Short release times (smaller than 20 ms) tend to produce loud noises and should be avoided.

the experiment are displayed in Figure 60.

The spectrogram of the keybed bottom collision is displayed in Figure 61. It has a duration of approximately 100 ms and it is relatively independent of the key involved and the attack time.

5.12 PID Tuning

PID tuning involves the adjustment of the control parameters in order to obtain a desired control response. Although many tuning methods are available, they normally involve the use of mathematical models or have high complexity. In this thesis, the PID controller was manually tuned.

Increasing the proportional gain K_p has the effect of decreasing the rise time, increasing the overshoot, decreasing the steady-state error and degrading the stability. Overshoots are unwanted in this system since they could lead to errors in the attack amplitude. The sound is produced over a short hammer-string collision and an overshoot error could considerably change the characteristics of the obtained recordings. Increasing the derivative gain K_d has a minor effect on the rise time, decreases the overshoot, decreases the settling time, has no effect on the steady-state error and improves stability for small K_d values. Finally, the integral gain K_i decreases the rise time, increases the overshoot, increases the settling time, eliminates

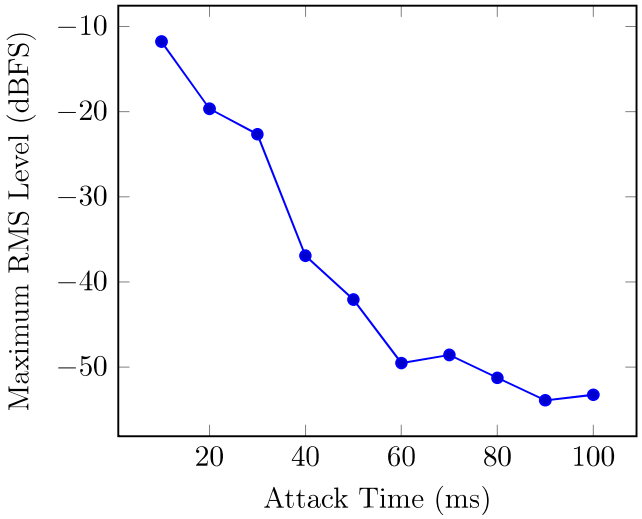


Figure 60: Keybed Noise. Maximum RMS level of keybed noise. The experiment was executed by ranging t_a from 100 ms to 10 ms in steps of 10 ms. The other parameters are $A_a = 9$ mm $t_s = 1000$ ms, $A_s = 8$ mm and $t_r = 100$ ms.

Parameter	Value
K_p	1,5 $\frac{A}{mm}$
K_d	3 $\frac{A.s}{m}$
K_i	0 $\frac{A}{s.mm}$

Table 8: PID Parameters. Values configured internally in the PID controller for the Piano Key Triggering.

the steady-state error and degrades the stability. K_i values different than zero make the position oscillate around the set-point. Again, this situation is undesirable and K_i is set to zero. That implies that steady-state errors occur, although they do not have an impact on the obtained results. As it was shown in Figure 48, the piano sounds do not have considerable changes in loudness in the range between 8 mm and 10 mm, so small steady-state errors in that range do not imply in loss of quality of obtained sounds.

Table 8 shows the values used for K_p, K_d and K_i. It is important to notice that the PID tuning considered the weights of the actuator slider (130 g) and the piano hammer attached to the slider (20 g).

To evaluate the tuning of the PID controller, an experiment was executed. The same ADSR curve was triggered two consecutive times, first triggering a piano key

