Acquisition and study of blowing pressure profiles in recorder playing

Francisco Garcia Diaz-Castroverde

MASTER THESIS UPF / 2010
Master in Sound and Music Computing

Master thesis supervisor:
Esteban Maestre
Department of Information and Communication Technologies
Universitat Pompeu Fabra, Barcelona

Abstract

The aim of this work is to acquire, study and model the blowing pressure profiles during the real performance of the recorder. Among all the possible performer gestures that take place in real performance, i.e. blowing pressure, fingering or embouchure, the blowing pressure has been selected as the most representative and measurable.

A direct acquisition system was developed with pressure sensors. Two prototypes were designed and the second one was improved by a recorder luthier in order to acquire more accurate and reliable measurements. After that, a professional recorder player performed a recording script that was designed in order to cover a representative subset of the different performance techniques. The recordings of the sound and the gesture shape the multimodal database.

The analysis of the blowing pressure profiles was done looking carefully at four analysis dimensions: pitch, articulation, dynamics and note duration. A set of features was defined and studied in order to characterize temporally the profiles. An initial temporal model was developed for certain features.

Keywords: performance, performer gestures, performer model, direct acquisition, sensor, wind instrument, recorder, multimodal database, sound synthesis, synthesis model

Acknowledgements

I would like to thank to all people who accompanied me during these two years of master study. First of all, I would like like to thank to my supervisor Esteban Maestre for his kindly guidance and indispensable support. Without his help this work would not have never been finished adequately. Special thanks to my Master colleague Leny Vinceslas for his patience and friendship.

There are three persons who have been essential in the development of this project. Joan Vives, a radio broadcaster and music history teacher who gave us all the necessary contacts to move forward the work. Josep Tubau, an experienced luthier who has been really involved in this project since we contacted him the first time. I would like to thank to him specially for the building of the final prototype used in the database construction. I think that his job is an art work. Finally, special thanks go to Joan Izquierdo, a professional recorder player and teacher who collaborated unselfishly in the recordings. His comments and advices have been very useful for the development of this work.

Special greetings go to all colleagues of the Master in Sound and Music Computing, some PhD students and the MTG researchers for a great time of sharing knowledge and interest in music technology. It has been a great time for me and I hope we can share a fruitful professional future in this sector.

Francisco Garcia Diaz-Castroverde

Barcelona

September, 2010

Contents

1 Introduction **1**

1.1 Conceptual framework 2

1.2 Main goals 3

1.3 Structure of this document 4

2 Background **7**

2.1 Sound production mechanisms in the recorder 7

2.1.1 Key elements in the sound production mechanism 9

2.1.2 Acoustic models and mixing region 12

2.2 Relevant instrumental gesture parameters 14

2.3 Performance techniques in the recorder 18

2.4 Previous studies in the acquisition of the blowing pressure 23

2.5 Applications of instrumental gesture acquisition 24

2.6 Specific goals of this work 25

3 Acquisition of blowing pressure **27**

3.1 Design approaches 27
3.1.1 Measurement from outside the instrument 28
3.1.2 Measurement inside the instrument 29
3.2 Prototype I: external acquisition 31
3.2.1 Design 31
3.2.2 Specifications 32
3.2.3 Problems 35
3.3 Prototype II: modified instrument 36
3.3.1 Design 36
3.3.2 Specifications 37

4 Database construction **41**
4.1 Recordings setup 41
4.1.1 Studio setup 42
4.1.2 LabView© acquisition program 43
4.2 Recording script 44
4.2.1 Repetitions 46
4.2.2 Scales 47
4.2.3 Musical pieces 48
4.2.4 Extra material 48
4.2.5 Statistical analysis of the database 50
4.3 Data pre-processing 51

4.3.1 Synchronization 52
4.3.2 Smoothing 53
4.3.3 Automatic segmentation 53

5 Data analysis 59

5.1 Feature description 59
5.2 Articulation shapes 61
5.2.1 Full legato 62
5.2.2 Legato 65
5.2.3 Soft staccato 67
5.2.4 Staccato 71
5.3 Fingering, dynamics and articulation 73
5.4 Attack and rise time 78
5.5 Attack slope 81
5.6 Steady state to note duration ratio 84

6 Conclusion 87

6.1 Summary of contributions 87
6.2 Future work 88

A Technical details of the Prototype II 91

B Scores of the musical pieces 97

References 100

List of Figures

2.1 Recorder family from Moeck© 8

2.2 Head section of a recorder flute. A: block, B: windway, C: lip labium 9

2.3 Mouth opening of the recorder. W: jet length, h: windway height, y_o: origin of the transverse position [21]. 11

2.4 Lips position and distance in a transverse flute [7]. 11

2.5 Flow visualization of the mouth of a experimental flue pipe in the first and second hydrodynamic modes [21]. 14

2.6 Energy distribution in the sound production mechanism 15

2.7 Classification of musical and instrumental gestures by Cadoz [3]. . . 16

2.8 Example of fingering chart for different recorders. 20

2.9 Behaviour of the blowing pressure in the transverse flute for different dynamics and fingerings [15]. 22

3.1 Measuring the blowing pressure from outside the instrument. 28

3.2 Measuring the blowing pressure inside the instrument: internal sensor 29

3.3 Measuring the blowing pressure inside the instrument: external sensor 30

3.4 Prototype I: complete system (left) and during performance (right). 32

3.5 Prototype I: detail of the pipe connection at the mouth opening of the recorder 33

3.6 Freescale pressure sensor MPVX4006GC6U. 34

3.7 Acquisition card Arduino Duemilanove©. 34

3.8 Details of the mechanical modification of the mouthpiece. Left: hole for sensing the blowing pressure. Right: connection for the pressure sensor. 37

3.9 Scheme of the different mechanical modifications of the mouthpiece. 38

3.10 Pressure sensor Honeywell© ACSX01DN series. 39

3.11 National Instruments© USB 6009 acquisition card. 39

4.1 General studio scheme for the recordings. 43

4.2 Detail of the Audio Technica© 350 condenser mic. 44

4.3 Example of the score used for the repetitions. Note A4. 47

4.4 Score of the first scale. 47

4.5 Score of the second scale. 48

4.6 Statistical distribution of the database for the four articulations. . . 50

4.7 Statistical distribution of the database for the three dynamics. . . . 51

4.8 Statistical distribution of the database for the five note durations. . 51

4.9 Raw and smoothed version of the blowing pressure. 54

4.10 Example of the automatic segmentation in a soft staccato recording. 55

4.11 Example of the automatic segmentation in a full legato recording. . 56

4.12 Example of the automatic segmentation in a legato recording. . . . 57

5.1 Features for a note in a staccato recording. 60

5.2 Features for a note in a legato recording. 61

5.3 Features for a note in a full legato recording. 62

5.4 Blowing pressure profiles for full legato and note duration of 0.666 seconds. 63

5.5 Blowing pressure profiles for full legato and note duration of 0.666 seconds: ZOOM. 64

5.6 Blowing pressure profiles for legato and note duration of 0.666 seconds. 66

5.7 Blowing pressure profiles for legato and note duration of 0.666 seconds: ZOOM. 68

5.8 Blowing pressure profiles for soft staccato and note duration of 0.666 seconds. 69

5.9 Blowing pressure profiles for soft staccato and note duration of 0.666 seconds: ZOOM. 70

5.10 Blowing pressure profiles for staccato and note duration of 0.666 seconds. 72

5.11 Blowing pressure profiles for staccato and note duration of 0.666 seconds: ZOOM. 73

5.12 Behavior of the blowing pressure with pitch and dynamics: Full legato. 74

5.13 Behavior of the blowing pressure with pitch and dynamics: Legato. 75

5.14 Behavior of the blowing pressure with pitch and dynamics: soft Staccato. 77

5.15 Behavior of the blowing pressure with pitch and dynamics: Staccato. 77

5.16 Rise time of the legatos for different pitches. 78

5.17 Attack time of the staccatos for different pitches. 79

5.18 Attack and rise times behavior with different note durations. 81

5.19 Attack slope values and curve fitting for different articulations: dots = average values, line = curve fitting. 82

5.20 Attack slope values and curve fitting for different dynamics: dots = average values, line = curve fitting. 83

5.21 Steady ratio behavior for three different articulations. 85

5.22 Steady ratio behavior for the different note durations. 86

A.1 General scheme of the whole contralto recorder. 92

A.2 Detail of the mechanical holes in the mouthpiece. 93

A.3 Dimensions of the mechanical modifications. 94

A.4 Blowing pressure: sensor connection. 95

A.5 Connection of the two pressure sensors. 95

A.6 Detail of the water tramp. 96

B.1 Score of the first piece. Piece 1 of the book. 97

B.2 Score of the second piece. Piece 8 of the book. 98

B.3 Score of the third piece. Piece 11 of the book. 98

B.4 Score of the fourth piece. Piece 12 of the book. 99

B.5 Score of the fifth piece. Piece 13 of the book. 99

List of Tables

5.1 Coefficients of the power relationship between blowing pressure and pitch: Full legato 74

5.2 Coefficients of the power relationship between blowing pressure and pitch: Legato. 75

5.3 Coefficients of the power relationship between blowing pressure and pitch: soft Staccato. 76

5.4 Coefficients of the power relationship between blowing pressure and pitch: Staccato. 76

Chapter 1

Introduction

The musical instrument sound synthesis follows nowadays two main trends: the sample-based models with spectral transformations and the physical models.

The sampling techniques offer a great sound quality, but they don't have much controllability. Therefore, although some sounds could be realistically synthesized, the transitions between one sound into another lack the prosodic rules of traditional instruments or speech [20], e.g. the excitation-continous instruments such as bowed strings or wind instruments. This issue requires more general sound transformations and they must be understood in terms of what we hear [20]. The best way to achieve this is to use spectral transformations [19].

The physical models are suitable for modeling families of instruments since they are based on the physical phenomena happening of each instrument. They have much controllability and expressivity because all the physical parameters could be tuned in order to change the synthesized sound. Their main problem is the lack of ability to fit with real observed data due to the high number of parameters involved, the fact that control parameters are not related to the produced sound in a trivial way and the radical non-linearities in the numerical schemes [19].

In order to improve the musical sound synthesis, it is necessary to pay attention

to how the instrument is played. The instrumental gestures acquire in this sense a significative importance because they provide better sound transformations that allow the synthesis of more realistic sounds.

1.1 Conceptual framework

With regard to the continuously excited instruments, such as the bowed strings or wind instruments, some researchers at the Music Technology Group (MTG) have developed a very interesting approach with the violin. This kind of instruments have been chosen because the physical models need a huge amount of control data, since the degree of control of these instruments is very high. Therefore, those models are not suitable for these instruments. However, in the case of impulsively excited instruments, such as hammered strings (e.g. piano) or percussive instruments, the physicals models provide a good solution, since the degree of control is much lower.

For the case of the continuously excited instruments, the different commercial approaches offer solutions with very low degree of controllability and expressive capabilities [17]. Therefore, the MTG researchers have developed a *control-based synthesis model* which takes into account the performer gestures in order to synthesize meaningful sounds [17, 13].

Furthermore, although much research effort has been done in the description of the acoustic properties of the instrument [12] and in the way it is played [16], there are no direct measurements that relate the musical expressivity to the performer gestures. Some researchers have tried to measure the influence of the vocal tract and the embouchure settings during the performance, relating these measurements to low-level parameters such as frequency and loudness [14]. But they have not established the relationship to musical expression. One good work that could be a good starting point is the one by Fletcher [10]. He studies the acoustical properties

of the transverse flute and he relates them to the musical performance techniques such as vibrato or dynamics. An actual continuation of this work [7, 15] also sheds some light to the relevant parameters of the transverse flute performance from a control point of view. All these works are the state of the art in the measurement of wind instruments, concretely in flute-like instruments.

1.2 Main goals

The main goal of this work is to improve the musical sound synthesis for wind instruments. In order to achieve it, the acoustic characteristics of different wind instruments will be studied in order to decide which instrument would be the most suitable for studying the performer gestures.

As it was commented in the previous section, the transverse flute has been studied in depth from a control point of view. However, the recorder, that appears as a simpler wind instrument in which the control is relatively easy, has not received enough attention. It has fewer control parameters than woodwind or brass instruments, but it keeps a great expressivity . As there is no reed in the mouthpiece of the instrument, the identification and measurement of the most relevant parameters for musical expression will be easier than in the case of more complex instruments. Therefore, the recorder has been selected as the instrument of study for this work.

The first goal to achieve is to decide which are the most important control parameters in the recorder and how they can be measured through a non intrusive way. These measurements should acquire all the controllability available to the performer, that is to say, they should represent all the parameters that a performer can control in order to produce all the nuances of the instrument. Therefore, the acoustics of the instruments should be understood in depth and different sensors and possible measurements will be tested, e.g. the blowing pressure in different

points of the instrument, the air flow, the fingering of the performer or the influence of the vocal tract.

Once the control parameters are well defined and the measurement points are established, an acquisition system must be built which is able to record in a non-intrusive way the most meaningful parameters and the produced sound, so that a multimodal database can be built with the sounds and the performer gestures. For this purpose, a professional musician should play different scores with different dynamics, articulations, etc. The sound and gestures should be synchronously recorded. After completing the database, a set of features should be defined in order to analyze the curves and find a parametric model which fits them.

From a bigger perspective, all this research has a more general goal that is the creation of a wind instrument synthesizer. Due to time constrictions, the whole synthesizer cannot be developed. Therefore, as a proof of concept, the measurements of the performer gestures could be used to fed a physical model of the instrument in order to see the quality of the synthesized sound. This sound could be compared to the one generated by the synthetic gestures created by the parametric model. In this way, the original gestures and the synthetic ones could be compared with the sound they generate in the same physical model. This comparison would serve as an evaluation of the performer model. This last goal is considered as an extra goal if there is time, since its implementation is not trivial.

1.3 Structure of this document

After this brief introduction, this document has five more chapters that are organized as follows:

Chapter 2 deals with the previous studies and related work in this field. It would allow us to understand how the recorder works acoustically and which are the main

control parameters. At the end of this chapter, we redefine the goals of the project in order to have a coherence with the work done by other researchers.

Chapter 3 deals with the design and implementation of the acquisition system. Three different approaches are explained and two prototypes are described in detail.

Chapter 4 covers the construction of the database: from the scripts to the final set-up. It also deals with the data pre-processing that we did on the raw data in order to prepare them to be analyzed in a subsequent stage of the project.

Chapter 5 deals with the analysis of the data. It describes the features used to model the gesture curves and it explains step by step the analysis developed with the data.

Finally, **chapter 6** summarizes all the contributions of this work and sheds some light on the future work that could be done in order to develop completely the whole wind synthesizer.

Chapter 2

Background

In this chapter previous studies regarding the acoustics of the recorder will be reviewed in order to identify the key elements that produce the sound and the main existing models that explain the physical phenomena happening in this instrument. Then, the main gestures that a performer can control will be analyzed and put in its research context. After that, the focus will be on the previous measurements of the blowing pressure in flute-like instruments. Afterwards, the main performance techniques and some of the applications that an acquisition system of the performer gestures could provide are commented. Finally, the specific goals of this project are reviewed in comparison to the state of the art in this field.

2.1 Sound production mechanisms in the recorder

The recorder belongs to the family of the aerophones. This kind of musical instruments produce sound primarily by causing a body of air to vibrate, without the use of membranes nor strings, and without the vibration of the instrument itself adding considerably to the sound i.e. the main reason for the sound production is the vibrating air column. [22].

Figure 2.1: Recorder family from Moeck©

There exist many types of recorders with different shapes and sizes. Figure 2.1 shows a small subsert of the recorder family: from left to the right, one can find the sopranino, soprano, alto, tenor, bass and great bass. All these recorders are baroque-type ones, although there are other kind of manufacturing styles. In this work we will focus on the soprano and the alto because they are very commonly used and they are considered as the standard recorders.

Normally, the recorders are made up of three separable sections: the head, the middle and the foot piece. The head is responsible for the sound production, as it will be explained later. The middle and the foot piece are the resonant cavity of the instrument, i.e. the body of the instrument. By opening and closing the holes in these parts, the player can change the pitch of the instrument, thus changing the effective length of the resonator and its corresponding resonant frequency.

In order to produce the sound, the player blows through the windway of the instrument (Figure 2.2,B) [12]. The resulting air jet travels then through the mouth opening and it strikes the lip labium (Figure 2.2, C). The interaction of the jet

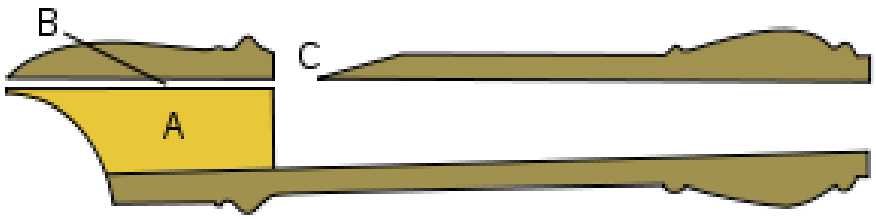

Figure 2.2: Head section of a recorder flute. A: block, B: windway, C: lip labium

with the standing air and with the lip labium creates a von Karman vortex street which makes the standing air inside the pipe to vibrate at its corresponding resonant frequency [12]. Finally, this vibrating column of air is the one that radiates the sound out of the instrument. During this process, there are many complex interactions which have been studied in depth from an acoustic perspective [14, 21], but we will only focus on the ones which are relevant to justify which are the main instrumental gesture parameters in the recorder and how they can be measured.

2.1.1 Key elements in the sound production mechanism

There has been a great amount of research in the acoustics of flute-like instruments, e.g. the organ pipe or the transverse flute [4, 10, 7]. Therefore, many of the approaches for studying the acoustic properties describing the sound driving mechanisms in the recorder has been adapted from the transverse flute or the organ pipe.

In the recorder the sound is produced by the interaction of an air jet with a sharp edge (called labium, Figure 2.2,C) placed in the opening of a resonator (body of the instrument). The jet is formed by flow separation at the end of a slit (flue exit) and travels along the mouth of the resonator towards the labium. At the flue exit, the jet is submitted to the transverse acoustic flow due to the oscillation in the pipe of the instrument. Because of its intrinsic instability, the jet is very sensitive to

these perturbations. They propagate on the jet and are amplified as they travel towards the labium. This results in a flipping of the jet on each side of the labium at the same frequency as that of the acoustic field. This motion of the jet around the labium provides energy to the acoustic field which enables its maintenance [21]. This energy can accumulate in standing waves corresponding to the resonances of the pipe of the instrument which favor oscillations at specific frequencies, i.e. the coupling between jet oscillations in the mouth and the acoustical resonances in the body provides self-sustained oscillations of the jet and maintains sound production [8].

In the case of the transverse flute, there are three main parameters which a player can alter to modify the timbre of the instrument [8, 7]: the jet length W, which is the distance from the flue exit to the lip labium; the windway height h, which is the height of the flue channel; and the origin of the transverse position of the labium y_o, which is normally in the middle of the flue exit. These parameters are not the only ones which can control the timbre of the instrument, but they are the most important and the most accessible.

In the case of the recorder, all these parameters are manually fixed by the manufacturers (figure 2.3). The ratio of the jet length W to the jet thickness h,

$$\frac{W}{h} \tag{2.1}$$

is normally fixed and equals approximately 4. The origin of the transverse position of the labium is also fixed in the manufacturing process. Therefore, one can claim that the expressivity of the recorder is more limited than that of the transverse flute. Furthermore, the recorder has a relatively poor acoustical response for higher harmonics compared to the transverse flute [8].

In order to understand the meaning of these parameters in the recorder, we can observe what they mean in a transverse flute. The jet length W refers actually to

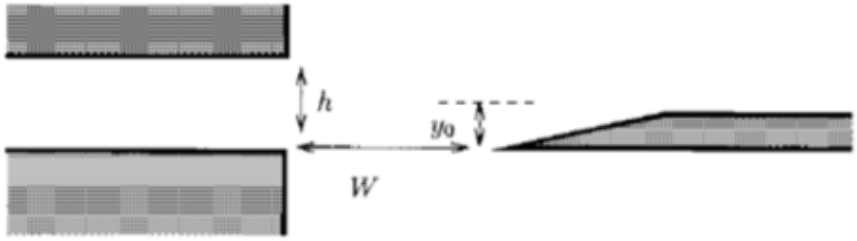

Figure 2.3: Mouth opening of the recorder. W: jet length, h: windway height, y_o: origin of the transverse position [21].

the distance between the lips and the labium, as it is shown in figure 2.4 through the help of a mirror, i.e. a mirror was placed next to the mouth in order to be able to measure the jet length W through computer vision techniques. The windway height refers to the lips opening h, which is also shown in figure 2.4. The transverse position of the labium would correspond to the relative position of the lips to the labium, but it is not shown in the figure. In the case of the transverse flute, the width of the lips opening H and the total opening surface Sm are also really important for the final timbre of the instrument, but it is not the case in the recorder.

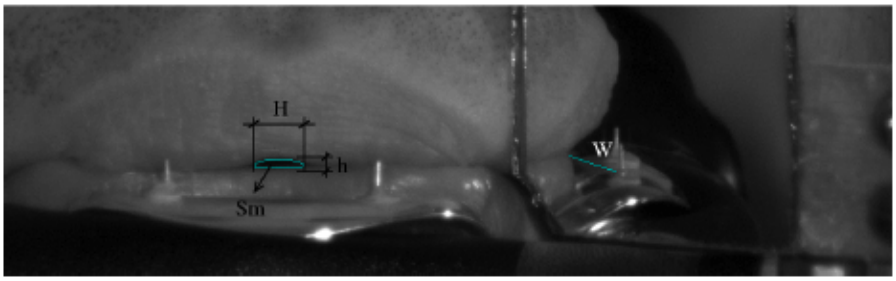

Figure 2.4: Lips position and distance in a transverse flute [7].

2.1.2 Acoustic models and mixing region

There are two accepted acoustic models that try to explain the complex phenomena happening in flute-like instruments. As it was mentioned before, at first they were developed by attending the organ flue pipe and the flute, but their results help to demonstrate the development of understanding of the sound production mechanisms common to the whole flute family. These two models are the **jet-drive model** and the **discrete vortex model**.

The jet-drive model was originally developed by Coltman in 1968 and it was later modified by Fletcher in 1975. It is a simplified model that explains the interaction of the air jet with the pipe. It assumes that the jet flow coming from the windway is separated at the labium into a flow entering the resonator and a flow leaving the resonator [11]. It also includes a momentum drive associated to the volume injection, although it was later demonstrated that this effect is weak and it can be neglected [5]. With this model, the sound driving mechanism of the recorder is usually represented by a feedback loop where the excitator is nonlinearly coupled to the pipe of the instrument which is represented as a linear acoustic filter. The steady-state amplitude of the wave coming from the jet is reached by saturation of the jet flow at both sides of the labium when the jet oscillations become greater than the half the width of the jet. The steady-state amplitude of the internal acoustic field of the resonator is later determined by a balance between the energy supplied by the jet oscillations and the loss mechanisms of the system, i.e. the sound radiation at the mouth and at the passive extremity of the pipe and also the thermal and viscous losses with the walls of the resonator.

Although this model has been very helpful for understanding the basic sound production mechanisms, it neglects many hydrodynamical details of the flow at the labium which appear to be fundamental in the performance of the instrument [21].

The jet-drive model does not allow to correctly predict basic features of the timbre such as the steady-state amplitude or the harmonic development. Therefore, the importance of the vortex shedding at the edge of the labium has to be taken into account in order to be able to determine the loudness of the instrument. The shedding of vortices at the edge of the labium constitutes the dominating nonlinear saturating mechanism for the fundamental of the acoustic oscillations of the instrument. While the separation of the flow at the labium mostly represents a damping mechanism for the fundamental frequency, it can also represent a source mechanism for the second and higher harmonics [21]. Furthermore, it also contributes to determine the timbre of the instrument and to trigger the attack transients.

For all these reasons, a more formal hydrodynamical model was developed by Verge in 1997: the discrete vortex model. Figure 2.5 shows the flow visualization of the mouth of a experimental flue pipe which was used to develop this model. We will not go through all the details because it goes out of the scope of this work, but we will highlight some important aspects which are very relevant for our purposes.

The discrete vortex model describes basically the timbre of the instrument depending on the dimensionless velocity of the air jet and the mouth geometry. Furthermore, it throws a very important conclusion about the energy transformations which happen during the sound production process. From the pneumatic energy coming from the air jet developed by the player, 95% is dissipated in the mixing region (a concept introduced by Elder in 1973 [9] which refers to the coupling zone at the exit of the windway and the lip labium), i.e. the shedding of vortices at the edge of the labium causes 95% of the energy to dissipate. From the remaining 5% that is transferred to the acoustic oscillation of the air in the pipe, around 3 or 4% is dissipated in viscous and thermal losses to the pipe walls, so that only about 1% of the initial pneumatic energy is radiated as sound, as it is shown in figure 2.6.

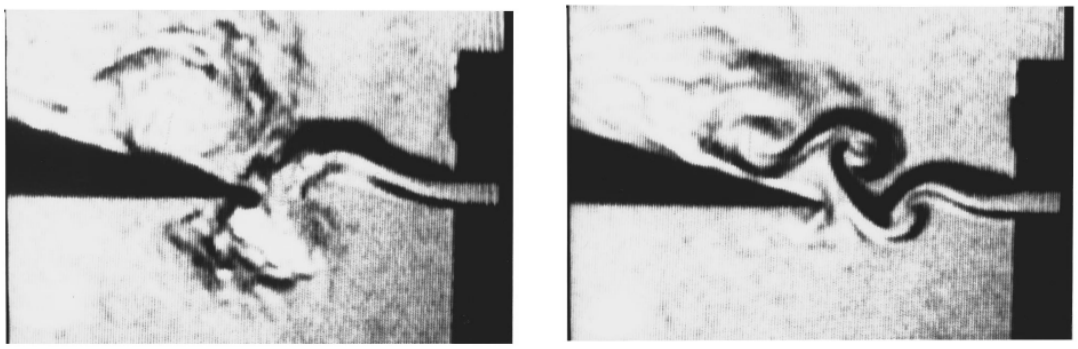

Figure 2.5: Flow visualization of the mouth of a experimental flue pipe in the first and second hydrodynamic modes [21].

For the aim of this work, this energy distribution has a huge impact. Since we are interested in the direct acquisition of instrumental gestures of the performer, we have realized that the unique places where we could establish reliable measurements would be in the mouth of the player or in the head piece of the instrument because after that point, most of the energy has been dissipated and the measurements would be much less representative of the performer controls. Therefore, we conclude that in order to achieve reliable and accurate measurements of the performer gestures we need to make them before the complex and nonlinear phenomena that happen in the lip labium, i.e. we have to measure in the head piece of the instrument.

2.2 Relevant instrumental gesture parameters

The analysis of instrumental gestures in music performance is a very active research field which has many practical applications. Among the musical gestures, one can find composition gestures and performance gestures. Composition gestures, while not physical, are already at the music creation step, constituting the composition itself. They can be understood as isolated notes, groups of notes, or annotations (e.g. crescendo, legato) referring to the manner in which some notes are to be

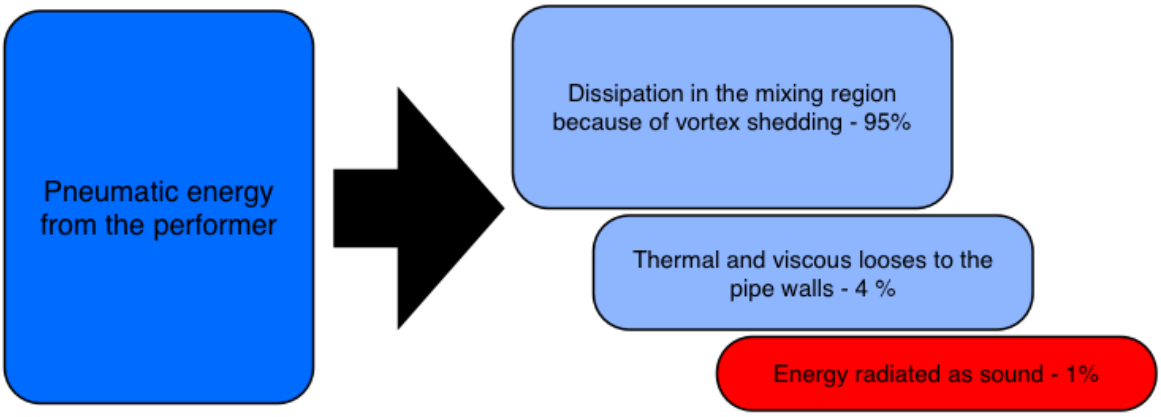

Figure 2.6: Energy distribution in the sound production mechanism

played. They convey a particular musical message and are expressed explicitly in the score. Conversely, performance gestures may not be explicit or quantified, and lay on the physical domain. They are understood as the voluntary (or constrained) gestures produced by the performer during the transformation of the score into musical sound [13].

Within performance gestures, Wanderley & Depalle [23] propose a classification into ancillary gestures and instrumental gestures. They refer to instrumental gestures as those involved in the sound production process (e.g., blowing with a certain pressure in a wind instrument), while ancillary gestures are considered to be produced by the performer as additional body movements, not involved directly in the sound production mechanisms, but linked to the performance and being able to communicate some emotional content, or even to slightly modify the sound properties (e.g., body movements of a clarinet player). Following the classification done by Cadoz in 1988 [3], which is shown in figure 2.7, a instrumental gesture typology is proposed according to their function:

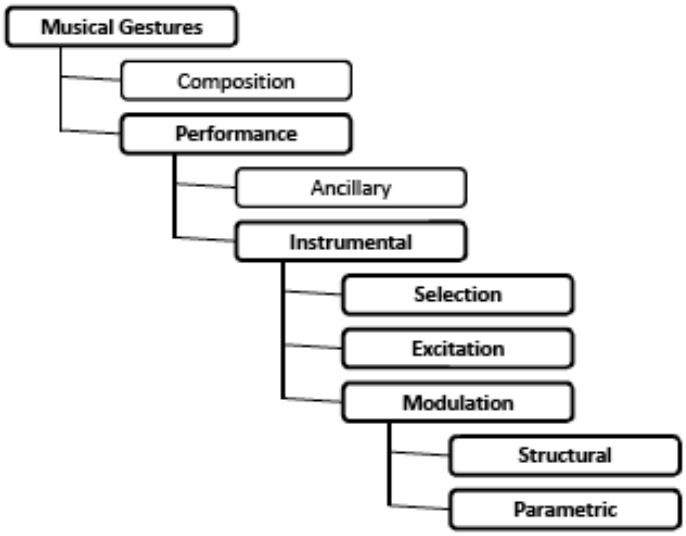

Figure 2.7: Classification of musical and instrumental gestures by Cadoz [3].

- Excitation gestures: they convey the energy that will be found in the sonic result (e.g. blowing in wind instruments).
- Modulation gestures: used to modify the properties of the instrument, but in which energy does not participate directly in the sonic result. These can be divided in parametric modulation gestures (e.g. change in the air flow in a wind instrument) and in structural modulation gestures (e.g. change the fingering for a certain note).
- Selection gestures: used to perform a choice among different but equivalent structures to be used during performance. There is neither energy transfer nor an object modification.

We are interested in studying the most important instrumental gestures that define the control a performer has on the recorder. As we will see later, all the performing techniques can be achieved through a combination of the modulation of

the blowing pressure, a fingering selection and the shape of the mouth, also called "embouchure". Although there are other physical actions that could be considered as performer controls, such as the air volume coming from the lunges or the shape of the vocal tract, we believe that the latter are the cause that make the blowing pressure to change, that is to say, the blowing pressure includes all these physical actions. Therefore, we could summarize the gestures of the recorder into three main control parameters, which the performer holds in order to produce all the nuances of the sound. These three parameters are the following:

- **Blowing pressure:** it is the most obvious control parameter in any wind instrument. It controls the dynamics and also the pitch of the recorder, as it was commented in the previous section. The blowing pressure is correlated with the pitch through an exponential relationship, see Equation 5.1 [14]. Its correlation with the dynamics is proportional[15], although it has not been quantified in an empirical way, since the dynamics can only be considered in a context. It is the main reason for the sound production of the instrument because once the air coming from the lunges is impulsed, the shape of the mouth and the oral cavitiy together with the position of the tongue cause a determined blowing pressure for that given air flow. In the case of flute-like instruments, it has been measured directly several times in different instruments [15, 10], as we will see in the next section.

- **Fingering:** this control allows the performer to change the pitch of the instrument, but it also permits him to slightly change the dynamics or even to perform a vibrato to a certain note. The interaction of the fingers with the instrument is really complex [1] and difficult to measure and it is out of the scope of this work due to time constraints.

- **Embouchure:** the influence of the configuration of the mouth and the vocal

tract is a very controversial issue in the case of the recorder. On the one hand, for Coltman (1968) [4] and Fletcher (1975) [10], the player's vocal tract is usually weakly coupled to the redorder and has only a limited influence on the sound. On the other hand, for Montgermont (2007) [15] and De la Cuadra (2005) [7], the player only acts on his mouth pressure and on the inner geometry of his vocal tract. Martin (1994) [14] analyzed the problem more in depth and he came to the conclusion that since the windway dimensions are fixed, the mouth windway resonant frequency depends only on the mouth volume. Therefore, he claims that the influence of the embouchure depends highly on the size of the recorder, but it is more apparent to the player than to his audience. For the purpose of this work, the measurement of this control parameter is really difficult, since it is only measurable through computer vision techniques or electromiography.

From all relevant control parameters which are available in the recorder, we will focus on the blowing pressure. We assume that this parameter is the effect of other actions or controls that modulate the sound, such as the tonguing and the teeth position. The fact that we restrict ourselves to measure only the blowing pressure would mean that we cannot cover all the possible performance techniques of the recorder. Therefore, we will limit our measurements to a representative subset which is coherent with our measurements.

2.3 Performance techniques in the recorder

The performance techniques in the recorder is a topic that has been studied by many researchers. There are many documents that review all kind of possible playing methods with this instrument, but we will focus on the traditional techniques besides the modern ones such as multiphonics or overblowing. Moreover, we are more

interested in the physical parameters a performer can control in order to achieve them than in the study of the techniques themselves.

The instrument range of the recorder is approximately about 2 octaves depending on the type of recorder we are talking about, e.g. treble, alto, bass, etc. A third octave can be also achieved through modern techniques.

The player is able to control how a note begins in order to create different attack transients which are very important to the overall perception of the subsequent sound [14]. The articulation range of the recorder covers from the briefest staccato to the broadest legato. The performer achieves the different articulations through a nearly equivalent syllable, such as “ta” or “da”. Here the consonant describes the required tongue action, while the vowel indicates the continuation of breath required to maintain the note. These syllables should be treated as hints for the performer, but not as prescriptions. For the staccatos, the typical syllables are “tu” and “du” as well as the sequences “tu-ku-tu-ku” or “du-gu-du-gu”. These syllables imply that the attack of the sound is more sharp. For the legatos, the syllables “du” (very soft) and ‘lu” are most used, as well as the sequences “lu-ru-lu-ru” or “du-ru-du-ru” [Joan Vives, private communication]. In this case, theses syllables imply a softer attack for each note. In order to achieve the different articulations, the performer has to change the shape of his mouth, the place where the tongue is positioned and finally the amount of air coming from the lunges. We believe that the combination of all these actions is showed in the shape of the blowing pressure curve.

The combination of the shape of the player’s mouth with the shape of the vocal tract is another performance technique. It is achieved through the manipulation of the size and shape of the oral and pharyngeal cavities [14]. However, its influence on the final sound of the recorder is a controversial topic.

The fingering is a relevant technique in the recorder. Its main purpose is to achieve different pitches, but it can also be used as a way of modulating dynamics on the sound. All notes can be played by means of several possible fingerings with slightly different timbres and dynamics, as it is shown in the fingering charts (figure 2.8). Most of the notes of the second octave and above are produced by partially closing the thumbhole on the back of the recorder, a technique known as "pinching". The placement of the thumb is crucial to the intonation and stability of these notes and it varies as the notes increase in pitch.

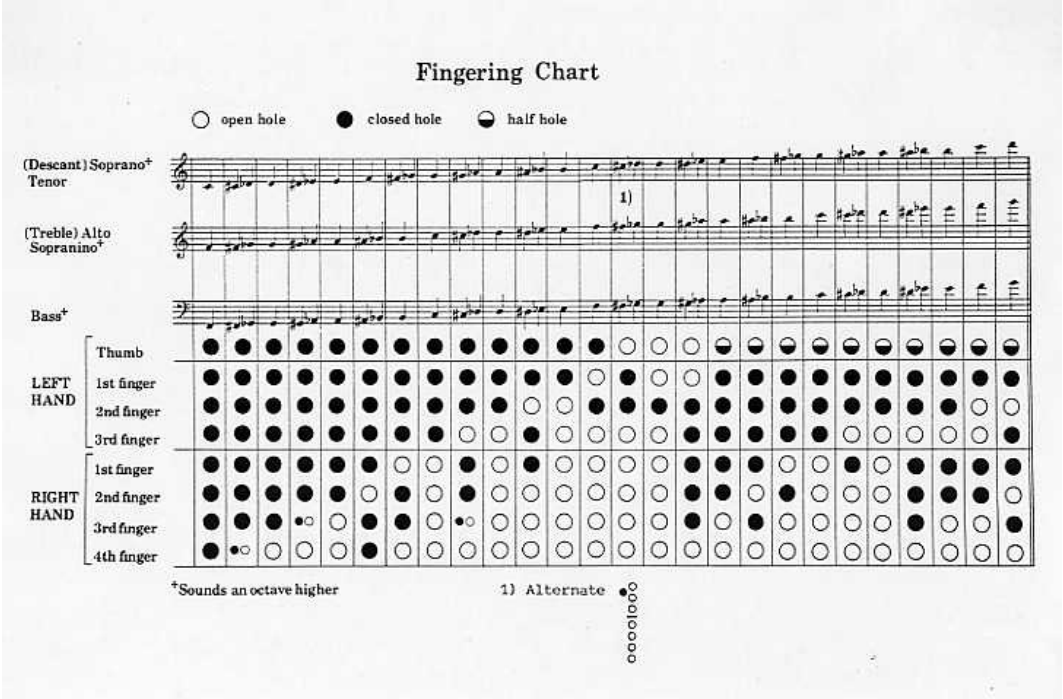

Figure 2.8: Example of fingering chart for different recorders.

The dynamics are achieved through a change in the blowing pressure, although this phenomenon also causes the pitch to change. The pitch of the instrument is correlated to the blowing pressure, as it was claimed by Bak in 1969, when he found

that a power relationship exists of the form [14]:

$$f = k \times P^{\alpha} \tag{2.2}$$

where k and α are constants. A similar relationship has been found in the transverse flute and it follows the relationship [10, 15]

$$P = 0.8 \times f \tag{2.3}$$

In order to keep the pitch between a certain acceptable range, the dynamics in the recorder are achieved through a combined effect of alternative fingerings and blowing pressure. The relationship between dynamics and blowing pressure is also approximately proportional, although it depends strongly on the musical context, i.e. dynamic can only be considered in a context. Montgermont [15] studied the relationship between dynamics and blowing pressure for the transverse flute. Figure 2.9 shows three different dynamics (pp, mf, ff) for different fingerings. For a given pitch, the amount of blowing pressure needed for achieving a higher dynamic is obviously higher. Basically, we can affirm that the blowing pressure, the pitch and the dynamics are highly correlated in the case of the recorder.

The fixed geometry of sound producing parts of the recorder limits the ways in which the player can add expression to the notes played. One resource available is the vibrato. In the recorder, there are two different techniques in order to achieve the vibrato on a note: the breath vibrato and the finger vibrato, i.e. flattement. The breath vibrato is achieved through a repetitive increase and decrease of the blowing pressure by means of a rhythmic contraction of the diaphragm. This produces a corresponding change in the pitch and loudness of the note, but as the harmonics change by differing amounts, the final result is also a variation in timbre[14]. The finger vibrato is executed by only partially covering an open hole somewhere below the lowest closed hole of the note normal fingering.

Another way of adding expression to a note could be with a tremolo. It can be achieved either by tonguing techniques (vibrato linguale) or enunciating an uvular "r". As it was commented before, the dynamics and the pitch are correlated in the recorder, therefore, from a purist point of view if we consider a tremolo a modulation in the loudness of a note, it seems that it is not possible to achieve it without slightly changing the pitch of the note at the same time.