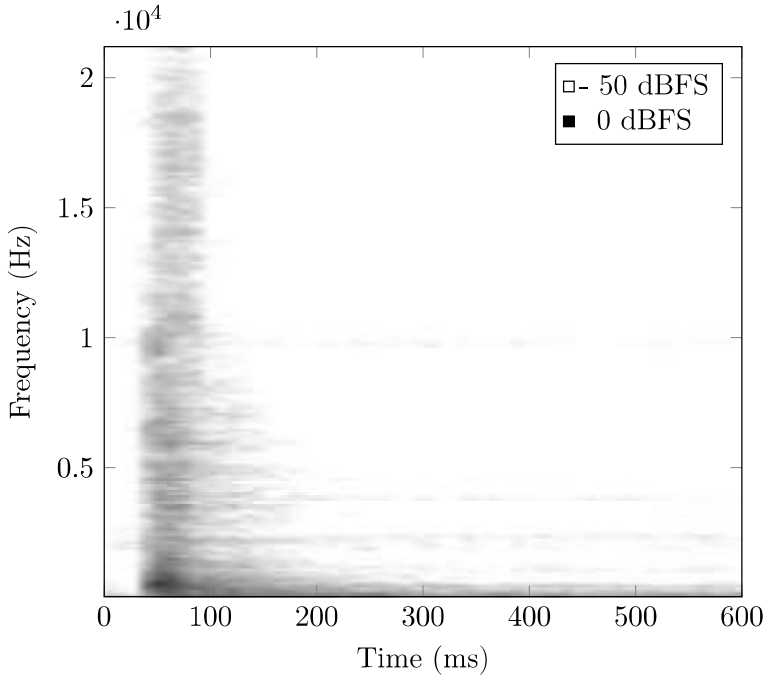


Figure 61: Spectrogram for the Keybed Noise. Spectrogram displaying the keybed noise for ADSR curve with $t_a = 10$ ms, $A_a = 9$ mm, $t_s = 1000$ ms, $A_s = 8$ mm and $t_r = 100$ ms. The FFT is calculated with window size of 1024 samples using hann window. The window sliding has an overlap of 50%. Sampling takes place at 44,1 kHz.

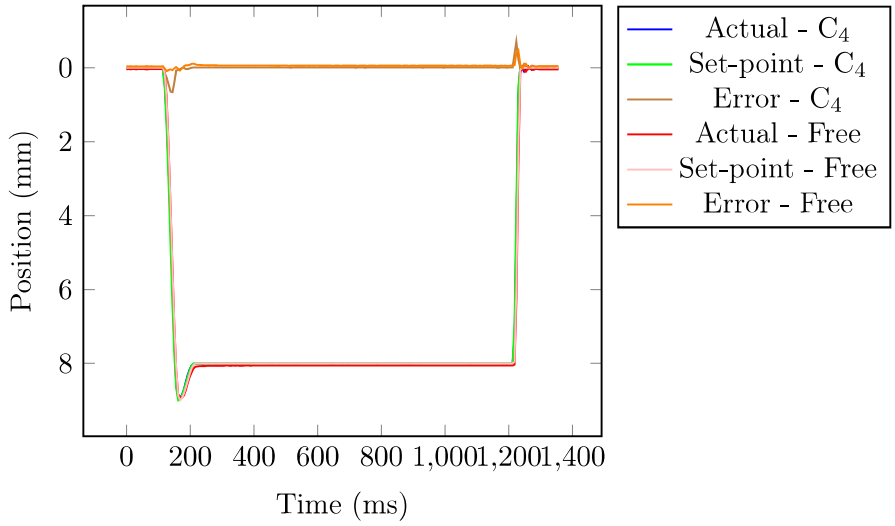


Figure 62: Actuator position with and without a piano key. A fixed ADSR curve with $t_a = 50$ ms, $t_d = 50$ ms, $t_s = 1000$ ms, $t_r = 20$ ms was used to trigger a piano key as well as no piano key, i.e. actuator moving freely.

and the after freely, i.e. away from the piano key. For each situation, the position set-point curves were compared with the actual curves read by the position sensor. Figure 62 shows the results of the experiment.

5.13 Synthesizers

The precise key triggering system is designed to operate with pianos. However, it can potentially work with other keyboard instruments, such as synthesizers and vintage pianos. An experiment was performed with a Chulek synthesizer, developed by the author of this thesis. The experiment setup is shown in Figure 63.

To understand this experiment, it is necessary to understand how velocity is interpreted in some types of digital synthesizers. Normally keyboards of synthesizers have a working principle similar to MIDI keyboards, i.e. when a key is pressed down, two non-latching on-off switches located alongside the axis of the key are sequentially closed. These switches are connected to a micro-controller using digital multiplexers and scanned within certain fixed time intervals. Finally, the time interval between closing the two switches is estimated by the micro controller's internal timers and converted to a MIDI velocity value by a mathematical curve implemented in the micro-controller, in this case, $\text{Velocity}_{\text{MIDI}} = -0.3706.\text{t}_{\text{interval}} + 130.7$. This velocity is sent via MIDI control change message to the synthesis engine. It is important to

Figure 63: Synthesizer Sampling. Experiment setup showing the precise key triggering system sampling a Chulek synthesizer. The goal of this experiment was to estimate the internal velocity curves implemented by the synthesizers keyboard.