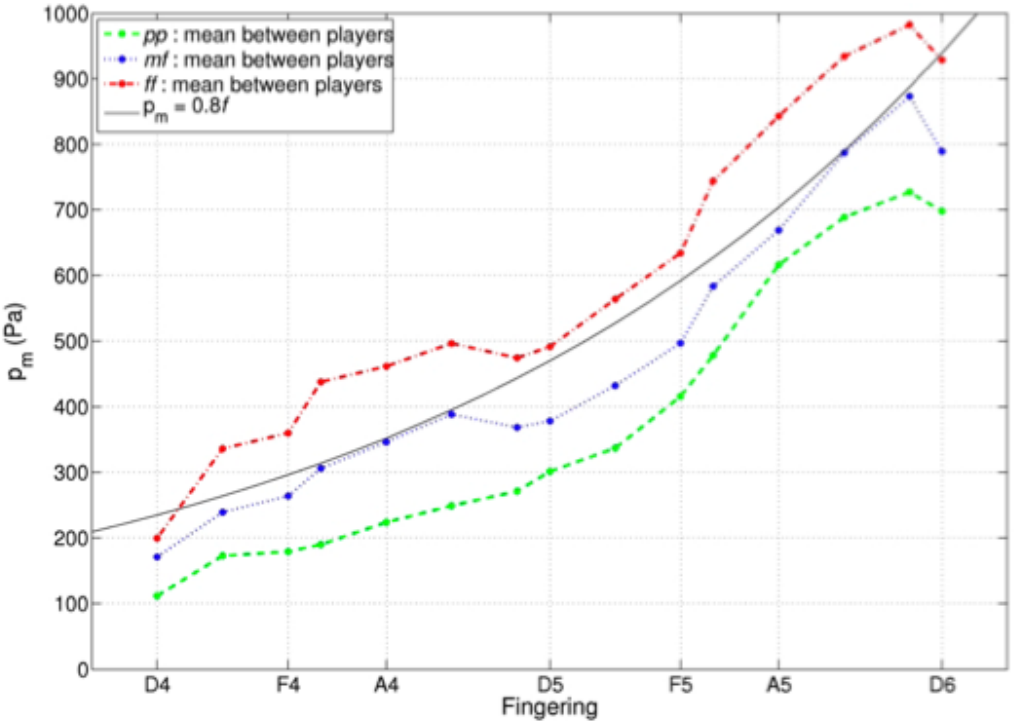

Figure 2.9: Behaviour of the blowing pressure in the transverse flute for different dynamics and fingerings [15].

2.4 Previous studies in the acquisition of the blowing pressure

In order to devise a system which must capture the interaction between an instrument and a player, there are three different ways that one could follow [23]:

- Direct acquisition: one or various sensors are used to monitor different physical features of a gesture. The signals coming from the sensors are synchronized to the recorded sound in order to understand the cause-effect relationship between gesture and sound. One independent data stream is obtained for each control parameter.

- Indirect acquisition: gestures are extracted from the structural, spectral and statistical properties of the sound produced by the instrument. The sound descriptors provide information about the performer actions in order to establish a sound to gesture relationship.

- Physiological signal acquisition: these signals provide information about muscle electrical signals. Therefore, they can be used in order to understand how the brain sends the necessary information to perform the different gestures.

For the scope of this work, we will focus on the direct acquisition. There has been many studies trying to measure directly the blowing pressure in flute-like instruments. Many of these studies cover the description of the acoustic properties of the instruments, but we are more interested in the performance of the musician than in those properties. Although, there are many studies that use a mechanical device in order to send a known blowing pressure to the flute [4, 21], we will focus on the ones that try to measure directly the blowing pressure of a human performer.

One of these early studies is the one by Fletcher in 1975 [10]. He tried to establish the relationship between the blowing pressure and the fingering, among other parameters. For that purpose, he used a 1mm catheter tube inserted into one corner of the lip opening. The tube was then connected to a sensitive aneroid pressure gauge which had been calibrated by comparison with a water column. His results will be commented in the next section. This approach with the catheter in the performer mouth has been used in many other studies: in the measurement of respiratory parameters during performance [6], in the measurement of performance techniques [15, 7] or in the measurement of frequency content of the breath pressure [18].

There are not many studies which cover the modification of the original instrument in order to allow a less intrusive and more accurate measurement of the blowing pressure [2]. This fact encourages us to try to devise an acquisition system based on a modified recorder, as we will see in the following chapter.

2.5 Applications of instrumental gesture acquisition

The measurement of instrumental gestures is a quite young research field, although its applications are very promising. With this computational framework, a great amount of information is extracted from the musical performance process. Therefore, a new set of applications are now possible. In this section, we will only comment a few of them.

From a sound synthesis perspective, this work could contribute to improve the controllability of sample-based models through spectral transformations that are meaningful to instrumental controls. At the same time, it could contribute to a better sound quality for physical models because of the availability of appropriate synthetic controls [13].

From an instrument design perspective, this framework would allow to improve the traditional design process or to augment the capabilities of the classical instruments. Since the gestures are measured in detail, one could devise a new instrument that allows new gestures or one could design an instrument that uses the traditional gestures to feed a different synthesis model, e.g. the virtually real flute [24].

Furthermore, another important application would be the pedagogical one. With the measurement of the gestures, a teacher would have detailed information about the performance techniques of his pupils. In this way, he could provide better and personalized advices to each one, since he would know how good the gestures are being executed. Moreover, the techniques could be described with the shape of the gesture curves instead of using different verbal approximations that make it really difficult to learn how to play the instrument.

2.6 Specific goals of this work

After revising the state of the art of the different fields related to this work, it is necessary to adapt the main goals of the project to the work done by other researchers. Therefore, the main goals will finally be the following ones:

- Study the most meaningful control parameter: the blowing pressure.
- Direct acquisition of the blowing pressure in the less intrusive way.
- Design a recording script including a balanced set of performance techniques.
- Build a multimodal database with sound and gesture.
- Find a good set of features that are useful for analyzing the recorded pressure profiles.

This list should be the core of the project. However, all this work could be evaluated through (1) the generation of synthetic gestures from the profile parameterization and (2) the sound that could be synthesized with the help of a calibrated physical model. Due to time constrictions, these two last steps are considered to be extra goals.

Chapter 3

Acquisition of blowing pressure

In this chapter the acquisition system designed for this work is described in detail. First, the different possible design approaches are commented in order to get an idea of which is the less intrusive and the most accurate measurement. Then two prototypes were built with different designs. Each one has its own advantages and drawbacks which are also analyzed in detail.

3.1 Design approaches

As we have seen in the previous chapter, the majority of the blowing pressure measurements in flute-like instruments have been done from outside the instrument, i.e. the structure of the instrument has not been changed. The main reason for this fact is that changing the structure could easily alter the timbre of the produced sound. Therefore, the only way in which one could measure the blowing pressure inside the instrument would be to ask to a professional instrument builder, i.e. a luthier, for some special mechanical alterations that would allow the insertion of some pressure sensors inside the instrument without altering the timbre. In the following sections, we will describe the outside and inside approaches and we will comment its advantages, drawbacks and limitations.

3.1.1 Measurement from outside the instrument

The majority of the blowing pressure measurements have been done with the following approach: a pipe (catheter) connects the mouth of the performer with a pressure sensor and then the pressure sensor is connected to an analog to digital converter (ADC) [10, 6, 15, 7, 18]. Figure 3.1 shows a scheme of this approach. There are several options with regard to the measurement point:

- The pipe enters directly the mouth of the performer.
- The pipe is connected in parallel to the beginning of the windway.

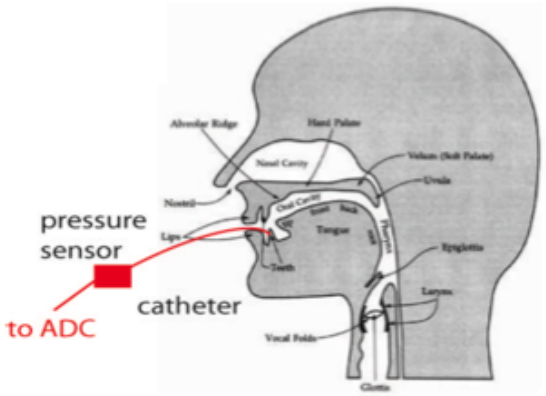

Figure 3.1: Measuring the blowing pressure from outside the instrument.

The main advantage of this design is that the instrument remains unaltered and it is really easy to implement. Nevertheless, this approach has several disadvantages. The first one is its intrusiveness. A performer would not feel comfortable with a PVC pipe inside his mouth and the techniques could not be executed naturally. Furthermore, the air volume that is inside the catheter would introduce an error in the measurement of the blowing pressure because there would be a small inertia in the transients. Therefore, the abrupt changes in the pressure would be smoothed

by the pipe and we would loose information about the shape of the pressure profile in the attacks or the releases of the notes. However, it is a good starting point and it has been used to build the first prototype in this work.

3.1.2 Measurement inside the instrument

If we would like to measure the blowing pressure inside the instrument, we would need to alter the main structure of the instrument, that is to say, we would have to make certain mechanical alterations that allow us to put some sensors inside the instrument or that allow us to connect the sensor to the instrument directly. For that purpose, we have come up with two design approaches that fulfill these requirements:

- A vertical hole should be done directly into the windway of the mouthpiece of a recorder, see figure 3.2 (in red). At the end of that hole a very small pressure sensor should be placed (in blue), so the blowing pressure is measured directly in the windway of the instrument.

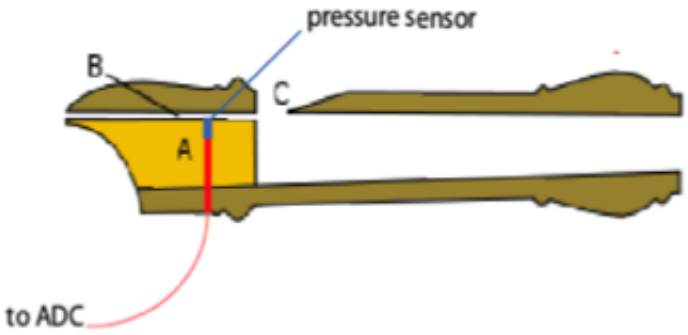

Figure 3.2: Measuring the blowing pressure inside the instrument: internal sensor

- In this case, the same vertical hole should be done as in the previous design,

but now the sensor is placed outside the instrument. The hole acts a as a connection between the measuring point and the sensor, see figure 3.3 . This approach would allows us to use a bigger sensor since it has not to be placed directly inside the hole.

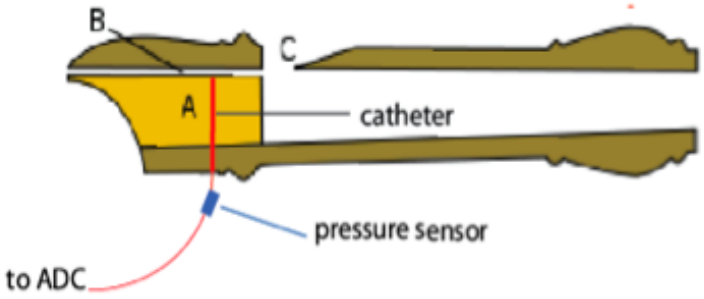

Figure 3.3: Measuring the blowing pressure inside the instrument: external sensor

Both approaches provide a better solution than in the case of the measurement from outside the instrument, since they are not intrusive at all, i.e. the musician can play the instrument normally. Moreover, the inertia error that would be introduced by the pipe could be neglected in the first design or drastically reduced in the latter, although this has not been measured in this work. However, both approaches need to alter the structure of the instrument and this is a very delicate task that could only be performed by an experienced luthier in such a way that the timbre of the instrument remains the same as the original one. Although the idea of inserting a sensor inside the instrument seems to be the most accurate measurement, the availability of such small sensors in the market is not very high. Therefore, the second approach showed in this section is the one used in the final prototype, as it will be explained later.

3.2 Prototype I: external acquisition

The first prototype developed for this work follows the approach of the measurement from outside the instrument. It is basically a reproduction of the work done by other researchers and it is a good first step in order to detect the problems in the measurement of the blowing pressure.

3.2.1 Design

A plastic soprano recorder made by Zen-On©[1] has been used for this first prototype. It is a light recorder that is normally used by students in order to learn how to play the instrument, i.e. it is a beginner's recorder. Figure 3.4 (left) shows a photography of the complete system designed for this prototype.

A PVC pipe was attached to the mouthpiece of the instrument. The opening of the tube was fixed at the same height as the mouth opening of the recorder. In this way, the blowing pressure could be measured at the same point in which the musician is blowing the instrument. The other extreme of the pipe was connected to a differential pressure sensor that is connected to an analog-to-digital converter (ADC). In order to minimize the length of the PVC pipe, the ensemble of acquisition card plus pressure sensor was attached to the body of the performer. In this manner, the error inertia, introduced by the the air volume inside the pipe, could be decreased. Figure 3.4 (right) shows a photography of a performer playing with this system. Figure 3.5 shows the detail of the pipe connection at the mouth opening of the recorder. It can be clearly observed that there is a small hole between the pipe and the recorder. This hole could be another source of errors in the measurement, so it has to be taken into account. The remaining technical details of the system are described in the following section.

[1] http://www.morizono.co.jp/ZEN-ON Recorders.htm

Figure 3.4: Prototype I: complete system (left) and during performance (right).

3.2.2 Specifications

Once the mechanical design is specified, the pressure has to be transduced to an electrical signal by the pressure sensor. This electrical signal is then digitalized in the ADC of the acquisition card and sent to the computer where it is stored in a text file. All the elements that participate in this chain are described in detail in the following sections.

Figure 3.5: Prototype I: detail of the pipe connection at the mouth opening of the recorder

3.2.2.1 Pressure sensor

The first constriction for the pressure sensor is the range. From the literature we know that the blowing pressure range goes from 0 Pa until approximately 3000 Pa. We also want to measure the short moments where the pressure is negative, i.e. when the performer is inhaling. Therefore we need a differential sensor that has at least a range of 0-3000 Pa. Furthermore, we need a relatively small sensor because in this prototype it is going to be attached to the body of the musician. From all the possible sensors of the market, we decided to select the piezoresistive transducers because they offer great accuracy together with a small size. Moreover, we selected the brand Freescale Semiconductor©[2] because they integrate on-chip signal conditioning and temperature compensation and calibration. The model selected is **MPVX4006GC6U** and it is shown in figure 3.6.

[2]http://www.freescale.com/

Figure 3.6: Freescale pressure sensor MPVX4006GC6U.

3.2.2.2 Acquisition card

The selected acquisition card is an Arduino©[3] board. These acquisition cards are very cheap and versatile boards that allow easily the implementation of prototypes. The board consist of a certain number of analog inputs and outputs as well as digital inputs and outputs, an on-board clock, an analog-to-digital converter (ADC), a digital-to-analog converter (DAC) and a communication system that allows the communication with the computer. The selected model is **Duemilanove** and it is shown in figure 3.7.

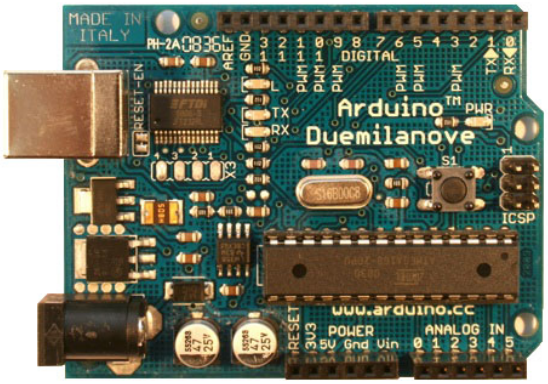

Figure 3.7: Acquisition card Arduino Duemilanove©.

[3]http://www.arduino.cc/

The most interesting feature for our purposes is the sampling rate. Although the clock of the converter provides very high sampling rates, the bottle neck of this board is the communication system. It has a virtual USB connection, i.e. it provides RS-232 communication through a USB port. Therefore, the maximum sampling rate depends on the bauds available on the serial port which is normally 39800 or 55600 bps. With these bauds we obtain a sampling rate for the analog signal of approximately 2000Hz and 3000Hz, respectively.

3.2.2.3 Software

In order to store the sampled signal, we used the API provided by Arduino© together with a PureData patch. The API simply read the value coming from the board and wrote it to the serial port. Then the PureData patch read that values and store them in the RAM memory, until the experiment had finished. In that moment, all the data were written to a text file in the hard disk. In this way, the digitalized pressure signal was stored for its latter analysis in Matlab©.

3.2.3 Problems

After some preliminary tests were done, we realized that this prototype has many limitations. The most important has to do with the sampling rate. Since the blowing pressure has a coupling frequency component related to the note that is being played in the instrument (we will explain this phenomenon in detail in the following chapter), higher sampling rates are needed than the ones provided by the Arduino© board.

Furthermore, the PVC pipe attached to the mouthpiece of the recorder is very intrusive and therefore a player cannot play naturally with it. This fact could alter drastically the measurements because the performer should adapt himself to this

new recorder. Moreover, the pipe introduces an error in the pressure measurement because there is an air volume inside the pipe that smoothes the transients in the pressure profile because of the inertia of the air volume.

For all these reasons, we decided to implement a better prototype that could solve all these issues. It is explained in the following section.

3.3 Prototype II: modified instrument

The second prototype follows the approach of measuring inside the instrument, i.e. the main structure of the recorder was modified in order to be able to connect some sensors directly in the instrument. For those modifications we counted on the help of a experienced recorder luthier, **Josep Tubau**. He modified the mouthpiece of a wooden contralto recorder that he has also designed himself. A brief explanation of the main modifications can be found in the following section and the technical details are placed in the Appendix A.

3.3.1 Design

The mouthpiece of a contralto recorder was mechanically modified by Josep Tubau. These modifications provide two measuring points: one for the blowing pressure and one for the internal pressure. The first measurement is achieved through a connection between a hole at the mouth opening (i.e. the beginning of the windway) with one side of the mouthpiece. In this way, the air coming from the lunges of the performer flows through the windway and at the same time it flows through the hole until the pressure sensor placed exactly outside the instrument. Figure 3.8 shows the details of this connection.

The second measurement is achieved through the connection between a hole placed

Figure 3.8: Details of the mechanical modification of the mouthpiece. Left: hole for sensing the blowing pressure. Right: connection for the pressure sensor.

at the beginning of the resonator pipe just after the lip labium, and a hole place at one side of the mouthpiece where the second pressure sensor is connected. In this manner, the internal pressure of the recorder can be measured. Figure 3.9 shows how these connections were done.

Although this measurement is not relevant for the modeling of the performer gestures because it is made after the lip labium, it provides meaningful information about the development of the sound in the recorder.

With these modifications, some problems of the first prototype were solved: the inertia error introduced by the PVC pipe can be practically neglected in this prototype and the intrusiveness is really decreased with regard to the first prototype since the performer can play the instrument normally.

3.3.2 Specifications

Besides the mechanical modifications, the rest of the components in the measurement chain were also changed in oder to overcome certain limitations of the first prototype.

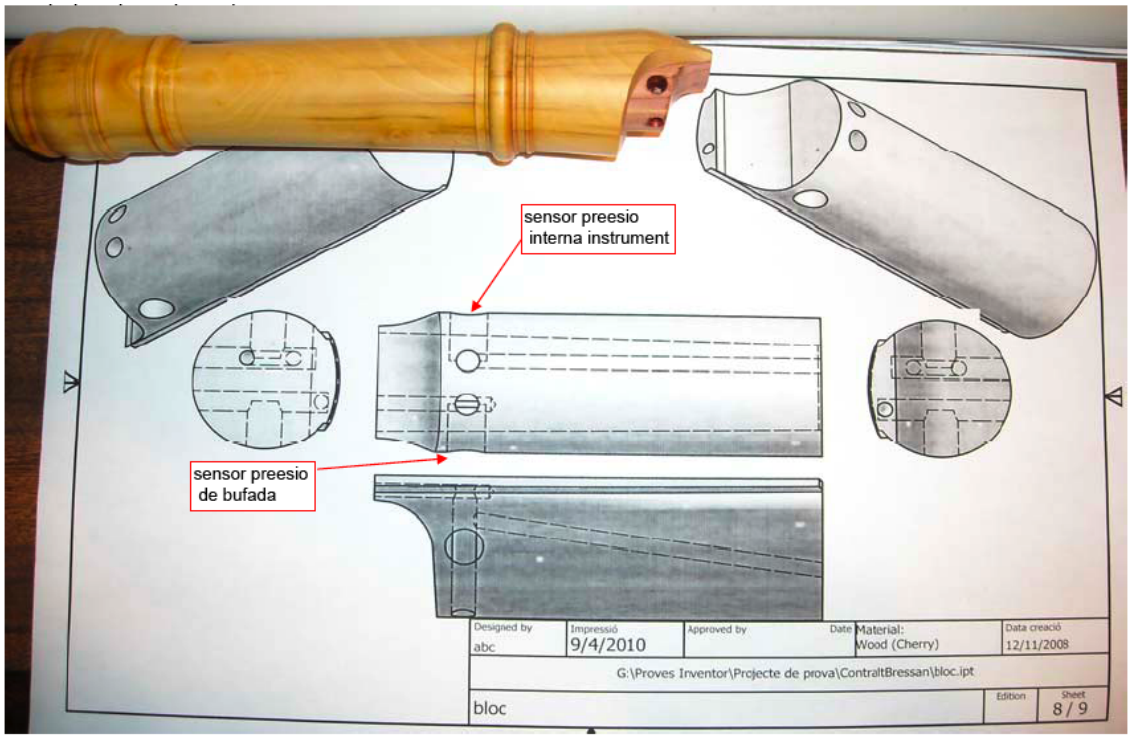

Figure 3.9: Scheme of the different mechanical modifications of the mouthpiece.

3.3.2.1 Pressure sensor

The pressure sensor had to be changed because the port connections had to match the hole size made in the recorder. Therefore, two sensors provided by the luthier Josep Tubau were used instead of the ones from the first prototype. The working principle of these sensors is very similar to the ones used before: a piezoresisitve transducer is attached to a silicon membrane that moves proportional to the pressure applied to it. They belong to the brand Honeywell©[4] and the model is the ACSX01DN series. Furthermore, the pressure range is enough for the one we want to measure: from 0 to 6.89 kPa.

[4]http://sensing.honeywell.com/index.cfm/ci_id/154366/la_id/1.htm

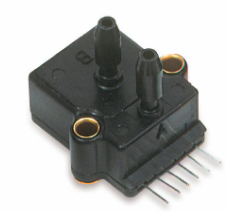

Figure 3.10: Pressure sensor Honeywell© ACSX01DN series.

Figure 3.11: National Instruments© USB 6009 acquisition card.

3.3.2.2 Acquisition card

In order to overcome the limitation of the 3000Hz of sampling frequency provided by the Arduino© board, a new acquisition card was employed. The selected model is the National Instruments© [5] USB-6009. It provides several analog inputs and outputs as well as digital input and outputs. The sampling frequency is 48000 Hz and it is multiplexed with the number of signals that are being acquired. The bit depth of the acquired signals is 14 bit. These features fulfill all our requirements and overcome the limitations we had in the former prototype.

[5] http://www.ni.com/

3.3.2.3 Software

The acquisition card provides a software that allows us to acquire and store analog signals easily. This software is LabView©. It is a graphical programming environment that provides the drivers for the communication with the acquisition board and allows the implementation of simple programs for storing the acquired signals. Furthermore, it permits full control on all the parameters of the acquisition card. A simple but robust acquisition program was developed in order to acquire all the desired signals. This program will be explained in the following chapter.

Chapter 4

Database construction

In this chapter the steps followed in order to build a multimodal database with sound and gesture are described. First, the studio setup for the recordings is explained and then the features of the acquisition program developed in LabView© are detailed. After that, the design of the recording scripts is discussed and a basic statistical analysis on the recorded database is performed. Finally, the pre-processing of the raw data is commented in detail, so the reader can move on to the data analysis on the next chapter.

4.1 Recordings setup

Once the final prototype was designed and tested, a studio session with a professional recorder player, **Joan Izquierdo**, was arranged in order to record certain scripts to build the database. He is an experienced musician and music teacher who has collaborated in several technological experiments. Therefore, he is used to work in these special conditions where the instrument has some sensors on itself. The recordings were performed in the studio of the Pompeu Fabra University.

4.1.1 Studio setup

The studio was setup in a special way because we had two different acquisition systems: the mixing desk and the acquisition card. Figure 4.1 shows a general scheme of the recordings. This configuration is designed to acquire the following three signals:

- **The blowing pressure**: this signal has two components. The first one is a DC value corresponding wih the pressure applied by the performer and it also has an oscillation with the frequency of the note that is being played, which is going to be explained later.
- **The internal pressure**: it is an oscillation around 0 Pa. Since it has not any DC component, it can be recorded using a conventional sound card without loosing any information. With the help of a direct injection box, this signal was directly connected to the mixing desk.
- **The microphone sound**: we are also interested in the analysis of the produced sound. Therefore, an Audio Technica© 350 condenser microphone attached to the recorder was used, as it is shown in figure 4.2.

The blowing pressure was acquired through the National Instruments© acquisition card and the internal pressure and the sound were recorded through the mixing desk. The fact that there are two different acquisition devices introduces a new problem: the synchronization. As it will be explained later in this chapter, a metronome had to be recorded in both devices in order to be used as a synchrony click.

The interest in recording the produced sound comes from the fact that this work is a collaborative project with another student Leny Vinceslas. His work covers the relationship between the produced sound and the gesture. Therefore, he will use

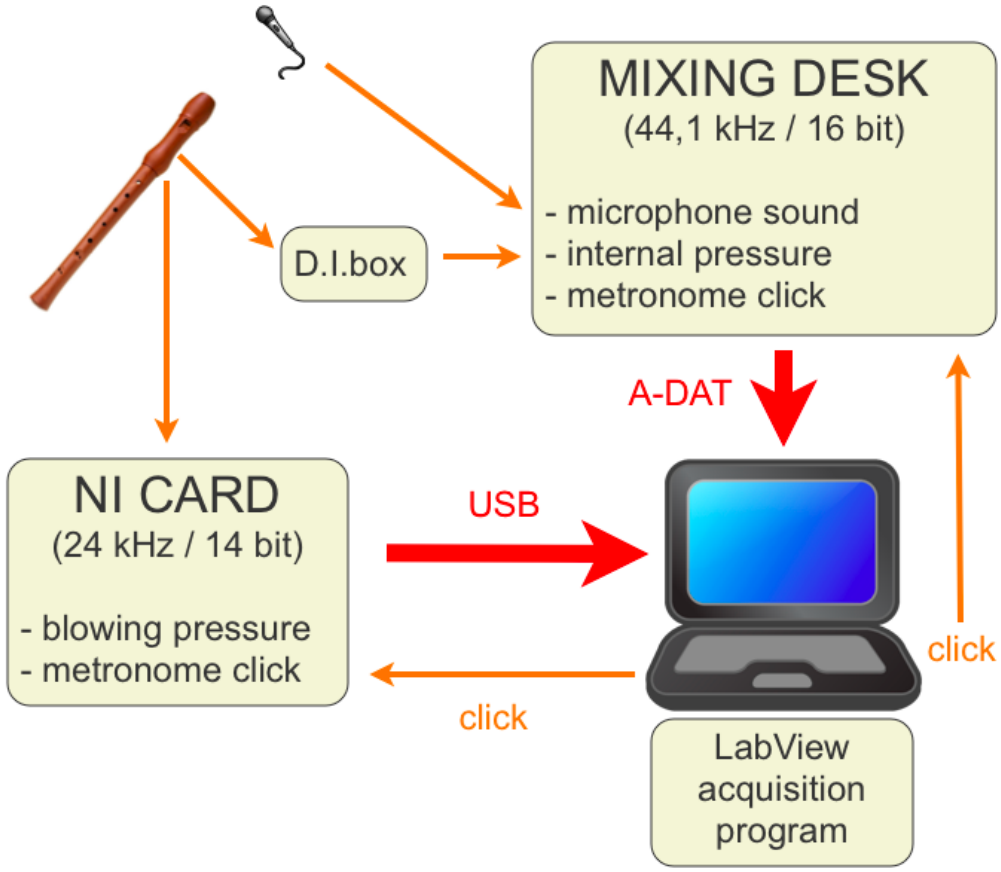

Figure 4.1: General studio scheme for the recordings.

the recorded data in order to build a model that is able to synthesize the sound for a given gesture curve.

4.1.2 LabView© acquisition program

This graphical programming environment enables the acquisition and storage of the three previous signals. It allows the configuration of the different acquisition devices independently. The sound, the internal pressure and the metronome were acquired through the mixing desk with a sampling rate of 44100 Hz and a bit depth of 16

Figure 4.2: Detail of the Audio Technica© 350 condenser mic.

bit. The produced sound and the metronome are saved in the same wave file as a stereo file whereas the internal pressure is saved in a different one as a mono file. The blowing pressure and the other metronome signal were acquired through the acquisition card with a sampling rate of 24000 Hz and a bit depth of 14 bit. These data were saved into a text file. The output of the acquisition program for each recording is a set of 3 files: 2 wave files with the signals recorded with the mixing desk and a text file with the signals acquired through the card. Both devices were configured with different software buffers in order to assure that any sample is not lost during the acquisition process.

4.2 Recording script

In the design of these scripts the main idea deals with the creation of an analysis space with some defined dimensions which covers a balanced set of performance

techniques. Since the blowing pressure is the only gesture that is being captured, the whole possible gesture set cannot be covered, but only a subset. Nevertheless, the definition of these dimensions will allow us to establish some relationships between the score and the gesture and to build a performer model that is able to synthesize the gesture curves for a given point in this analysis space. The dimensions of that space are the following:

- **Pitch:** the recordings cover different pitches from the three registers of the instrument. As it was commented in chapter 2, there is a clear correlation between the blowing pressure and the performed pitch in the recorder.
- **Dynamics:** each score is recorded with three different dynamics: pianissimo, mezzo-forte and fortissimo.
- **Note duration:** five different note durations were recorded: 0.66, 0.33, 0.166, 0.25 and 0.125 seconds. These durations correspond to the duration of a quarter, eighth and sixteenth note at 90 BPM (beats per minute), and a eighth and sixteenth note at 120 BPM, respectively.
- **Articulation:** this dimension refers to how the performer uses the shape of the mouth and the tonguing in order to provide a certain attack and release to the played notes. Four different values for this dimension were defined. Each one of these values refers to a certain syllable sequence that is a guide to the performer for articulating each note. These four values are:
  - Full legato: no tonguing is used to articulate the notes in this case. The performer was specifically asked not to use the tonguing so that this value can be considered as an extreme of this dimension.
  - Legato: the syllable sequence “LU-RU-LU-RU” is used in this case to articulate.

  - Soft staccato: the performer uses the syllable sequence "DU-GU-DU-GU".
  - Staccato: in this case the syllable sequence "TU-KU-TU-KU" is used to articulate the note.

- **Interval:** this dimension refers to the semitone difference between one note and the following. Two different scales with different intervals between the notes were recorded: 1, 3, 5, 7 and 9 semitones were the selected values for this dimension.

With all these dimensions, a recording script made up of four main parts was designed. In the first part, named the repetitions, the same note is recorded with different dynamics and articulations for 16 different pitches. In the second part, named the scales, the focus is on the intervals and they were achieved through the definition of a musical scale. In the third part, five different musical pieces were recorded in order to use them for a later evaluation of the gesture model. Finally, some extra material was recorded for different purposes that go beyond the scope of this work.

4.2.1 Repetitions

In the first part of the recordings, the performer played the score shown in figure 4.3 for 16 different pitches, the three different dynamics, the four different articulations and two different tempos: 90 BPM and 120 BPM. The selected pitches were:

- F4, F4#, G4, A4#, C5, D5, E5, F5, G5, A5, A5#, C6, D6, E6, F6 and A6

In this way, the three registers of the recorder were equally covered because the contralto recorder used for the recordings is tuned to F. The main purpose of this

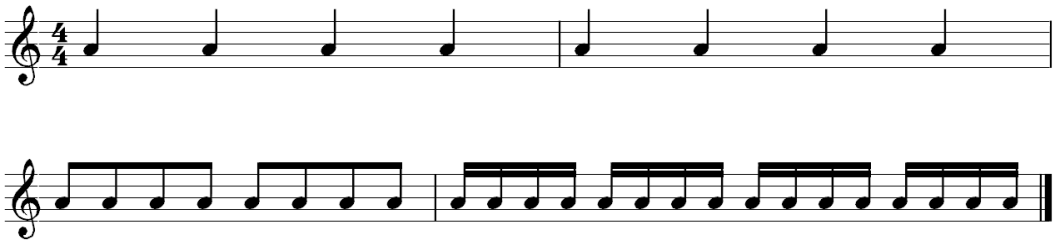

Figure 4.3: Example of the score used for the repetitions. Note A4.

part of the recordings script is to devise the shape of the blowing pressure profile for each dimension independently.

4.2.2 Scales

Two different scales were designed with the same intervals, but different note durations. The second scale has half notes with regard to the first one. The composition of the scale tries to cover the same number of ascending and descending intervals being at the same time musically meaningful. Both scales were recorded at 90 BPM. Figures 4.4 and 4.5 show the scores of the scales.

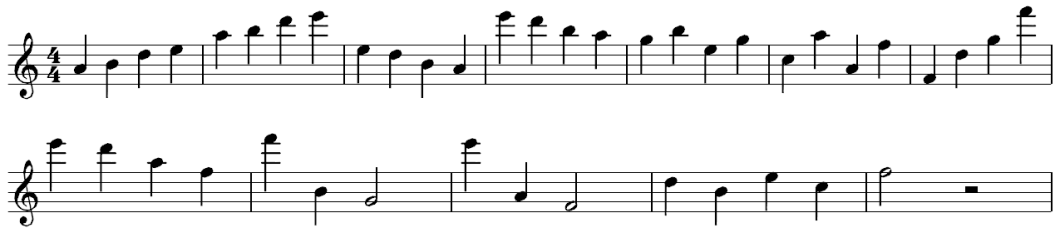

Figure 4.4: Score of the first scale.

Both scales were also recorded with the three different dynamics and the four articulations. In this manner, the influence of the interval can be compared later with the repetitions and the change of the blowing pressure profile can be observed, if there is any. The musical context is taken into account in this case whereas the

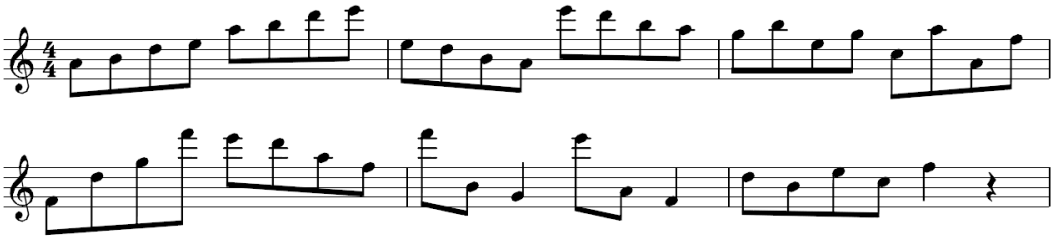

Figure 4.5: Score of the second scale.

repetitions are used as a reference point.

4.2.3 Musical pieces

Five musical pieces were selected by the musician Joan Izquierdo. They were extracted from the book "Preludes and Voluntaries". It is a typical learning book used in his recorder lessons and it has a great collection of compositions for recorder solos. The main purpose of these recordings is to analyze the shape of the blowing pressure profile in a real performance. The repetitions and the scales could be used to build a gesture model as a training set and these pieces could be used to evaluate it.

The musician Joan Izquierdo was asked for playing each piece with the tempo, articulation and dynamics that best fit them in each case. In fact, there are different articulations and dynamics during the same piece and the tempo varies slightly from a nominal value. The scores of the five pieces can be found in the appendix B.

4.2.4 Extra material

These recordings were done for the project of Leny Vinceslas and for the interest of the luthier Josep Tubau. Leny is interested in the relationship between sound and gesture. Therefore, he needed some data about the overblowing and the attacks.

Josep is interested in the influence of the vocal tract on the final sound produced by the recorder. The "Vowels" and "Voiced" recordings were done for him. These recordings cover some issues that are out of the scope of this work. Therefore, a brief explanation of each one is provided here.

4.2.4.1 Timbre

This section covers the overblowing phenomenon. The performer was asked for increasing the pressure slowly while playing the same note. In this way, the vibration mode of the instrument jumps from the first mode to the second one. This transition and the spectral distribution of each mode are really important data for the modeling of the sound-gesture relationship. A large set of different pitches was recorded with two different fingerings: the normal one and the alternative one. Whenever a note was playable with two different fingerings, it was recorded with both. The fingering is another necessary parameter for building the model that synthesizes the sound for a given gesture, which is the case of Leny's project.

4.2.4.2 Attacks

The performer was asked for playing ten different notes with three different attacks each one. These data are useful for modeling the sound-gesture relationship because they provide meaningful data for training the model in the special case of the note attacks.

4.2.4.3 Vowels and voiced

In this part of the recordings, the musician Joan Izquierdo was asked for playing ten different notes with five different mouth shapes each corresponding to one vowel, i.e. each note was played while the performer was saying the vowel while blowing.

Furthermore, for the voiced recordings he was asked for saying some words while blowing. The purpose of these recordings is to analyze if the shape of the vocal tract affects somehow the timbre of the produced sound.

4.2.5 Statistical analysis of the database

For the scope of this work, the focus will be on the repetitions and the scales. Both of these recordings formed a set of approximately 10500 single notes. A basic statistical analysis was performed on these notes over three different dimensions in order to see if the database was uniformly distributed. Figures 4.6, 4.7 and 4.8 show the distribution for the articulations, the dynamics and the note durations.

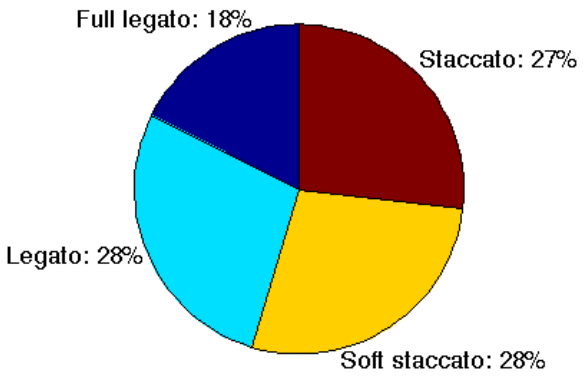

Figure 4.6: Statistical distribution of the database for the four articulations.

As it can be appreciated in the figures, the database is approximately uniformly distributed in these dimensions. It can be observed that the full legato has a percentage lower than the other articulations. This is due to the fact that during the recordings of the repetitions at 120 BPM the full legato was not recorded because it was too fast for the performer to play those notes without articulation. Furthermore, the note durations corresponding to the 120 BPM (0.125 and 0.25 seconds) have a

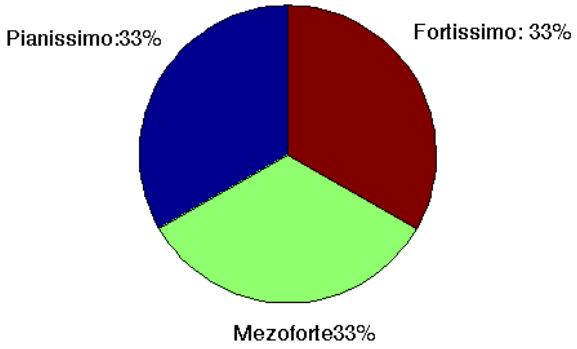

Figure 4.7: Statistical distribution of the database for the three dynamics.

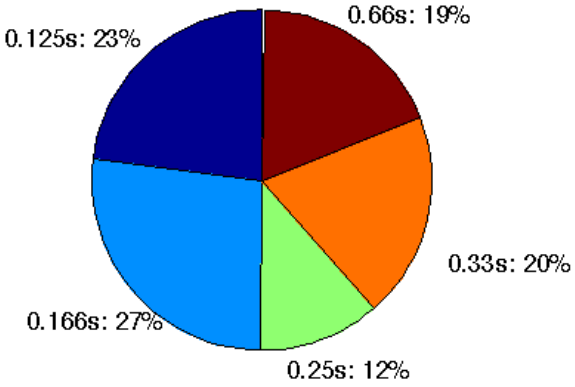

Figure 4.8: Statistical distribution of the database for the five note durations.

percentage lower than the other durations because the scales were only recorded at 90 BPM. However, the distribution seems to be reasonably distributed. Therefore, it can be affirmed that the data are statistically significative.

4.3 Data pre-processing

There are several issues with the raw data which have to be solved before the analysis can be performed successfully. These problems deal with the synchronization of both acquisition systems, i.e. acquisition card and mixing desk, the smoothing of

the blowing pressure curves and the automatic segmentation of each recording into single notes.

4.3.1 Synchronization

As it was commented in section 4.1.1, two different acquisition systems were used during the recordings: the National Instruments© acquisition card saved the blowing pressure and the audio card of the mixing desk saved the produced sound and the internal pressure. The main problem in this configuration is the fact that the acquisition card introduces a time-variable delay in comparison to the audio signals due to the USB asynchronous communication bus and due to software buffers. Therefore, the metronome click had to be recorded in both systems in order to perform a synchronization between the signals acquired with both devices.

In the synchronization process the signals recorded with the audio card are assumed to be the "ground truth" and the signals recorded with the acquisition card are delayed. Therefore, the metronome click recorded with the NI card (named acquisition click) was aligned with the metronome click recorded with the audio card (named audio click). For that purpose, the peaks of both metronome clicks were detected and then the first peak of the acquisition click was aligned with the first peak of the audio click. Once the first peaks are aligned, a loop runs through all the detected peaks taking in each iteration the blowing pressure signal between two consecutive peaks of the acquisition click and resampling it to the time of the audio click between the corresponding peaks. In this manner, the delay of the blowing pressure signal is corrected at each pair of clicks and the signal is resampled to 44100 Hz.

The output of this process is the synchronized blowing pressure signal resampled at the same sampling frequency as the audio signals.

4.3.2 Smoothing

The blowing pressure signal has two components: a DC level that is the main interest of this work and an oscillation with the frequency of the note that is being played. The played note is changed with different fingerings, i.e. different effective lengths of the resonator pipe. The blowing pressure has a coupling frequency that depends on the effective length of the resonator. Since this coupling frequency is not very interesting for our purposes, the blowing pressure was smoothed in order to keep only the DC level. The smoothing algorithm windowed the original signal and approximated each window through a quadratic regression. Two parameters allow the control of this algorithm: the window size and the overlap factor.

There is a trade-off between the window length and the loss of the transient shape. The longer the window, the bigger the transient loss. Several combinations of window length and overlap facotr were tested for different notes. Finally, the selected values for the smoothing algorithm were:

- Window length: 11 milliseconds.
- Overlap factor: 0.5.