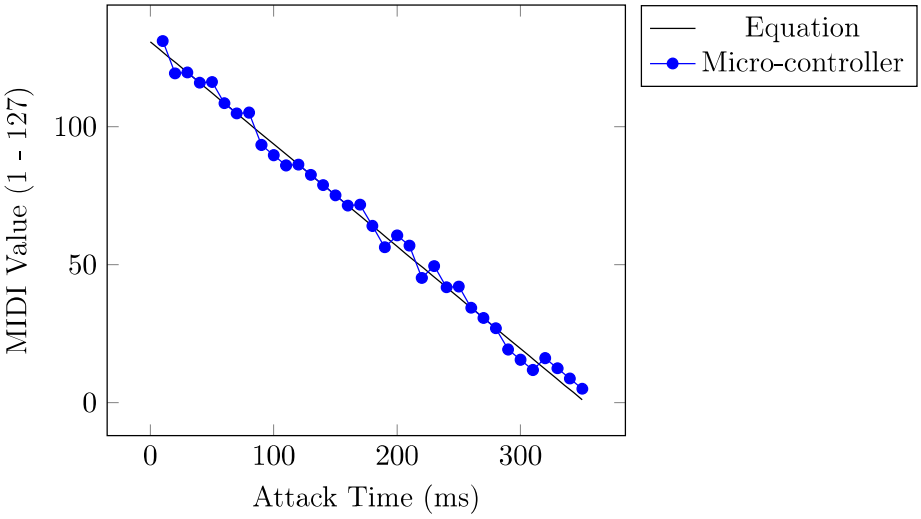


Figure 64: MIDI velocity vs attack time. The actuator slider presses the synthesizer key at different attack times. The time interval between closing the two switches is measured by a micro-controller operating at 8 MHz. Due to the larger amount of switches than available digital inputs, the switches are read via digital multiplexers. As a consequence, each key is scanned every 1 ms. The micro-controller implements the equation $\text{Velocity}_{\text{MIDI}} = -0.3706.\text{t}_{\text{interval}} + 130.7$ internally.

notice that analog synthesizers normally work based on a different working principle. Keyboards of analog synthesizers often use control voltages and are not covered in this experiment. In this experiment, the linear actuator triggers the synthesizer key at different attack times and the calculated MIDI velocities are compared to the theoretical values.

One conclusion of this experiment is that the linear actuator can be used to sample synthesizers that have the keyboard based on two on-off switches. Another conclusion is that the internal curves implemented by the keyboard's micro-controller can be estimated with the linear actuator.

6 Conclusions

Piano sample libraries aim to provide authenticity and genuineness in digital implementations of pianos. The creation of piano sample libraries is a time consuming and repetitive task that is prone to numerous types of errors. In this context, a precise key triggering system based on a linear actuator and a following state-of-the-art approach can reduce the overall time required in a recording session as well as improve the quality of the obtained samples. Such a system allows reproducing the different dynamics of a trained pianist's finger, as well as estimate start and end markers used for precise sample cropping. Also, such a system allows triggering the different types of sounds required for the creation of piano sample libraries, i.e. notes without sustain, notes with sustain, notes with sympathetic resonances and release sounds. Most of the advantages of this system lay in the high repeatability provided by the use of a linear actuator with precision in the micrometer scale. A piano sampling system with these features has not been described in the literature.

Literature research suggests that a model of the piano action mechanism could, for instance, use the key movement as input and the sound as output. However, modeling such a system would be a highly complex task and would not generalize well in different types of pianos. Invasive methods are commonly described in the literature but often require modifications of the piano's action mechanism due to the installation of sensors into the piano mechanics. These changes in piano mechanics are highly unwanted since they lead to changes in the characteristics of the sounds. Non-invasive methods, like the one described in this thesis, do not require the modification of the piano action mechanism and show a great commitment to the authenticity of sounds. Non-invasive methods tend to generalize well under different types of pianos. In this thesis, such a non-invasive system is achieved by analyzing the signal captured by a microphone and controlling a linear actuator positioned on the top of the key surface.

Since the system aims to achieve the different dynamic levels of a pianist's finger, the use of audio descriptors of loudness is essential. The maximum RMS level of a recorded signal provides a good estimation of the perceived loudness and is widely used in this system. When estimating the loudness using RMS levels, it is necessary

to know information about the key, such as the fundamental frequency of the note being played. When calculating the RMS level, a sampling window with a duration of two times the period of the fundamental component of the note provides satisfactory results of loudness. The RMS level has also the advantage that the calculations are performed in the time domain of the signal.

Spectral analysis requires proper spectrum estimation. In the scope of audio signals, the FFT algorithm associated with Hann or Hamming window provides a good spectral estimation. To ensure an equal spectral resolution for all keys, window size is not key-dependent and can be set at fixed value. The window size has to consider the worst-case scenario, i.e. the lowest fundamental frequency of the piano. Spectral audio descriptors, such as spectral disparity and spectral centroid can be useful in the assessment of different types of information. The spectral disparity provides an estimation of the beginning of sound and can be used as an indirect estimator for the hammer-string collision, although a time offset correction is required.