Figure 4.9 shows the raw and the smoothed version of the blowing pressure signal. It can be observed that the peaks in the transients are slightly smoothed as it was commented previously. For the analysis of the blowing pressure profiles in the following chapter the smoothed version will be used.

4.3.3 Automatic segmentation

Once the blowing pressure curves have been synchronized and smoothed, each recording has to be cut into single notes. For that purpose, an automatic segmentation algorithm was developed in Matlab©. The algorithm is made up of two

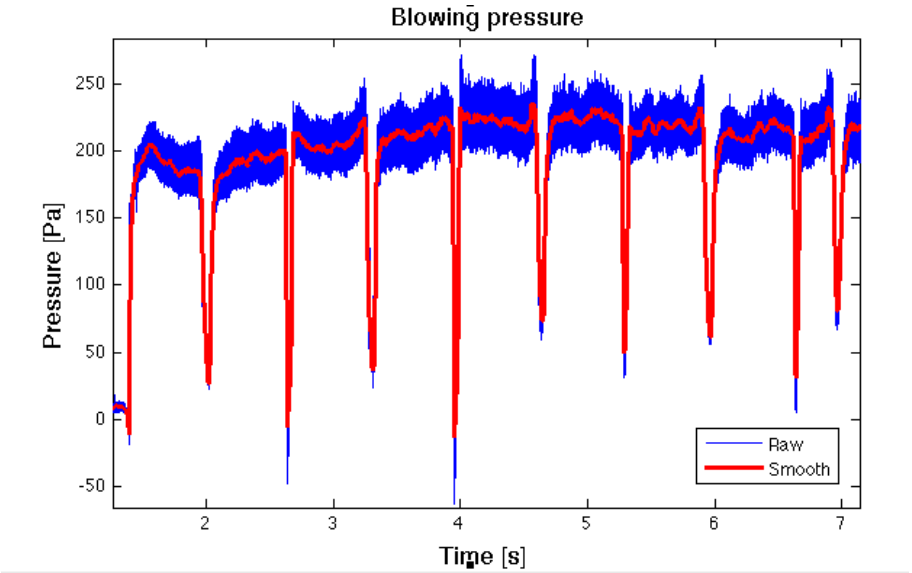

Figure 4.9: Raw and smoothed version of the blowing pressure.

stages: in the first one, a set of onset candidates is generated; in the second one, an evaluation function determines which candidate is the best.

Two criteria have been used for the generation of candidates: the minimum peaks below a certain threshold of the smoothed pressure and the maximum points of the first derivative which were computed as the zero-crossing points of the second derivative. In this way a set of candidates is generated for each single note. In the case of very low pitches, the pressure curve has a ripple (it will be explained in the next chapter) that causes this algorithm to generate a high number of candidates. Therefore, it was necessary to clean them in order to be able to perform a coherent evaluation of each candidate. The cleaning process involved the comparison of the pressure value of each candidate with the pressure value 100 milliseconds later. If the difference between these two values was bigger than a certain threshold, then it meant that the pressure curve was increasing fast and the candidate was accepted.

For the evaluation of all the candidates, an adaptive algorithm that takes into

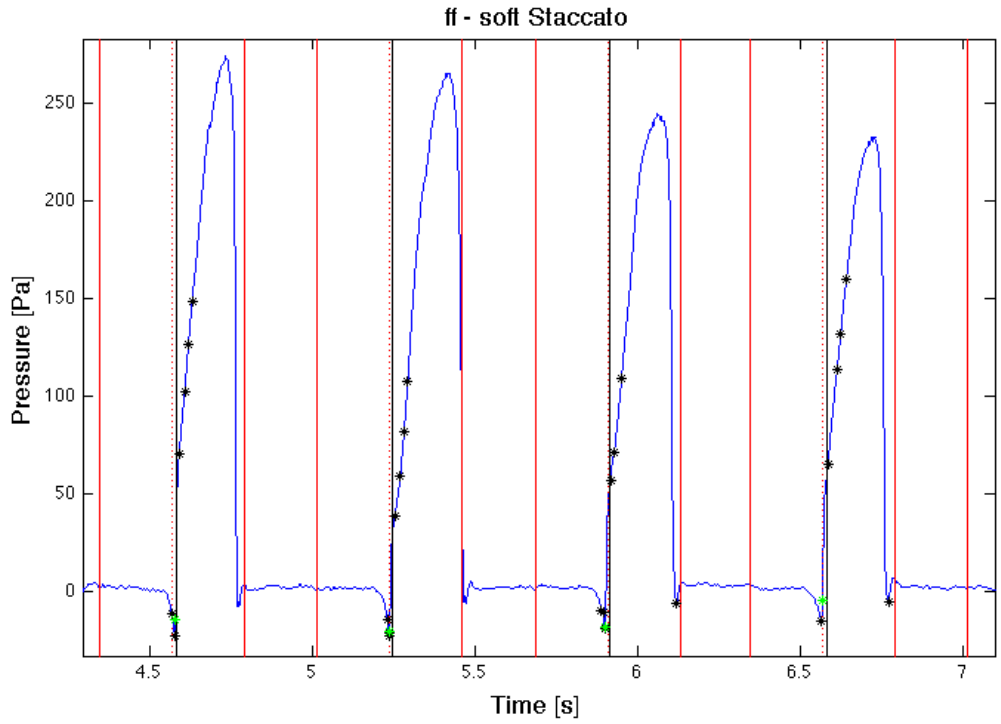

Figure 4.10: Example of the automatic segmentation in a soft staccato recording.

account the natural deviation of the performer from the theoretical onsets was developed. In the first step of the evaluation, a weighted history of that deviation is computed for the last three onsets. Afterwards, a window was created which had its center in the weighted deviation and a size of $\frac{2}{3}$ of the theoretical note duration. All the candidates inside that window were evaluated as possible onsets. For each candidate, a window of ± 11 milliseconds is also created in order to compute the slope between the pressure value of the candidate and the extremes of this window, i.e. each candidate generates two slopes, one regarding past samples of the pressure and the other one regarding future samples. The evaluation function computes then the ratio between the slope of the "future" pressure and the slope of the "past"" pressure. The best candidate is the one that has the highest ratio.

Figure 4.10 shows an example of this process: the blue line represents the pressure

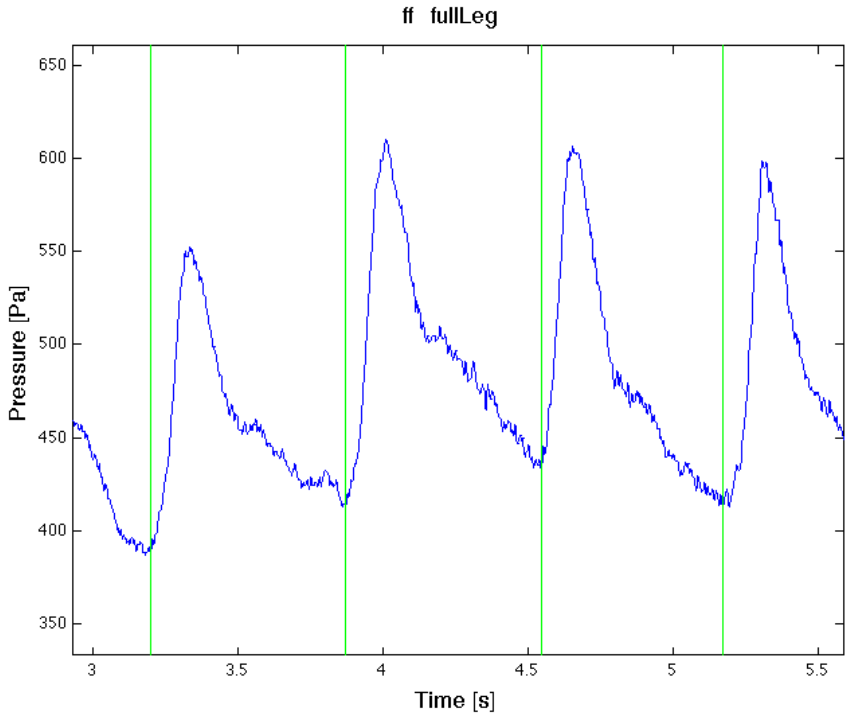

Figure 4.11: Example of the automatic segmentation in a full legato recording.

curve; the vertical black lines are the theoretical onsets; the dashed vertical red lines are the center of the history window for each theoretical onset and the vertical red lines are the extremes of the same window; the black dots are the cleaned candidates and the green dots are the selected onset for each note.

Figures 4.11 and 4.12 show the final results of the automatic segmentation for two different articulations: full legato and legato. The vertical green lines show the onsets of each note.

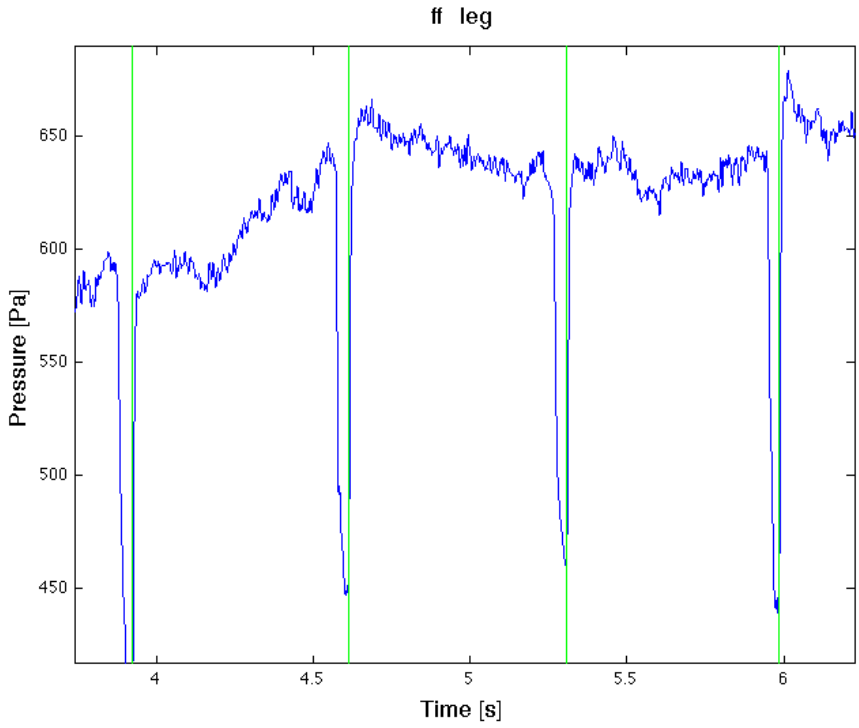

Figure 4.12: Example of the automatic segmentation in a legato recording.

Chapter 5

Data analysis

This chapter describes the analysis performed on the recorded database. First, a set of features is defined, which describes the temporal properties of the blowing pressure profiles of one single note. After that, the shape of these profiles for each articulation is described independently, so that the reader can understand better how the articulation affects the blowing pressure. Later, the relationship between the pitch (i.e. performed fingering), the dynamics and the articulation is analyzed in detail. Afterwards, different features such as the the attack time or the attack slope are analyzed with regard to different analysis dimensions: pitch, articulation, dynamics and note duration. These analysis provide the first results for the temporal modeling of the performer gestures.

5.1 Feature description

For the description of the temporal properties of the blowing pressure profiles a set of features is defined. Each recorded note is depicted by a set of states, times and slopes. Two states are defined: the steady state and the release state. For the state computation the difference between the maximum and the minimum pressure value is calculated and named $\triangle P$. The steady state is computed as the pressure values

bigger than 95% of this $\triangle P$. The release state is computed as the pressure values smaller than 5% of $\triangle P$ at the end of the note. Three times are also defined in order to delimit these two states. The time t_1 defines the beginning of the steady state. It is also defined as the attack time for the staccato articulations and as the rise time for the legato articulations (see 5.4). The time t_2 defines the end of the steady state and the time t_3 defines the beginning of the release state. Finally, two slopes are defined: the attack and the decrease slopes. Both slopes are computed as a lineal regression of the pressure curve between the beginning of the note and the time t_1 for the attack and between t_2 and t_3 for the decrease slope.

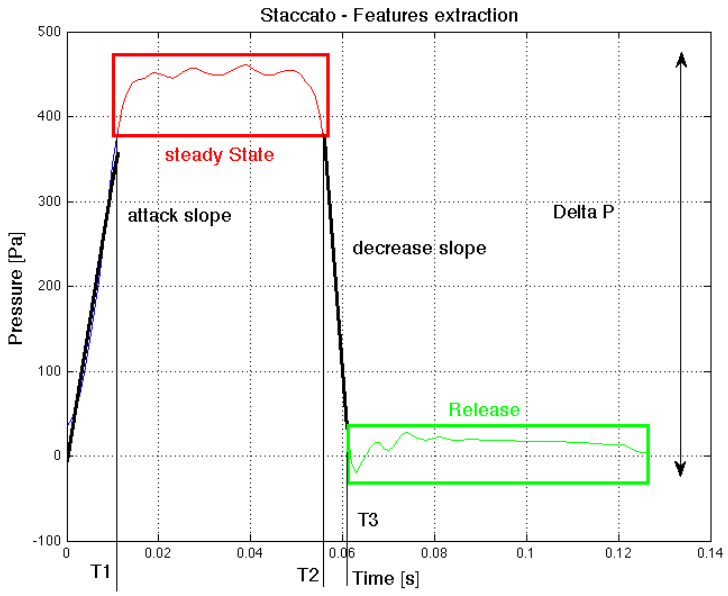

Figure 5.1: Features for a note in a staccato recording.

Figures 5.1, 5.2 and 5.3 show three examples of blowing pressure profiles for three different articulations. In each figure all the features have been drawn in order to understand better the previous explanation. As we can observe in these figures, the three articulations provide very different values for the features. For example, the staccato articulation, shown in figure 5.1, has a steady and release state that are very

similar in length, while the legato articulation (figure 5.2) has a much longer steady state than release state. Therefore, these features can be very useful for classifying the articulations as well as other analysis dimensions such as the dynamics or the note duration, as we will see later in this chapter.

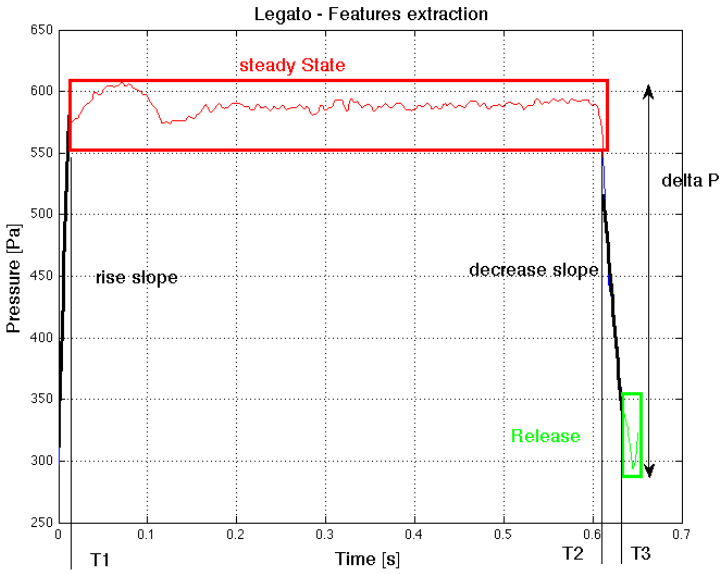

Figure 5.2: Features for a note in a legato recording.

5.2 Articulation shapes

In this section, the different shapes for different articulations of the blowing pressure profiles are described in detail. For each articulation, the note duration is fixed and equal to 0.666 seconds. The longest note was selected for a better representation, but any note duration would give the same results. Three different pitches (i.e. performed fingering) are selected for the representation, one per register. All the recorded pitches cannot be shown simultaneously because the representation would be very difficult to understand. Furthermore, the three different dynamics are also shown in the following figures. Basically, it can be observed how the blowing pressure

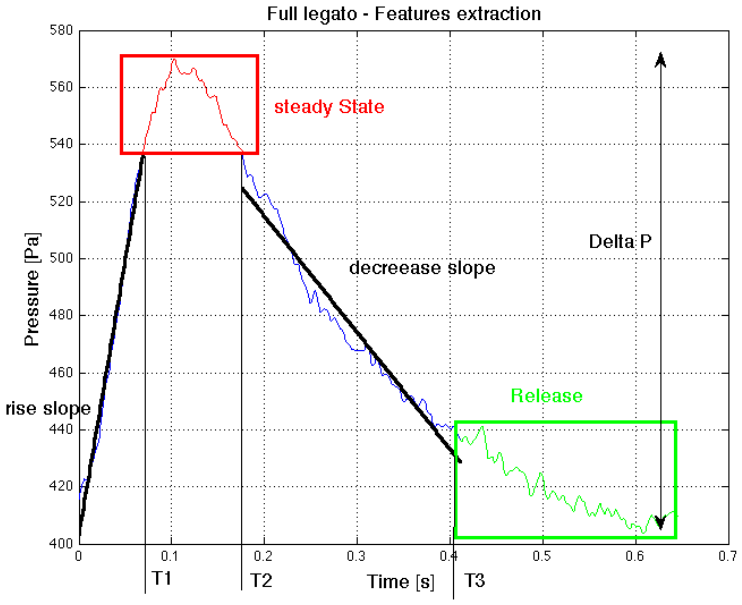

Figure 5.3: Features for a note in a full legato recording.

profiles change for a given articulation and a given note duration with the pitch and the dynamics. The pitch (i.e. performed fingering) and the dynamics are represented in the X and Y axis, respectively.

5.2.1 Full legato

As it was commented in the previous chapter, this articulation was performed without using the tongue to articulate the notes. The musician was specifically asked for doing it in this way. Therefore, we consider this articulation as one extreme of this analysis dimension.

Figure 5.4 shows how the blowing pressure profiles change with increasing pitch (G4, F5 and F6) and dynamics (pp, mf and ff). Each one of the nine diagrams represent a point in the analysis space. The bottom left diagram represents the point of pitch G4 and dynamics pianissimo and the top right diagram represents

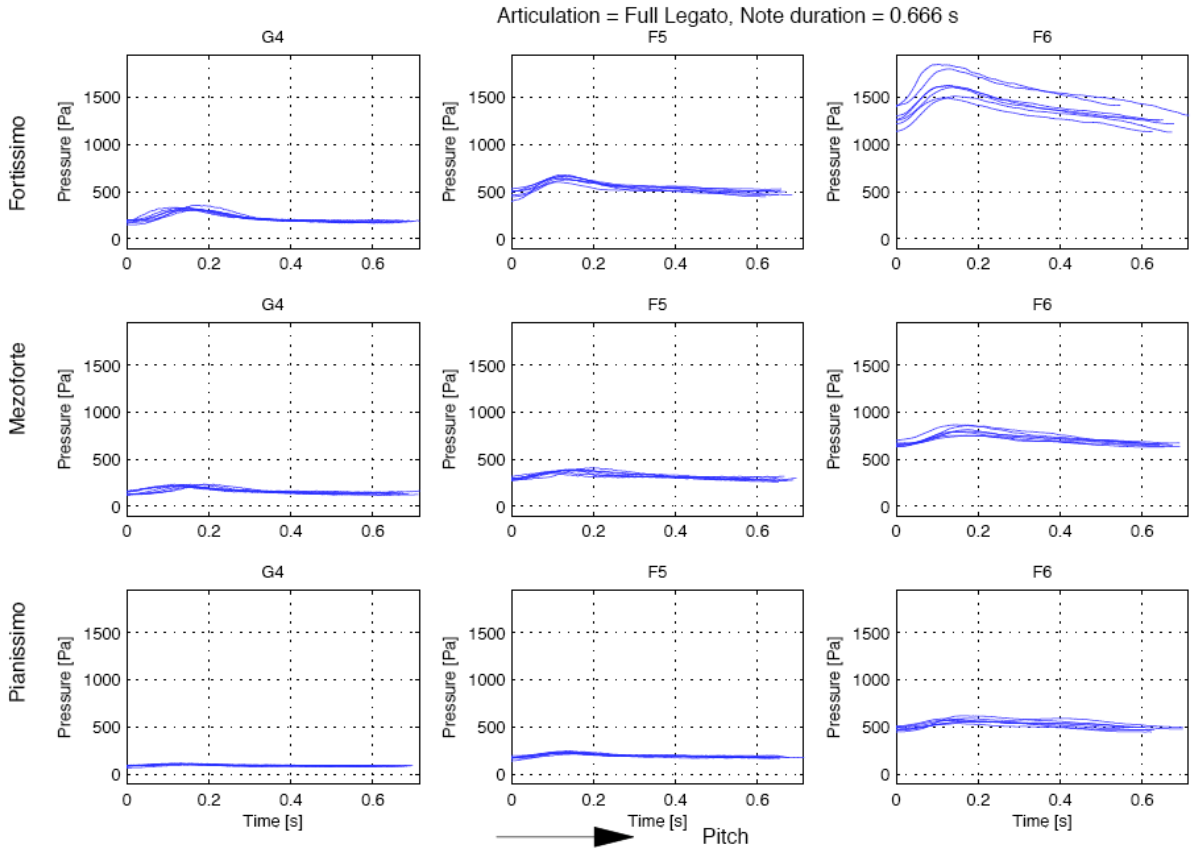

Figure 5.4: Blowing pressure profiles for full legato and note duration of 0.666 seconds.

the point of pitch F6 and dynamics fortissimo. There are several recordings that fulfill the conditions of each point in the analysis space and they are shown with the different curves of that diagram. It can be clearly appreciated how the mean value of the pressure profiles in each diagram increases with increasing dynamics and pitch. This fact is a reproduction of the results by Fletcher and Montgermont [10, 15].