Different dynamic levels cause changes in the spectral components of the harmonics, i.e. the brightness of a sound. Instead of working loudness, one could potentially use the brightness of a sound to determine the different dynamic levels. In this context, the spectral centroid provides the center of mass of the spectrum and an estimation of the brightness of a sound. The spectral centroid behaves as a good estimator of the different dynamic levels only for short values of attack times. For long attack times, the spectral centroid suffers a considerable influence of noises and does not provide useful information about the different dynamic levels. The use of the spectral centroid as an estimator of the dynamic level requires association with stabilization methods, such as averaging of centroids.

The use of a PID controller is adequate for applications requiring accurate and precise position control of a linear actuator. Proper PID tuning by trial and error method ensures the proper execution of set-point curves with small position errors under different circumstances. As an example, the controller shows similar results with and without the piano, a sign of robustness. Modeling the actuator position over time by using ADSR curves turned out to be a convenient choice since it allows playing all types of sounds in the piano. The use of a coordinate system positioned on the top of the piano also turned to be an appropriated choice. Linear-shaped ADSR curves produce loud noise due to discontinuities in kinematic variables and can significantly impact the quality of the recorded samples. Cosine-shaped curves produce less noise and allow a smoother operation of the PID controller. The attack amplitude A_a of the ADSR curve can be estimated using the result of the keybed bottom distance

algorithm, which monitors the current flowing into the linear actuator. The attack times t_a of the ADSR curves proved to be directly proportional to the loudness of a sound and, as a consequence, one of the most important parameters of the system. The decay time t_d of the ADSR curve does not have a considerable impact on the loudness of sounds and can be set at a certain fixed value. Similarly, the release time t_r allows achieving different types of release sounds, although low release times (smaller than 20 ms) lead to high noises and should be avoided. The escapement point can be estimated from the maximum speed of the linear actuator, which can be derived from the ADSR curve. However, the estimation of the escapement point still suffers a major influence of UDP jittering.

Time constraints play a major role in a precise key triggering system for piano sample recordings. Sample cropping requires sample-based precision and as a consequence, proper synchronization between the linear actuator controller, the control software and the DAW. Synchronization using MIDI time-code messages sent from the DAW ensures that the suggested start and end markers required for sample cropping follow a single clock source. The sequential execution of position set-points directly from the control software does not meet the time constraints required in such a system. This result is mostly related to the unpredictable delivery times of UDP packages. Applications involving precise key triggering need to have a time resolution as low as 1 ms and the use of external embedded hardware ensures the proper execution of the ADSR curves.

The pianos show a linear relationship between the attack time of the ADSR curve and the maximum RMS level in the range of 8 mm to 10 mm. The calibration procedure collects pairs of values composed by attack time and maximum RMS level and, after that, creates a loudness profile of the key by fitting those data-pairs in a LLS model. Knowing a linear model of the key, the attack times required for triggering the system at different dynamic levels can be calculated. In this context, the algorithm for detection of the quietest sound turned to be essential since it filters unnecessary noise data before fitting the linear model. The values of attack time provided by the model allow achieving the different dynamics of a key.

6.1 Future Work

Several tests, improvements, and experiments have been left for future work mostly due to time constraints. Suggestions for future work include the study of better descriptors for the escapement point, better methods for muting the string sounds

while recording keybed noises, hardware synchronization using optical methods, treatment of electromagnetic noise created by the windings in the stator, among others.

The estimation of the escapement point provided in this work is reasonable, i.e. it always takes place immediately before the sound onset. However, it is still prone to random errors mostly due to the jigger created by the use of a UDP-based protocol. To avoid this, a possible future study could consider triggering the actuator curves via analog signals, such as voltage or current. The method could potentially provide a better estimation of the escapement point due to the low latency of analog signals.

Recording keybed noise sounds is an intrinsically hard task mostly due to the difficulty of damping the vibration of the strings. A possible future study could describe how to obtain good keybed noise sounds without changing the mechanics of the piano, such as by removing the hammer.

Another suggestion of future work includes the study of spectral slopes for the analysis of the changes in overtones as an alternative to the spectral centroid. This study could focus on analyzing angles formed between the spectral bins of the harmonics for different dynamic levels. Higher dynamic levels are expected to have smaller angles when compared with low dynamic levels.

Continuing, another work could also consider the use of string exciters for recording piano resonances. String exciters work by changing the direction of the magnetic field and allow exciting metal strings without any contact. Their use could lead to a direct method for obtaining resonances, i.e. without the need for subtracting sounds with and without resonances.