In order to understand better how the pressure behaves with this articulation, figure 5.5 shows a zoom of the previous figure. In this way, it can be appreciated

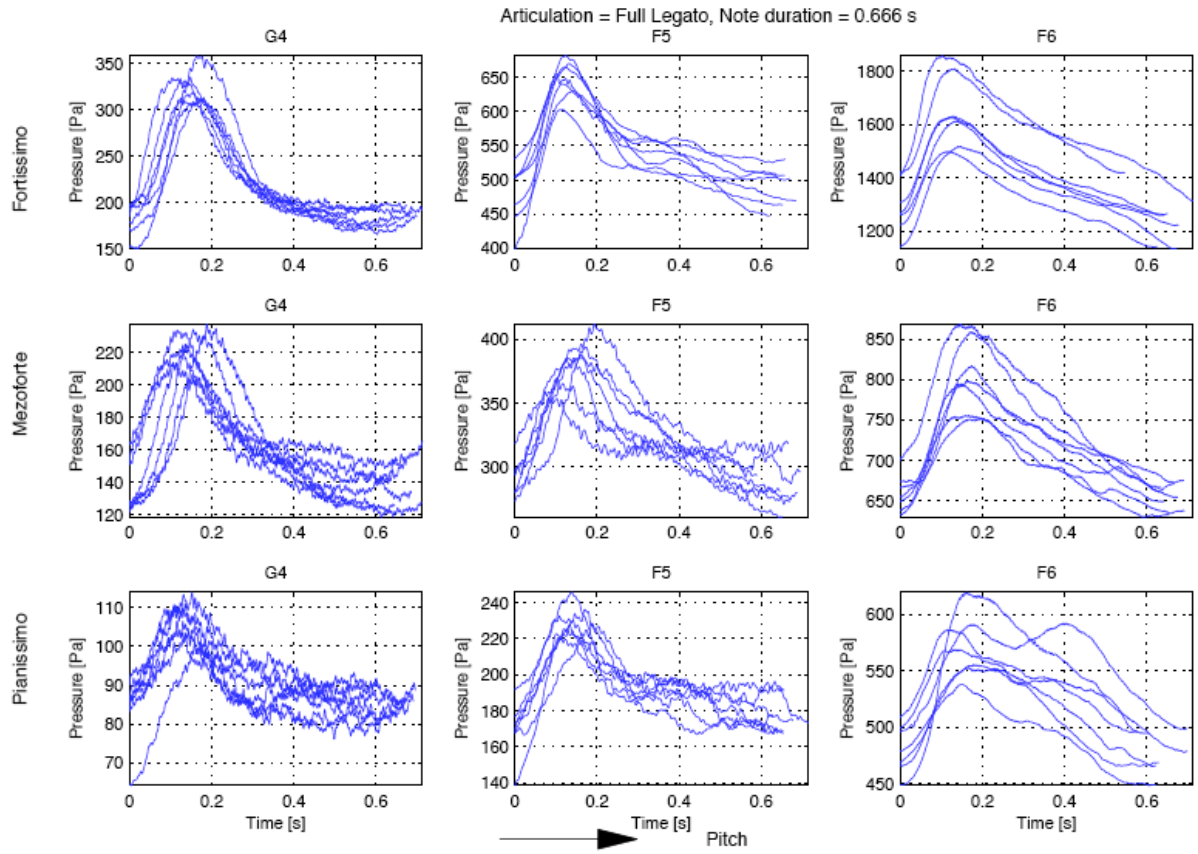

Figure 5.5: Blowing pressure profiles for full legato and note duration of 0.666 seconds: ZOOM.

the different curves for each point of the analysis space. Moreover, it can also be observed that each single curve does not start nor end at zero Pascals, i.e. the pressure level never reaches the zero and the curve has a slight peak for each note. This fact is very characteristic of the legato articulation where the note is played during the whole nominal note duration without any silence, i.e. the performer never stops blowing and this is clearly represented in the pressure curves. Moreover, the attack and decrease slopes are very soft in comparison with other articulations, as it will be shown later.

Another important observation is the ripple of the pressure curves with low pitch and low dynamics (bottom left zone in figure 5.5). The lower the pitch and the dynamics, the stronger is the ripple of the blowing pressure curve. The explanation of this phenomenon deals with the fact that the performer has to blow very soft in the low pitches because the jump from the first acoustic mode of vibration to the second one takes place at very low pressures. Therefore, he has to keep the pressure level very low for not jumping to the next octave. Furthermore, if the performer wants to play a pianissimo note, then this level is even lower. In this case of low pitch and low dynamics, the ripple of the blowing pressure deals with the difficulty of the performer to keep a very low pressure level without jumping to the second mode of vibration of the recorder.

5.2.2 Legato

The legato is articulated with the syllable sequence: "LU-RU-LU-RU". By definition, in the legato each note is attached to the following one and the note duration equals the nominal duration written in the score. Therefore, the blowing pressure profiles of this articulation have always a constant pressure level with a short cut between one note and the following. Figure 5.6 shows the three pitches selected for the representation, i.e. G4, G5 and D6, and the three dynamics (pp, mf and ff). As it was commented with the full legato, the average level of the blowing pressure profiles increases with increasing pitch and with increasing dynamics as well.

In order to appreciate the small details of these legato curves, figure 5.7 shows a zoom from the previous figure. The start and end pressure level for each single note is not zero Pascals, i.e. the pressure level never reaches the zero Pascals as it happened with the full legato articulation. The shape of each single note resembles the half of a square wave period. The attack and the decrease slopes are very sharp

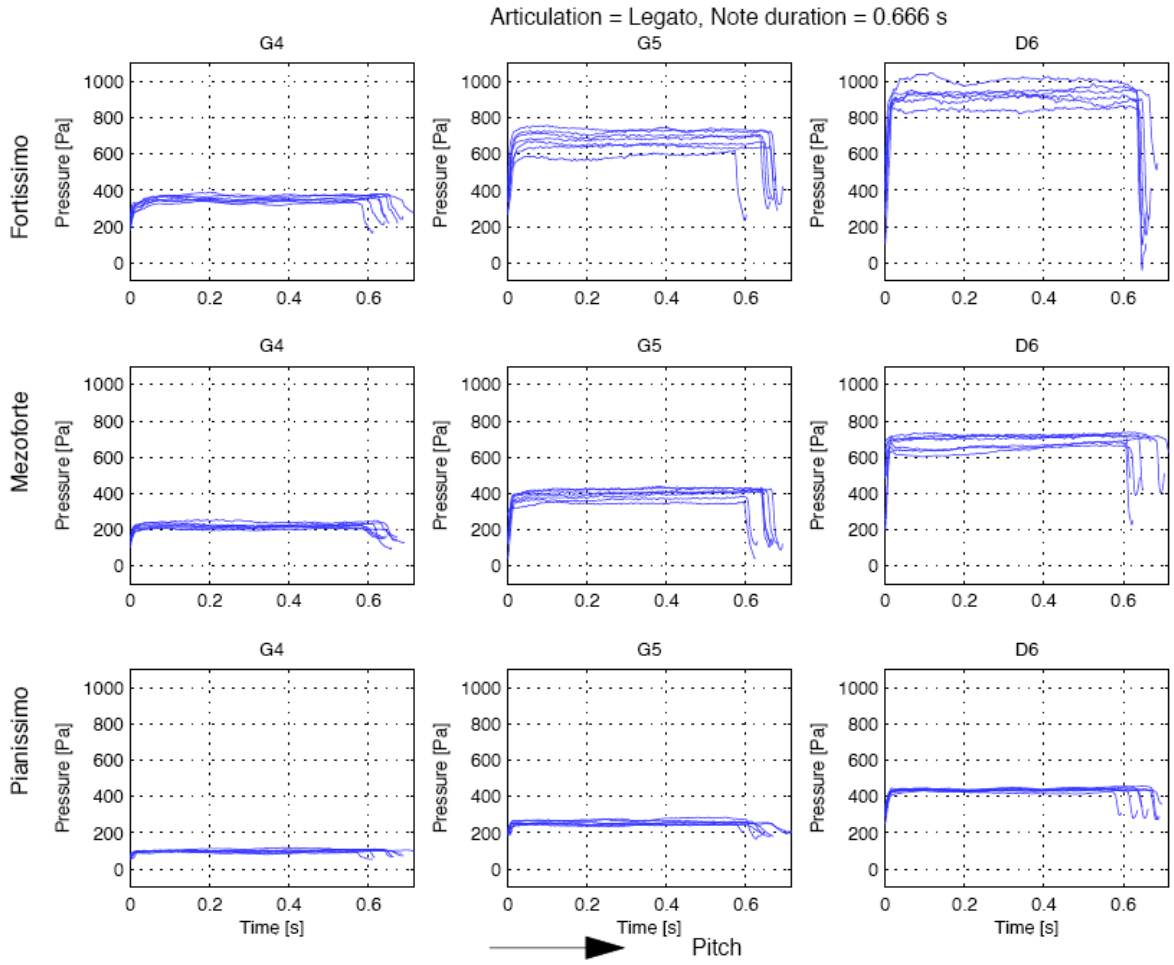

Figure 5.6: Blowing pressure profiles for legato and note duration of 0.666 seconds.

and the steady state covers almost the whole note duration.

Furthermore, the ripple in the curves with low pitches and low dynamics is more pronounced than in the full legato. In the left column of the figure which corresponds with the three dynamics of the pitch G4, it can be observed that the ripple decreases with increasing dynamics. As it happens in the case of the full legato, the main reason for this ripple is the very low pressure levels that the performer must achieve for not overblowing the note at the very low pitches. It is extremely difficult to keep constant a very low pressure level and therefore the average level of the pressure oscillates.

5.2.3 Soft staccato

The soft staccato was achieved with the syllable sequence "DU-GU-DU-GU". The staccato technique means that the notes are detached one from the following one. Therefore, some performers shorten the nominal note duration and introduce a silence between the notes. This is the case of the musician Joan Izquierdo. Therefore, in the soft staccato and the staccato articulations, the blowing pressure profiles start and end at zero pascals of pressure, which is shown in figure 5.8. This figure shows 3 pitches (G4, F5 and F6) and the three dynamics. As it happens with the legatos, the average value of the blowing pressure increases with increasing pitch and dynamics.

If we zoom in each diagram, then we can appreciate some details as it is shown in figure 5.9. The shape of this articulation is completely different than the legatos. The attack is very sharp followed by a short steady state and a fast decrease. The steady state of the note is very short in comparison to the note duration, i.e. it is about $\frac{1}{3}$ of the note duration, but we will come back on this fact later in this chapter. The ripple that has been observed in the legatos is also present in this articulation, although it is very slight. It can be only appreciated in the pianissimo

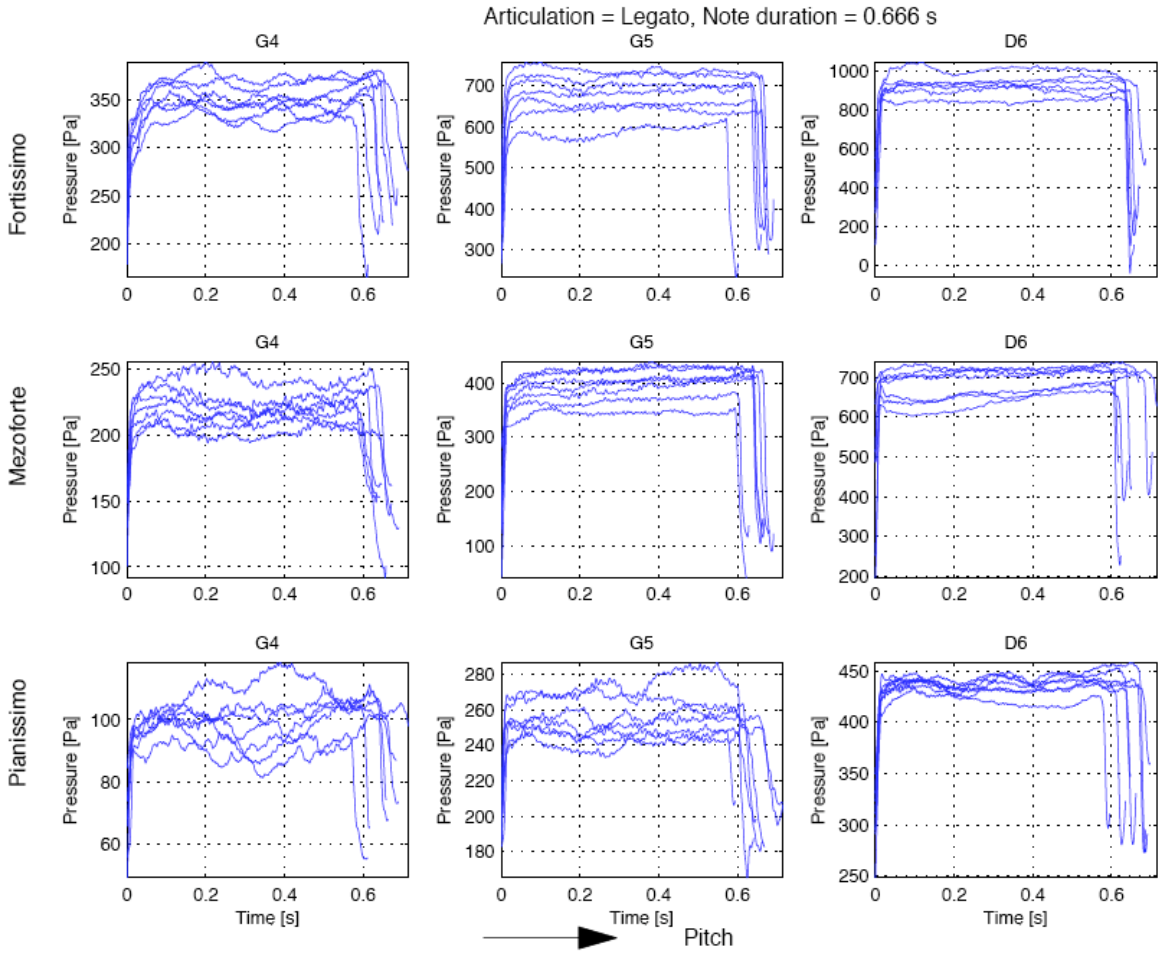

Figure 5.7: Blowing pressure profiles for legato and note duration of 0.666 seconds: ZOOM.

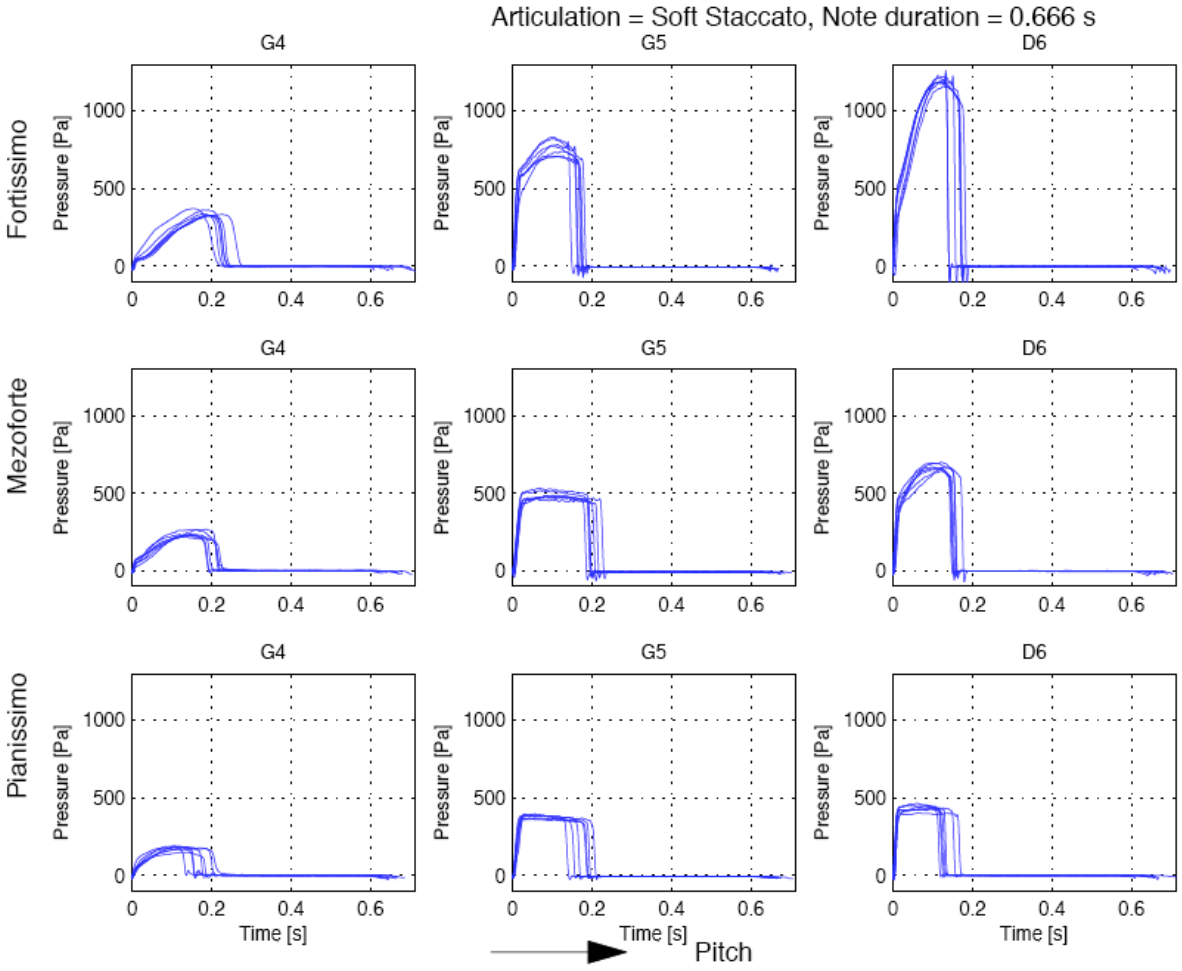

Figure 5.8: Blowing pressure profiles for soft staccato and note duration of 0.666 seconds.

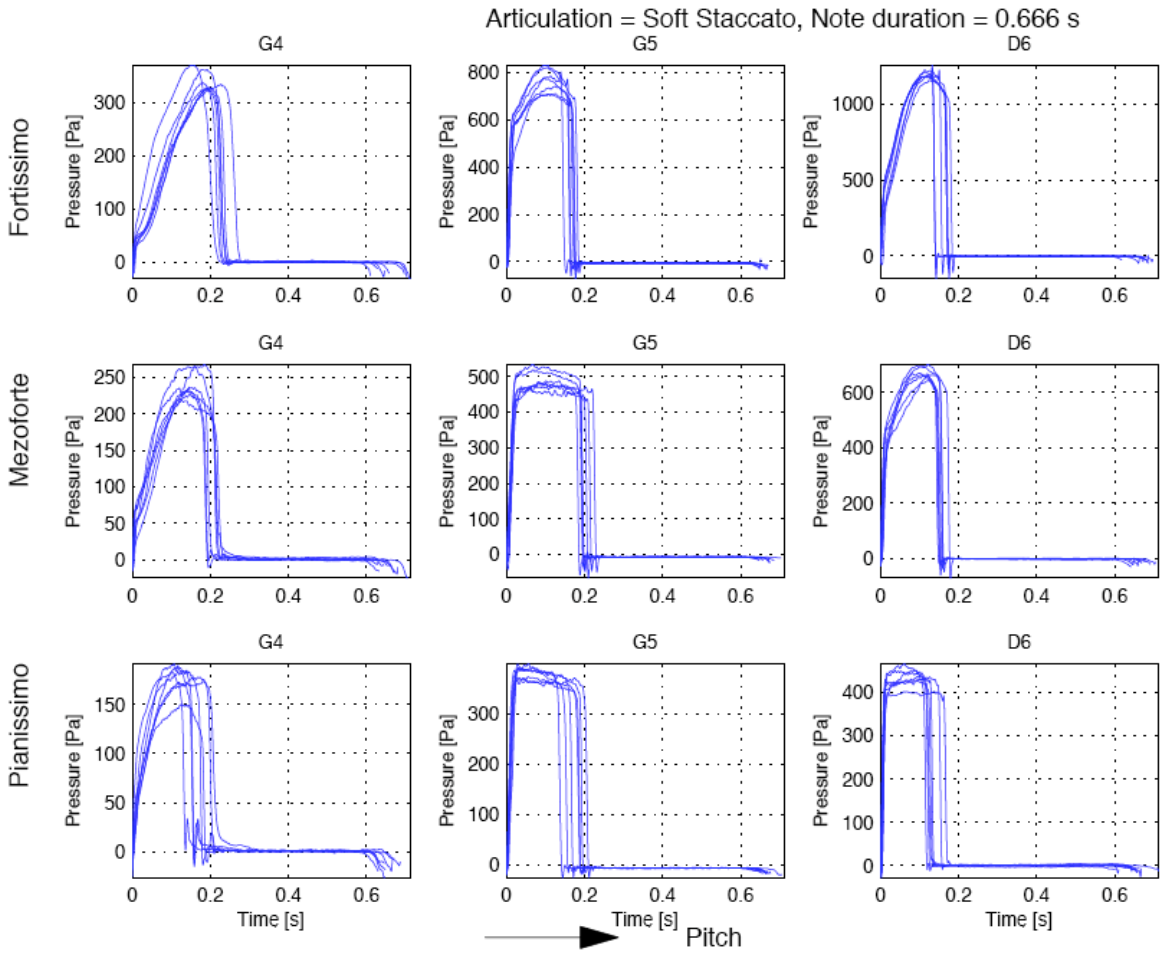

Figure 5.9: Blowing pressure profiles for soft staccato and note duration of 0.666 seconds: ZOOM.

dynamics of the low pitches and a little bit in the mezzo-forte dynamics. Since the effective note duration is very sort, the ripple has not time to develop fully as it happens in the legato.