Another potential study could involve the analysis of formants for different dynamic levels. Formants represent the concentration of energy in a fixed frequency ranges. The formants could be used as an estimator of the different dynamic levels of a key as a replacement of the loudness.

Electromagnetic noise produced by the windings of the stator still takes place. Simple damping solutions do not isolate the noise well. A possible future work could explore alternative methods for attenuating the noise created by the linear actuator.

Furthermore, a similar approach used in this work could be used to trigger percussion instruments, such as drums, and pipe instruments, such as organs. Drums could potentially be triggered with linear actuators, although the characteristics of the collision are considerably different from the piano case. Pipe organ produces sounds by controlling airflow through pipes and could also be triggered via a linear actuator. In either case, it is important to analyze the physics behind the instrument as well as

the characteristics of the sound produced.

7 Appendix A

List of important functions used to interface with the embedded controller. The parameter and return types are shown.

```
Bool LMav_SetCurrentCommandMode(String, Int)
Bool LMav_ResetCurrentCommandMode(String)
Int LMcf_getCurveProgress(String)
Bool LMcf_LoadCurve(String, Int, Int, String, Byte, Byte, Int, Int, Int, Int, Int[])
Bool LMcf_isCurveLoading(String)
String LMcf_GetErrorTxt(String)
String TLMcf_GetErrorTxt(String)
Int LMcf_GetErrorCode(String)
Int LMcf_GetWarningCode(String)
String LMcf_GetWarningTxt(String)
Bool isMasterSlaveOperationEnabled(String)
Bool MasterSlaveHoming(String)
Bool setBit15(String, Bool)
Int LMcf_StartStopDefault(String, Int)
Int getROM_ByUPID(String, Int)
Int getRAM_ByUPID(String, Int)
Int getMinVal_ByUPID(String, Int)
Int getMaxVal_ByUPID(String, Int)
Int getDefault_ByUPID(String, Int)
Int SetRAM_ByUPID(String, Int, Int)
Int SetROM_ByUPID(String, Int, Int)
Int SetRAM_ROM_ByUPID(String, Int, Int)
Bool LMmt_GoToPosFromActPosAndActVel(String, Float, Float, Float, Float)
Bool LMmt_MoveAbs(String, Float, Float, Float, Float)
Bool LMmt_MoveRel(String, Float, Float, Float, Float)
Bool LMmt_Stop(String, Float)
Bool LMmt_WriteLivePar(String, Int, Int)
```

```
Bool LMmt_VAJIGoToPos(String, Float, Float, Float, Float, Float)
Bool LMmt_PStreamSlaveGeneratedTimestamp(String, Float)
Bool LMmt_StopStreaming(String)
Bool LMav_Mod16BitCTPar(String, Int, Int, Int)
Bool LMav_Mod32BitCTPar(String, Int, Int, Int)
Bool LMmt_StartCTCommand(String, Int)
Bool LMmt_ClearEventEvaluation(String)
Bool LMav_RunCurve(String, Int, Int, Int, Int)
Bool LMav_MoveBestehorn(String, Float, Float, Float, Float)
Bool LMav_MoveBestehornRelative(String, Float, Float, Float, Float)
Bool LMav_MoveSin(String, Float, Float, Float)
Bool LMav_MoveSinRelative(String, Float, Float, Float)
Bool LMfc_ChangeTargetForce(String, Float)
TimestampData getDemandPosWithTimestamp(String)
TimestampDataUTC getDemandPosWithTimestampUTC(String)
TimestampData getDemandCurrentWithTimestamp(String)
TimestampDataUTC getDemandCurrentWithTimestampUTC(String)
TimestampMonitoring getMonitoringChannelWithTimestamp(String, Int)
TimestampMonitoringUTC getMonitoringChannelWithTimestampUTC(String, Int)
Int get_PseudoScopeSamples()
Void set_PseudoScopeSamples(Int)
Bool enablePseudoScopeTrace(String)
Bool isPseudoScopeSampling()
Int getPseudoScopeProgress(String)
StateMachineStates getStateMachineState(String)
Bool isSwitchOnActive(String)
Bool isEventHandlerActive(String)
Bool isSpecialMotionActive(String)
Bool isInTargetPosition(String)
Bool isHomed(String)
Bool isFatalError(String)
Bool isMotionActive(String)
Bool isRangeIndicator1(String)
Bool isRangeIndicator2(String)
Bool isOperationEnable(String)
Bool isEnableOperation(String)
```

```
Bool isError(String)
Bool isSafeVoltageEnable(String)
Bool isQuickStop(String)
Bool isSwitchOnLocked(String)
Bool isWarning(String)
Bool isNotReadyToSwitchOnSM(String)
Bool isSwitchOnDisabledSM(String)
Bool isReadyToSwitchOnSM(String)
Bool isSetupErrorSM(String)
Bool isErrorSM(String)
Bool isHWTestsSM(String)
Bool isReadyToOperateSM(String)
Bool isOperationEnabledSM(String)
Bool isHomingSM(String)
Bool Active(String)
Bool SwitchOn(String)
Bool setSwitchOnBit(String, Bool)
Bool Homing(String)
Bool setHomingBit(String, Bool)
Bool AckErrors(String)
Bool SetErrorAcknowledgeBit(String, Bool)
Bool JogPlus(String)
Bool JogMinus(String)
Bool setJogPlus(String, Bool)
Bool setJogMinus(String, Bool)
Void Dispose()
Void deleteDebugLogFile()
Void setDebugMode(DebugModes)
Void addAppLogToDebug(String)
Void setDebugLog(String)
Double getMeanResponseTime(String)
Int getControlWord(String)
Int getStateVar(String)
Int LongRAWDataToInt(Int)
Void setHostMAC(String, String)
String getHostIP()
```

```
String getVersion()
Void CreateTargetAddressList()
Bool SetTargetAddressList(String, String)
String SetTargetAddressListByMAC(String, String)
Void ClearTargetAddressList()
Double getTimerCycle()
Bool setTimerCycle(Int)
Bool ActivateConnection(String, String)
Bool CloseConnection()
Bool isNetworkRunning()
String getDLLError()
Bool clearDLLErrors()
String getDriveType(String)
Bool isConnected(String)
Bool isResponseUpToDate(String)
Bool isRealtimeConfigUpToDate(String)
Int getDatagramCycleTime(String)
Double getActualPos(String)
Double getCurrent(String)
Double getDemandPos(String)
Int getMonitoringChannel1(String)
Int getMonitoringChannel2(String)
Int getMonitoringChannel3(String)
Int getMonitoringChannel4(String)
TimestampData getActualPosWithTimestamp(String)
TimestampDataUTC getActualPosWithTimestampUTC(String)
```