5.2.4 Staccato

Finally, the last articulation was performed with the syllable sequence "TU-KU-TU-KU". It is the strongest and fastest articulation from all the above. It has the same properties as the soft staccato with regards to note duration and to the average pressure value with increasing pitch and dynamics. The start and end pressure values are for each single note 0 Pascals. Figure 5.10 shows all these details for the staccato notes. A closer look of these pressure profiles is shown in figure 5.11.

Basically, this articulation is very similar to the soft staccato. The attacks and release of each note are sharper than in the soft staccato, as it will be shown later in this chapter. Moreover, the effective note duration with regards to the nominal note duration is shorter than in the soft staccato. The ripple seen in all the other articulations is almost not appreciable in this case because the effective note duration is really short.

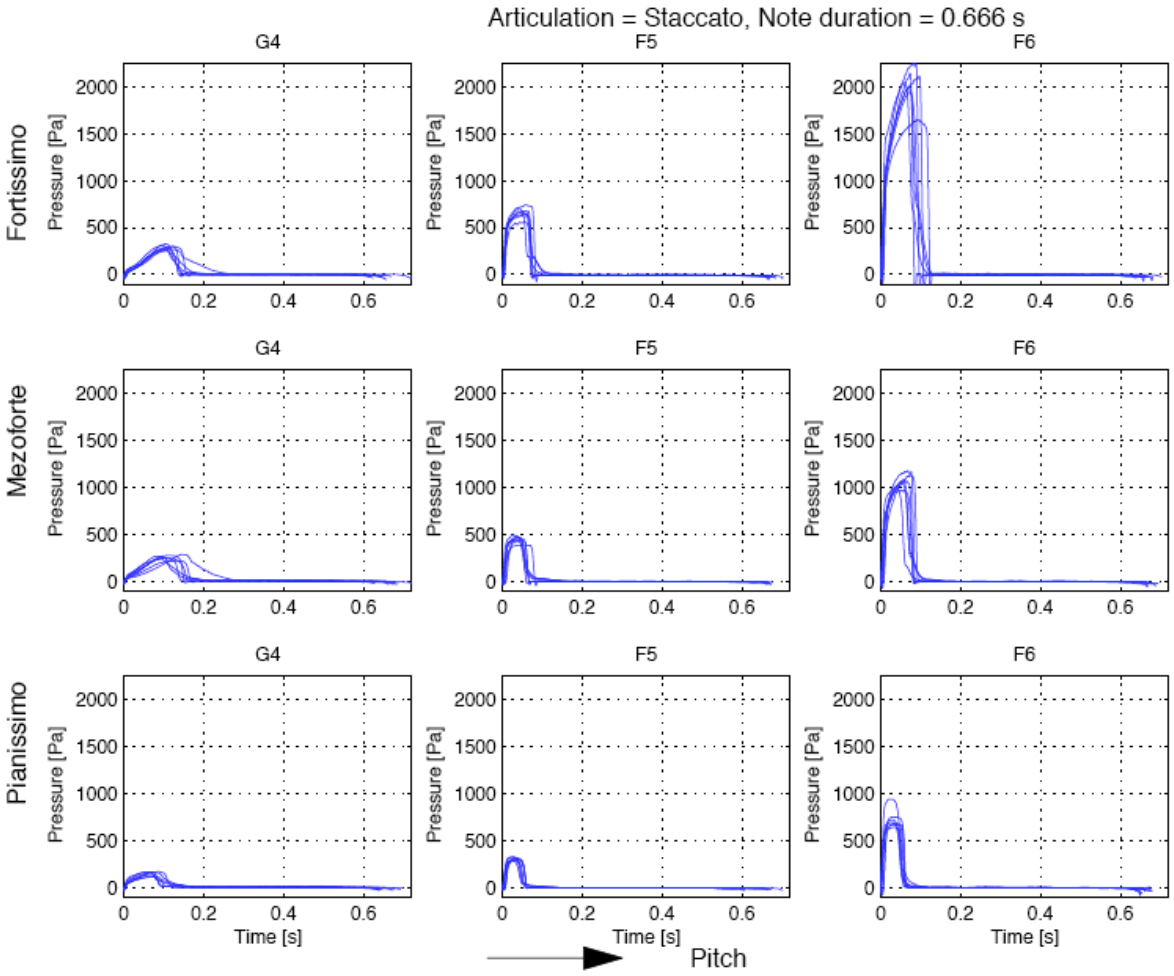

Figure 5.10: Blowing pressure profiles for staccato and note duration of 0.666 seconds.

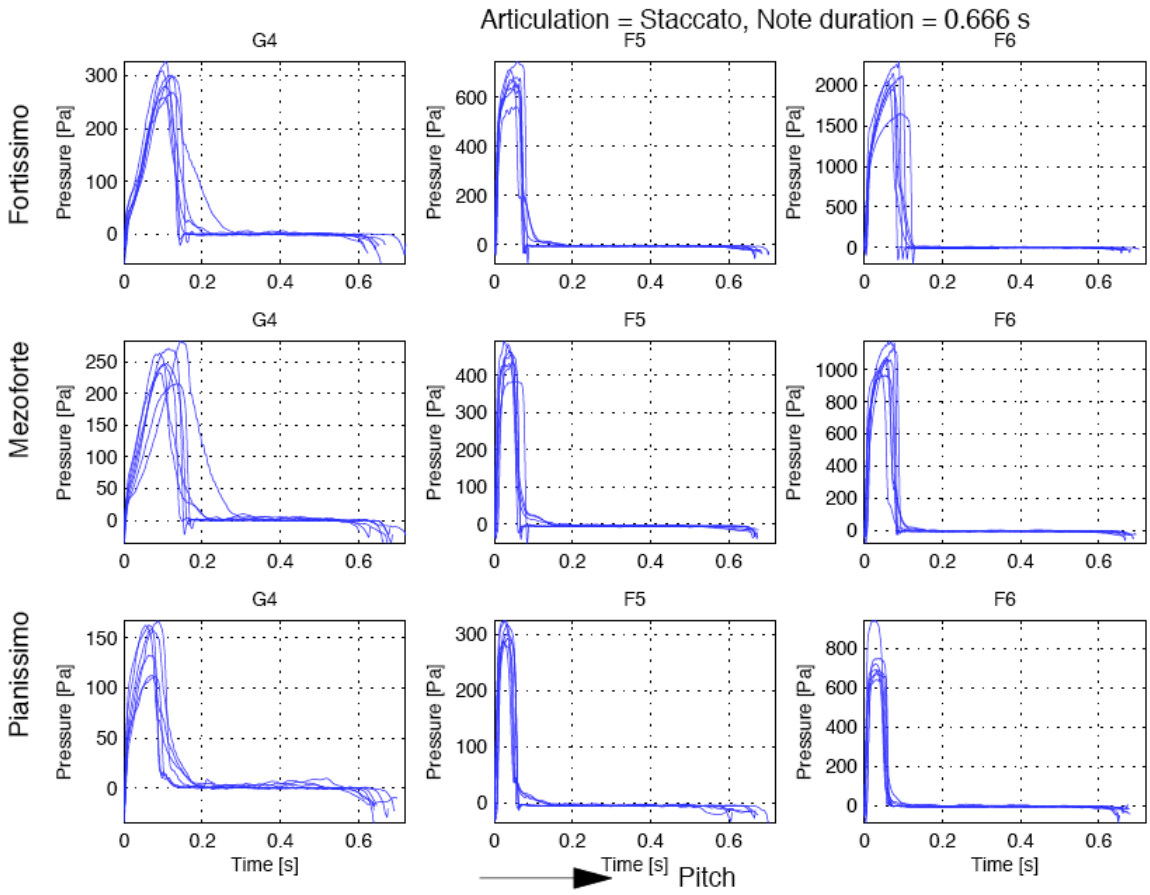

Figure 5.11: Blowing pressure profiles for staccato and note duration of 0.666 seconds: ZOOM.

5.3 Fingering, dynamics and articulation

In order to summarize the results shown in the previous section, the relationship between the blowing pressure, the pitch (i.e. performed fingering) and the dynamics will be shown in the following figures. These results are a reproduction of the results by Fletcher and Montgermont with the transverse flute [10, 15], where they studied the relationship between blowing pressure, pitch and dynamics (see Chapter 2). In this work, we have extended these results for different articulations.

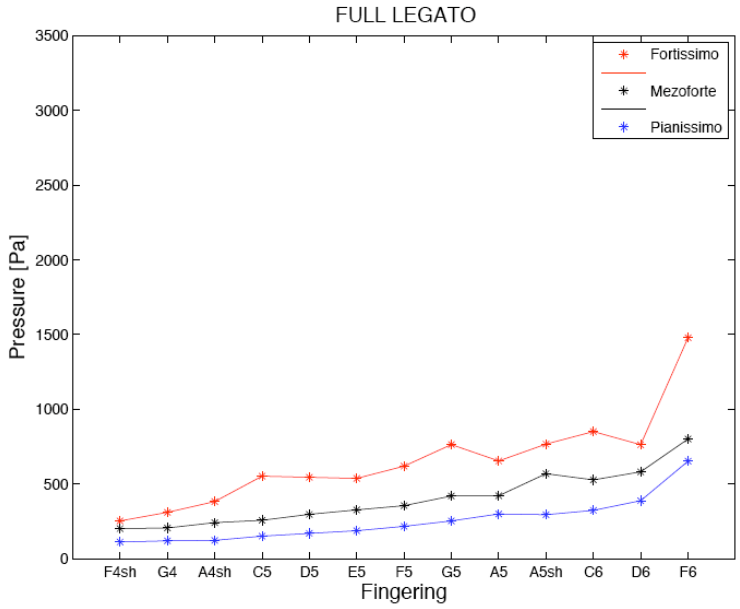

Figure 5.12: Behavior of the blowing pressure with pitch and dynamics: Full legato.

	k	α
pp	0.0135	1.473
mf	0.3235	1.072
ff	0.4712	1.088

Table 5.1: Coefficients of the power relationship between blowing pressure and pitch: Full legato

Figure 5.12 shows how the blowing pressure evolves when the pitch is increasing for three different dynamics. The blowing pressure for fortissimo is higher than the one for mezzo-forte and the pressure for mezzo-forte is higher than the one for pianissimo, as expected. Moreover, the blowing pressure increases with the pitch (i.e. performed fingering). This behavior keeps the same for the different articulations, although there are small changes in the slopes of the curves. If one tries to quantify the relationship between blowing pressure and pitch, one could follow the approach proposed by Bak [14]. He claims that these entities follow a

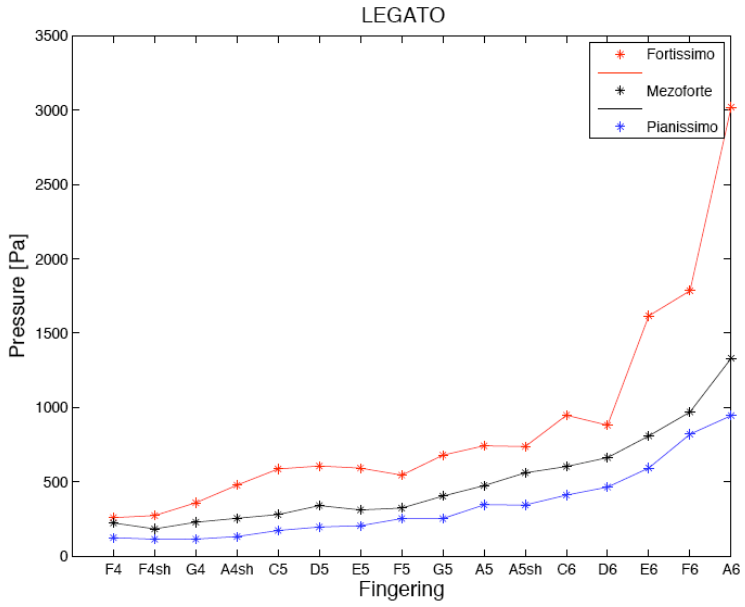

Figure 5.13: Behavior of the blowing pressure with pitch and dynamics: Legato.

	k	α
pp	0.0127	1.504
mf	0.0584	1.337
ff	0.0038	1.807

Table 5.2: Coefficients of the power relationship between blowing pressure and pitch: Legato.

power relationship of the form:

$$P = k \times f^{\alpha} \tag{5.1}$$

By means of curve fitting techniques the coefficients k and α have been computed for the three different dynamics. The results are shown in table 5.1. For the case of the full legato, one can observe that the coefficient k increases with increasing dynamics and the coefficinet α decreases with increasing dynamics. The same approach has been performed on the other three articulations. Figures 5.13, 5.14 and 5.15 show the behavior of the blowing pressure for the articulations legato, soft staccato and staccato, respectively. It is very similar for all articulations, although

	k	α
pp	0.0012	1.847
mf	0.0009	1.945
ff	0.0012	1.961

Table 5.3: Coefficients of the power relationship between blowing pressure and pitch: soft Staccato.

	k	α
pp	0.0131	1.488
mf	0.0026	1.795
ff	0.0009	2.018

Table 5.4: Coefficients of the power relationship between blowing pressure and pitch: Staccato.

there are slight differences that are quantified with the coefficient values in tables 5.2, 5.3 and 5.4.

The case of the legato (see table 5.2) is not very well characterized with these coefficients because the values for k and α do not follow any specific trend while increasing the dynamics. However, the staccatos work much better. In the case of the soft staccato (table 5.3), the coeffcient α increases with increasing dynamics, but the coefficient k oscillates. In the case of the staccato (table 5.4), α increases with increasing dynamics like the soft staccato and k decreases. The behavior of these coefficients in the staccato seems to be completely the opposite as in the full legato. In the staccato articulation, when the dynamics are increased the power coefficient α also increases, which is the one that dominates the relationship. But in the full legato articulation, the power coefficient decreases with increasing dynamics. The opposite behavior happens with the proportional coefficient k. It decreases in the staccatos and it increases in the full legatos.

The computation of these coefficients sheds some light on the slight differences that exist between each articulation and its relationship with the pitch and the dynamics.

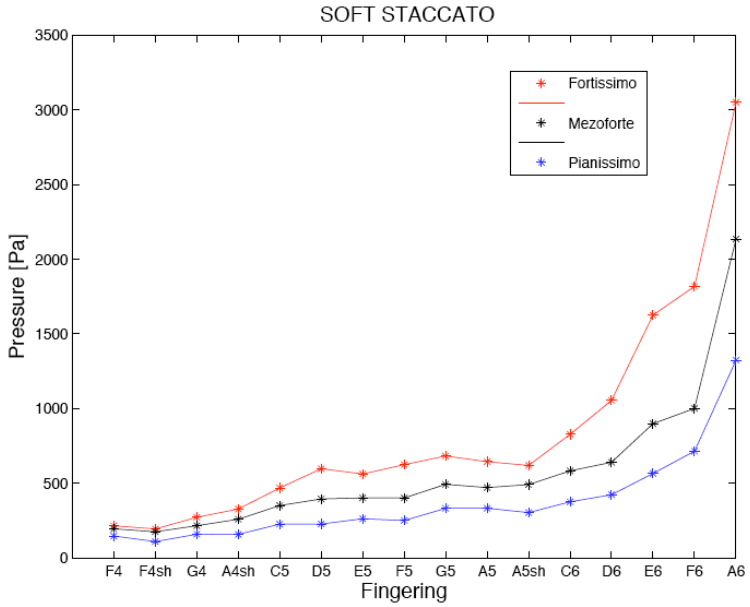

Figure 5.14: Behavior of the blowing pressure with pitch and dynamics: soft Staccato.

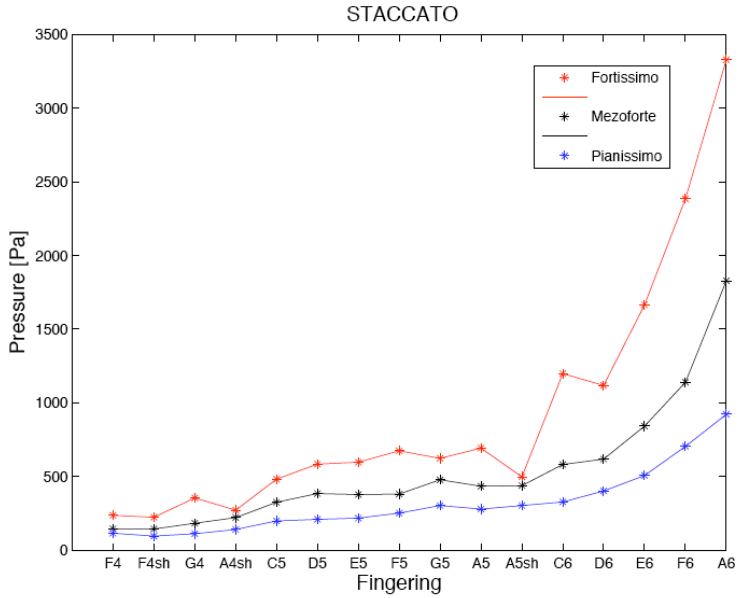

Figure 5.15: Behavior of the blowing pressure with pitch and dynamics: Staccato.

5.4 Attack and rise time

In section 5.1 the time t_1 was defined as the time between the beginning of the note and the pressure level that reached 95% of the $\triangle P$ in a certain note. This time is going to be defined as attack or rise time depending on the articulation that one is dealing with. For the legatos, t_1 is going to be a rise time because the pressure curves of these articulations never start from zero Pascals, but from a certain value depending on the dynamics and the pitch. Therefore, strictly speaking, the time t_1 in the legatos is just a rise time, but not an attack time, since it does not start from zero. However, for the staccatos this time can be considered an attack time because it coincides with the time that takes to go from 0 Pascals to 95% of the maximum pressure. That is the reason why the rise times and attack times are evaluated separately.

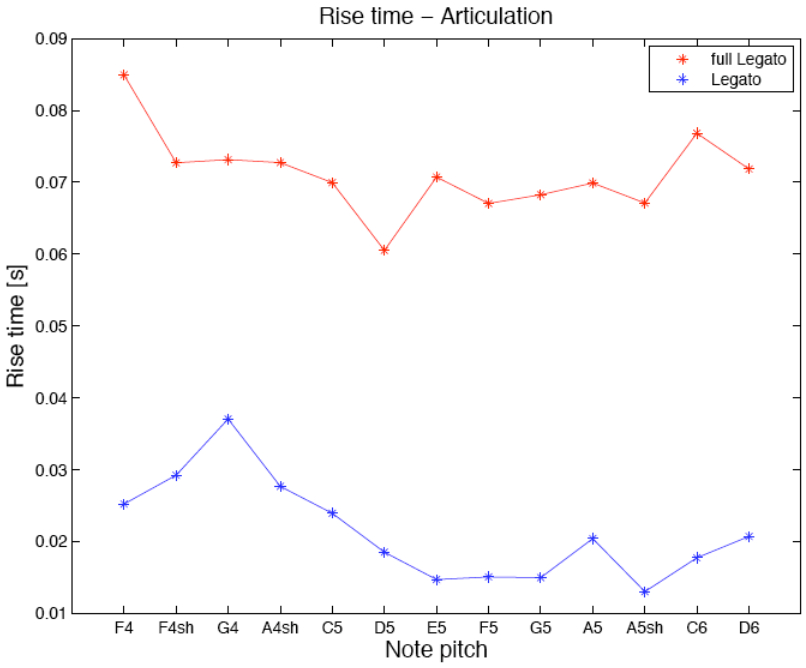

Figure 5.16: Rise time of the legatos for different pitches.