8 Acknowledgments

I would like to say thank you to (in alphabetical order):

- Clyde Sendke - Director of Markets Group at Steinberg
- Leonhard Reindl - Professor at University of Freiburg
- Matthias Klag - Instrument and Sound Designer at Steinberg
- Moritz Diehl - Professor at University of Freiburg
- Michael Ruf - Project Manager at Steinberg
- Philippe Bono - Release Engineer at Steinberg
- Ralf Kuerschner - Director of Engineering at Steinberg
- Sebastian Breiter - Instrument and Sound Designer at Steinberg
- Tilmann Mueller - Software Engineer at Steinberg

My special thanks to Matthias Klag and Sebastian Breiter for closely monitoring my activities.

Bibliography

[1] N. Giordano, *Physics of the Piano*. OUP Oxford, 2010.

[2] W. Goebl, R. Bresin, and A. Galembo, "The piano action as the performer's interface: Timing properties, dynamic behaviour and the performer's possibilities.," 07 2003.

[3] B. Bank, "Model-based digital pianos: From physics to sound synthesis," 2019. `https://ieeexplore.ieee.org/abstract/document/8588429`, accessed 2019-02-01.

[4] P. Newell, *Recording Studio Design*. Taylor & Francis, 2013.

[5] J. Gát, *The Technique of Piano Playing*. Collets, 1980.

[6] R. Kratzert, *Technik des Klavierspiels: ein Handbuch für Pianisten*. Bärenreiter, 2002.

[7] J.-C. Risset and S. V. Duyne, "Real-time performance interaction with a computer-controlled acoustic piano," *Computer Music Journal*, vol. 20, no. 1, pp. 62–75, 1996.

[8] Y. Li and L. Chuang, "Controller design for music playing robot — applied to the anthropomorphic piano robot," in *2013 IEEE 10th International Conference on Power Electronics and Drive Systems (PEDS)*, pp. 968–973, April 2013.

[9] J. Lin, H. Huang, Y. Li, J. Tai, and L. Liu, "Electronic piano playing robot," in *2010 International Symposium on Computer, Communication, Control and Automation (3CA)*, vol. 2, pp. 353–356, May 2010.

[10] D. Zhang, Jianhe Lei, Beizhi Li, D. Lau, and C. Cameron, "Design and analysis of a piano playing robot," in *2009 International Conference on Information and Automation*, pp. 757–761, June 2009.

[11] J. L. Anders Thorin, Xavier Boutillon, "Modelling the dynamics of the piano action: is apparent success real?," *Acta Acustica united with Acustica*, p. 10., 2014.

[12] J. Chabassier, A. Chaigne, and P. Joly, "Modeling and simulation of a grand piano," *The Journal of the Acoustical Society of America*, vol. 134, no. 1, pp. 648–665, 2013.

[13] A. Askenfelt, "From touch to string vibration. ii: The motion of the key and hammer," *J. Acoust. Soc. Am.*, vol. 90, pp. 2283–2293, 1991.

[14] W. Goebl, R. Bresin, and A. Galembo, "Touch and temporal behavior of grand piano actions," *The Journal of the Acoustical Society of America*, vol. 118, pp. 1154–65, 09 2005.

[15] W. Goebl, R. Bresin, and I. Fujinaga, "Perception of touch quality in piano tones," *The Journal of the Acoustical Society of America*, vol. 136, p. 2839, 11 2014.

[16] H. Suzuki, "Spectrum analysis and tone quality evaluation of piano sounds with hard and soft touches," *Acoustical science and technology*, vol. 28, no. 1, pp. 1–6, 2007.

[17] M. Flückiger, T. Großhauser, and G. Tröster, "Precision finger pressing force sensing in the pianist-piano interaction," in *Proceedings SMC 2016*, pp. 137 – 142, Zentrum für Mikrotonale Musik und Multimediale Komposition (ZM4), 2016. An optional note.

[18] S. F. Hiroshi Kinoshitaa, "Loudness control in pianists as exemplified in keystroke force measurements on different touches," *Acoustical Society of America*, vol. 121, no. 5, p. 2959–2969, 2007.

[19] H. Kinoshita, S. Furuya, T. Aoki, H. Nakahara, and E. Altenmüller, "Characteristics of keystroke force in the piano," *Journal of Biomechanics - J BIOMECH*, vol. 40, 12 2007.

[20] V. S. Library, "Vienna software," apr 2019. `https://www.vsl.co.at/en/Software/Video_Demos`, accessed 2019-02-01.

[21] Modartt, "Pianoteq," 2019. `https://www.pianoteq.com/`, accessed 2019-05-02.

[22] S. Robot, "sample Robot." `https://samplerobot.com/`, accessed 2019-07-03.

[23] Yamaha, "B series," 2019. https://de.yamaha.com/de/products/musical_instruments/pianos/disklavier/enspire_pro/index.html, accessed 2019-07-30.

[24] T. R. Neville H. Fletcher, *The Physics of Musical Instruments*. Springer, 1998.

[25] N. Giordano and J. P. Winans, "Piano hammers and their force compression characteristics: Does a power law make sense?," *The Journal of the Acoustical Society of America*, vol. 107, no. 4, pp. 2248–2255, 2000.

[26] N. Giordano, "Sound production by a vibrating piano soundboard: Experiment," *The Journal of the Acoustical Society of America*, vol. 104, no. 3, pp. 1648–1653, 1998.

[27] J. Davidson, "Average vs rms meters for measuring noise," *IRE Transactions on Audio*, vol. AU-9, pp. 108–111, July 1961.

[28] E. A. Dietrich Parlitz, Thomas Peschel, "Assessment of dynamic finger forces in pianists: Effects of training and expertise," *Journal of Biomechanics*, vol. 31, pp. 1063–1067, 1998.

[29] A. Askenfelt and E. V. Jansson, "From touch to string vibrations. i: Timing in the grand piano action," *The Journal of the Acoustical Society of America*, vol. 88, no. 1, pp. 52–63, 1990.

[30] P. Dijksterhuis, "De piano," *Nederlandse Akoest. Genootschap*, vol. 7, pp. 50–65, 1965.

[31] T. Cheng, S. Dixon, and M. Mauch, "Modelling the decay of piano sounds," *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 594–598, 2015.

[32] S. Papetti and C. Saitis, *Musical Haptics: Introduction*. Cham: Springer International Publishing, 2018.

[33] A. Askenfelt, "Observations on the transient components of the piano tone," in *Proceedings of the Stockholm Music Acoustics Conference*, pp. 297–301, ACM, Aug 1994.

[34] S. Liasi, "What are linear motors?," 05 2015.

[35] S. Dalla Bella and C. Palmer, "Rate effects on timing, key velocity, and finger kinematics in piano performance," *PloS one*, vol. 6, p. e20518, 06 2011.

[36] LinMot, "Hp motoren," 2019. `https://linmot.com/de/produkte/linearmotoren/`, accessed 2019-05-10.

[37] K. Astrom and T. Hagglund, "Pid controllers: theory, design and tuning," *Instrument Society of America*, 1995.

[38] Steinberg, "Ur-22 user manual." `http://download.steinberg.net/downloads_hardware/UR22/UR22_documentation/UR22_OperationManual_de.pdf`, accessed 2019-08-07.

[39] Naudio, "Naudio." `https://github.com/naudio/NAudio`, accessed 2019-02-01.

[40] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*. San Diego, CA, USA: California Technical Publishing, 1997.

[41] F. J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proceedings of the IEEE*, vol. 66, pp. 51–83, Jan 1978.

[42] J. M. Grey and J. W. Gordon, "Perceptual effects of spectral modifications on musical timbres," *The Journal of the Acoustical Society of America*, vol. 63, no. 5, pp. 1493–1500, 1978.

[43] M. Diehl, "Lecture notes in modelling and system identification," October 2016.

[44] C. Hansen, S. Snyder, X. Qui, L. Brooks, and D. Moreau, *Active Control of Noise and Vibration*. 01 2012.

[45] Yamaha, "B series," 2019. `https://de.yamaha.com`, accessed 2019-05-10.

Glossary

Bass Lower frequencies in the audible frequency range. In a piano they are created by the keys located on the left side of the keyboard. 20

Dynamics Musical term used to describe the perception of loudness. Different dynamic levels represent different loudness.. 23

Escapement point Moment when the hammer loses contact with the piano action mechanism. 3

Fundamental frequency The lowest harmonic component of a piano sound.. 9

Grand piano Piano with the hammers and strings in horizontal position. 2, 12

Hammer A padded wooden component in a piano action that strikes the string. 1

Keybed Term used to describe the bottom of the lever formed by the key. 3

Octave The distance in frequency between two notes whose fundamental frequencies differ by a factor of two.. 9

Piano action The mechanism that allows the piano the create sounds. The main elements are the key, hammer, string, damper, bridge, soundboard, among others. 7, 12

Pitch Physically can be interpreted as the fundamental frequency, although other definitions are available. 9

Sample libraries Group of files containing the sounds produced by instruments. 3

Spectrogram Representation of frequency spectrum over time. 92

Staccato Term used to describe that musical notes are disconnected, dis-joined, i.e. played with periods of silence between them. 30

Sustain resonance Resonance achieved when the hammer strikes the string while all dampers are kept away by pressing the sustain pedal. 3, 29

Sympathetic resonance Resonance achieved when the hammer strikes the string while other key is kept down by the pianist's finger. 3, 29

Timbre The timbre gives the unique perception of a note. Physically, it is produced by the combination of the fundamental and harmonic components of given key. 14

Treble Higher frequencies in the audible frequency range. In a piano they are produced with the keys located on the right side of the keyboard. 20

Upright piano Piano with the hammers and strings in vertical position. 2, 12, 14

Velocity A measure of how fast a key on a piano is pressed or how loud a sound is perceived. 3

Acronyms

ADSR Attack, Decay, Sustain and Release. vii, viii, xi, xii, 2–4, 25–27, 30, 42, 44, 47, 48, 54, 56–58, 63, 64, 67, 68, 70, 75–77, 79–83, 87–89, 91, 93–95, 100, 101

CD Compact Disk. 53

CSV Comma Separated Value. xiii, 69

DAW Digital Audio Workstation. 38, 69, 70, 89, 101

DFT Discrete Fourier Transform. 60

DLL Dynamic Link Library. 51, 56

FFT Fast Fourier Transform. 54, 60, 62, 63, 84, 94, 100

GUI Graphical user interface. 54

LLS Linear Least Squares. 3, 66–68, 82, 83, 85, 101

MIDI Musical Instrument Digital Interface. xiii, 1, 7, 8, 12, 37–40, 70, 95, 97, 101

PCM Pulse Code Modulation. 53, 76

PID Proportional-Integral-Derivative. iii, v, ix, xiii, 2–4, 41–43, 48–51, 54, 57, 58, 75–77, 89, 92, 93, 100

PLC Programmable Logic Controllers. 51

RAM Random-Access Memory. 47, 51, 58

RMS Root Mean Square. viii, xii, 23, 54, 59, 60, 63, 65–69, 77–88, 91–93, 99–101

SCADA Supervisory Control and Data Acquisition. 54

UDP User Datagram Protocol. 51, 89, 101, 102

UML Unified Modeling Language. 56

VST Virtual Studio Technology. 38

WAV Waveform Audio File. 53, 54, 60, 76