Figure 5.16 shows the evolution of the rise times for the legatos with different pitches. The values for the full legato (in red) lie around 70 ms for almost every pitch, i.e. the rise times is approximately constant for all pitches. The legato articulation (in blue) shows a different behavior. First, the rise times are shorter than the full legato for all pitches, that is to say, the full legato articulation articulates the notes slower than the legato articulation. Moreover, in the legato the low pitches have a higher rise time than the high pitches, i.e. the rise time decreases for higher pitches in the case of the legato.

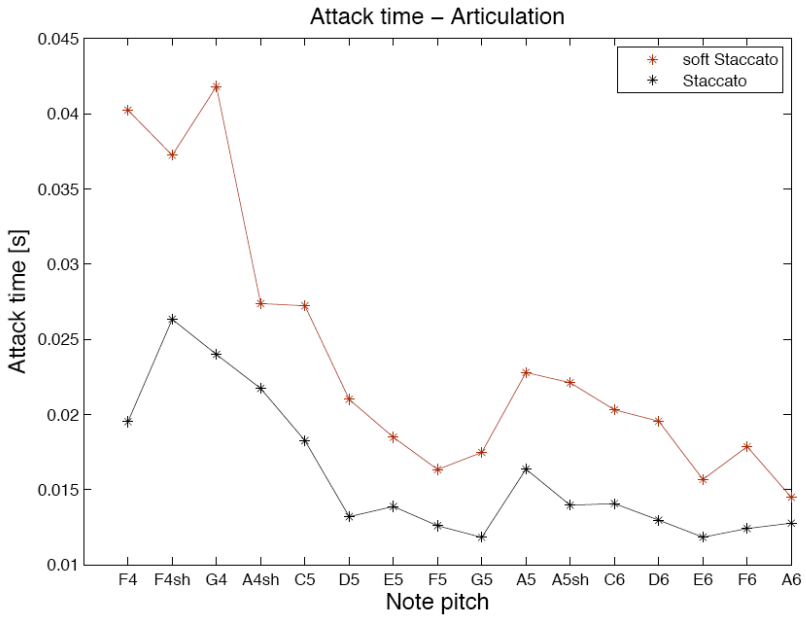

Figure 5.17: Attack time of the staccatos for different pitches.

In the case of the staccatos (see figure 5.17), the behavior of the attack time is similar to the legato articulation, i.e. it decreases with higher pitches for both articulations. The soft staccato articulation (in red) is slower than the staccato (in black). Furthermore, there are some peaks that do not fulfill this decreasing trend. These peaks (F4#, G4, A5 or F6) are probably caused by the design of the contralto recorder used in the database construction.

As a general rule for the articulations, one could conclude that the attack times decrease for higher pitches in the case of the staccatos and the rise times decrease in the case of the legato. The full legato seems to be approximately constant for all pitches.

The behavior of the attack and rise times with regard to the note duration has also been studied. As it is shown in figure 5.18, the shorter the note duration, the lower is the attack or rise time, i.e. the shorter notes are articulated faster than the slower notes for different dynamics and articulations. The reason that explains this phenomenon is the fact that if one wants to reach a certain pitch with the recorder then he has to blow with a certain pressure that is independent from the note duration. Therefore, if one wants to make a very short note, then he has a very short time to achieve the same pressure level compared to the time needed for a longer note. In order to achieve the same level in less time, the rise time has to be necessarily shorter for short notes than for long notes. This can be appreciated in figure 5.18, where the evolution of the attack/rise times for the five note durations from the database are represented with regard to different pitches. The dark blue line represents the note duration of 125 milliseconds. Its attack/rise time is around 20 milliseconds while the attack/rise time for the note duration of 666 milliseconds (cyan blue) is around 40 and 80 milliseconds. In the middle of these extreme values lie the rest of the note durations, but they all accomplish the rule commented above: for shorter note durations, the attack/rise time is shorter.

Finally, the behavior of the attack and rise times has been studied for different dynamics. There was not found any correlations with the rest of the analysis dimensions, i.e. articulations, note durations or pitch. The behavior of the attack and rise times seems to be independent from the dynamics.

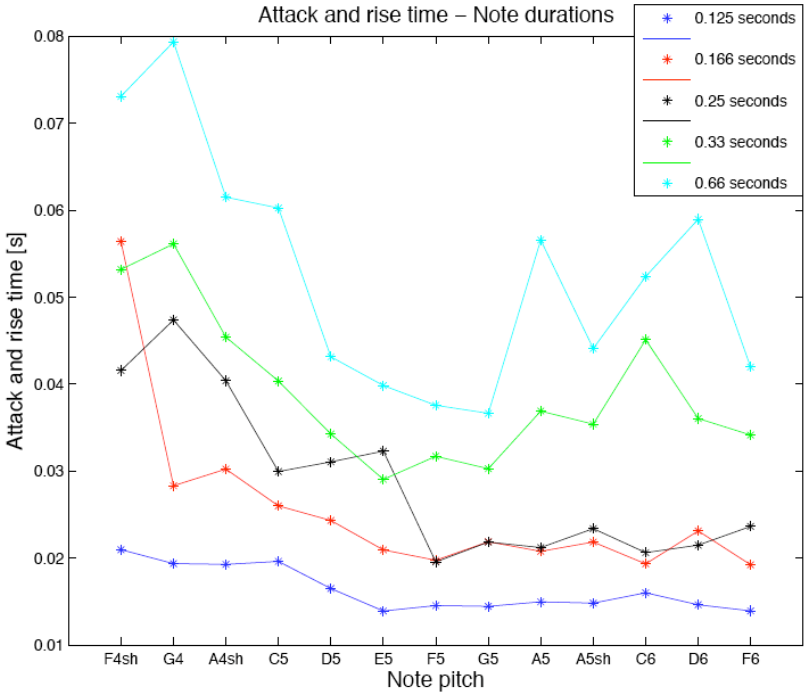

Figure 5.18: Attack and rise times behavior with different note durations.

5.5 Attack slope

In section 5.1 the attack slope was defined as the slope of the linear regression of the pressure curve between the beginning of each note and the time t_1. This feature describes how fast the blowing pressure reaches the steady state. First, the dependency of this feature on the performed articulation is studied in depth. From the previous section it came out that the shortest attack time was achieved with the staccato articulation. Therefore, the expectations for the highest slopes are also with the staccatos. Figure 5.19 shows how the attack slope evolves for the different articulations when the pitch is increased. The dots represent the mean values of the attack slope for each pitch and each articulation. As it was expected, the staccato articulation reaches the highest slopes, followed by the soft staccato, the legato and

finally the full legato. Moreover, for all articulations except the full legato, the increasing pitch results in a higher slope. This result is coherent with the behavior of the attack/rise time, e.g. the rise time of the full legato does not substantially change with the pitch and the same happens with the attack slope.

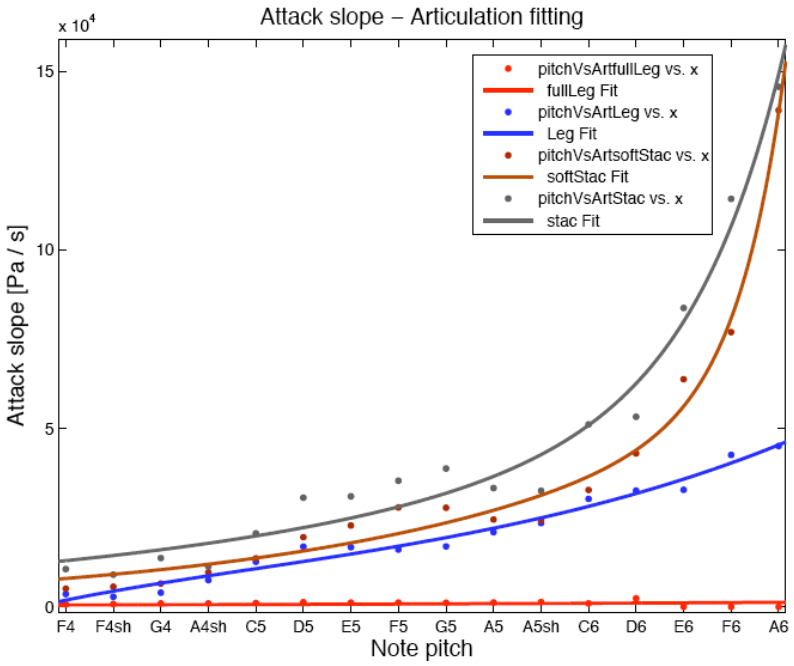

Figure 5.19: Attack slope values and curve fitting for different articulations: dots = average values, line = curve fitting.

As a way to model the behavior of the attack slope, the mean values were approximated by means of curve fitting techniques through exponential curves. The resulting curves are shown in figure 5.19 as lines, one for each articulation. This exponential regression could be a temporal model of the attack slope depending on the pitch and the articulation. Obviously, the modeling of the full legato articulation through this exponential curve does not return satisfactory results because it does not change with the increasing pitch. The rest of the articulations behave relatively well through this approach.

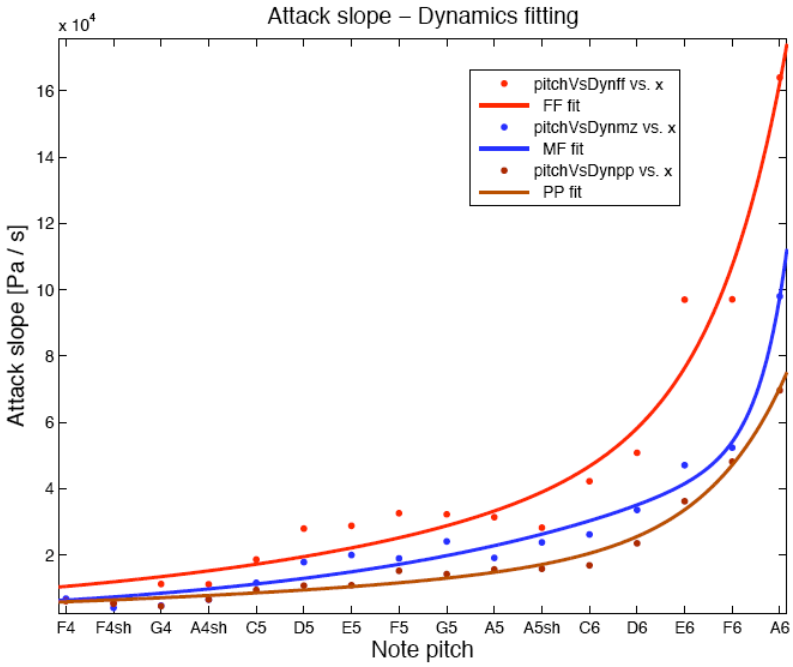

Figure 5.20: Attack slope values and curve fitting for different dynamics: dots = average values, line = curve fitting.

The attack slope was also studied with regard to the dynamics applied by the performer. Although the attack/rise times did not provide concluding results for its dependency on the dynamics, the relationship between the attack slope and the dynamics is much more clear. Figure 5.20 shows the average values (dots) of the attack slope for different dynamics. As it is expected, the fortissimo has higher slopes than the pianissimo for all fingerings. Furthermore, when the pitch is increased the attack slope is also increased. In this case, this behavior was also modeled through curve fitting techniques. The exponential curves are shown as colored lines for each dynamic in the figure and they seem to model the attack slope correctly.

Finally, the dependency of the attack slope on the note duration was investigated. There was not any clear trend or correlation between the fingering and the note

duration with regard to this feature. Although the attack/rise time has a clear correlation with the note duration, in the case of the attack slope that relationship is not trivial.

5.6 Steady state to note duration ratio

In section 5.1 two states were defined as features: the steady state and the release state. In this section, the duration of the steady state with regard to the nominal note duration will be studied. The steady ratio is defined as:

$$SR = \frac{Steady\ state\ [s]}{Note\ duration\ [s]} \tag{5.2}$$

This ratio should give an idea of how much time of the nominal note duration the blowing pressure provides its highest level inside a single note. First, it will be analyzed how this ratio changes with the different articulations. In this analysis the full legato articulation has been excluded because the steady state in this articulation does not give meaningful information since the pressure is almost all the time at its highest level. Therefore, the steady ratio should be 100% for all the notes in full legato. For the rest of the articulations, figure 5.21 shows how the ratio evolves for an increasing pitch. It can be appreciated that the legato varies from 60% to 85% while the soft staccato moves around 40% and the staccato changes below 30%. These results are coherent with the articulation shapes that were shown in section 5.2 and they confirm the way the performer plays the staccato, i.e. he shortens the note more than 50% from its nominal duration. Furthermore, the variation of the ratio with the pitch is due to the design of the contralto recorder used specifically for the construction of the database, as it happened wit the observed peaks for the attack time (see section 5.4). This ratio could be used as a valuable classifier in order to differentiate the articulations.

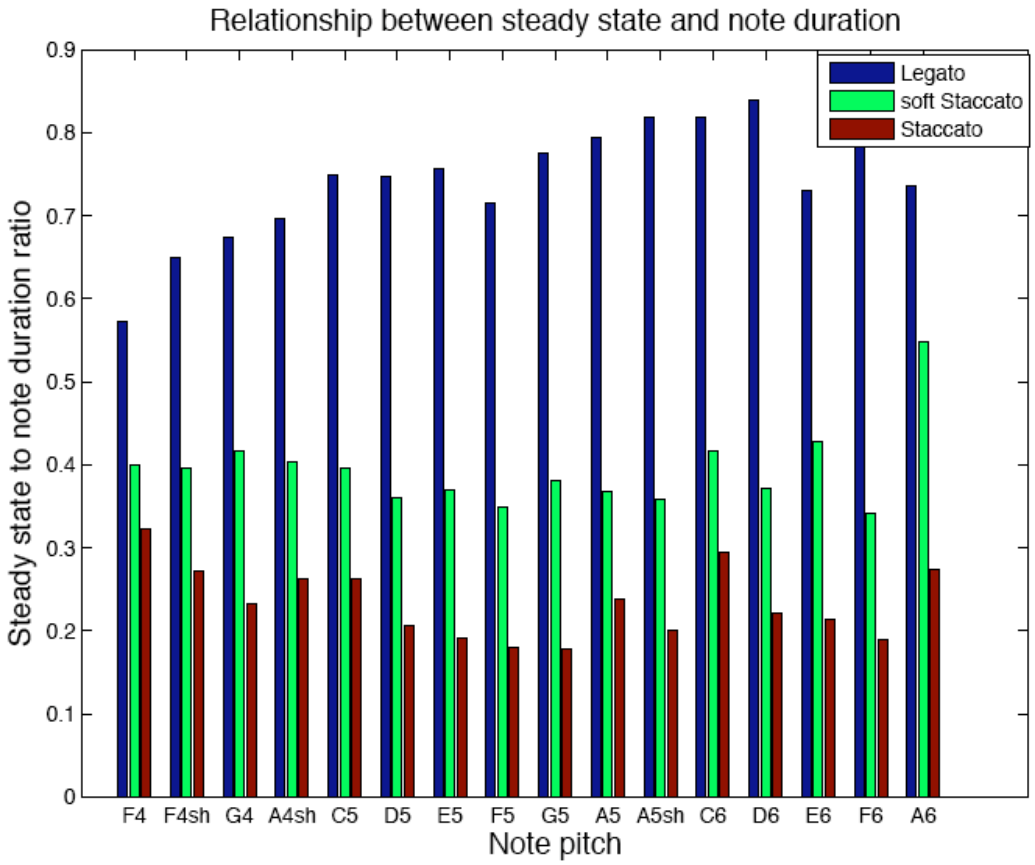

Figure 5.21: Steady ratio behavior for three different articulations.

A second analysis sheds some light on the relationship of the steady ratio with the note duration. For almost every analyzed pitch, the shorter notes provide a higher ratio than the longer notes. Figure 5.22 shows this behavior. Every single pitch presents five bars that correspond to each note duration. It can be observed that for almost all the pitches the ratio decreases from left to right in each pitch, i.e. the shorter notes have a higher steady ratio than the longer notes. The main explanation for this phenomenon is the fact that longer notes have a slower attack/rise time. Therefore, the steady state is smaller in the longer notes because they need more time to reach it. The shorter notes have a faster attack/rise time and they remain

in the steady time more time than the longer notes.

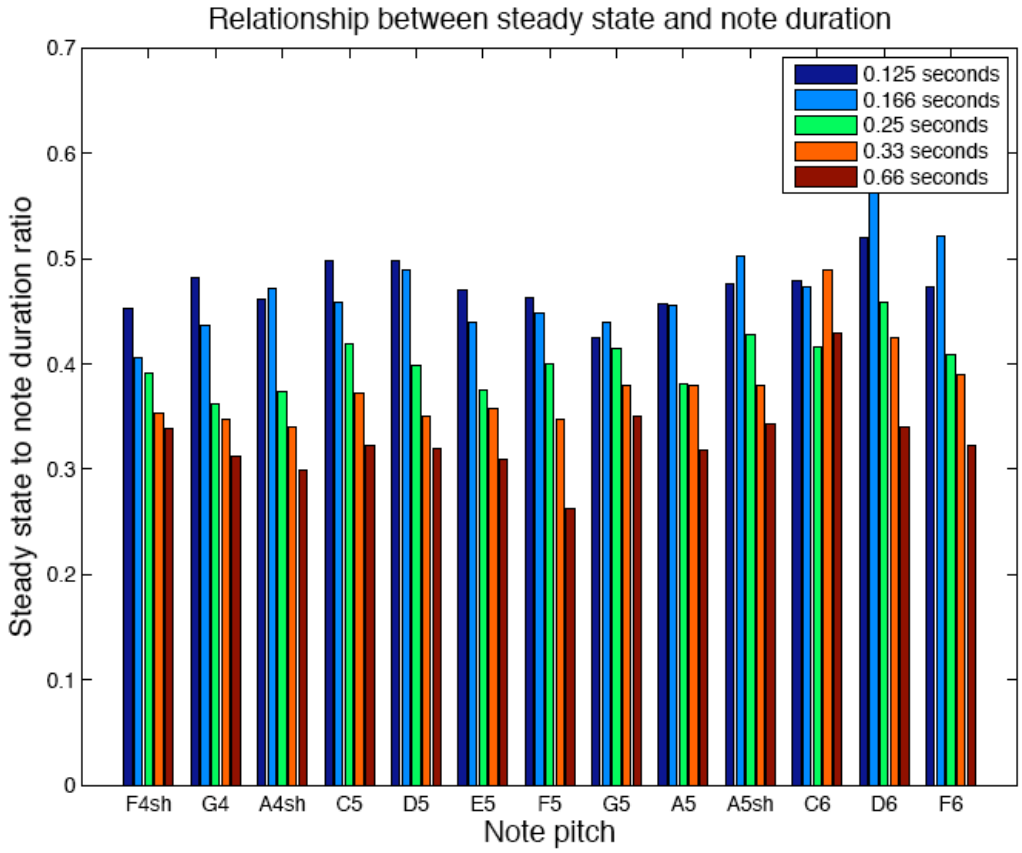

Figure 5.22: Steady ratio behavior for the different note durations.

The last analysis performed on the steady ratio tries to explain its relationship with the dynamics. As it happened with the attack/rise times, there was not found any clear correlation. Therefore, one can conclude that the dynamics do not affect this ratio directly.

Chapter 6

Conclusion

In this chapter, the most important contributions and achievements of this work will be highlighted. Afterwards, a brief look into the next steps that could follow this project is done. These steps go towards the synthesis of performer gestures and sound and, from a more general point of view, towards the construction of a wind synthesizer.

6.1 Summary of contributions

The first and maybe the most important contribution of this work is the construction of the multimodal database with sound and gesture. Many different analysis can be performed on these data, although all recordings have not been analyzed exhaustively because of lack of time. Therefore, it has been a great contribution to create this huge database in order to foment further studies and analysis.

Another important contribution has been the identification and description of a unique shape of the blowing pressure profiles for each articulation. This performance technique is a very expressive tool in the case of the recorder, so it is really important to know how the pressure profile evolves in each articulation in order to synthesize

successfully the performer gestures.

A reproduction and extension of the results by Fletcher and Montgermont [10, 15] in the case of the recorder flute has been achieved. The relationship between dynamics, pitch and blowing pressure has been studied qualitatively for each articulation and it has been quantified through the Bak approach [14]. Small differences between the different articulations were found in this relationship. This fact promotes also the further research in this direction.

A temporal parameterization of the blowing pressure profiles has been done through a set of successfully defined features. The behavior of these features with regard to the analysis dimensions (i.e. pitch, dynamics, articulation and note duration) has been studied comprehensively. These analysis are the starting point in the building of a temporal features model according to the analysis dimensions, i.e. annotations.

Finally, a small subset of features (i.e. attack slope, attack time) has been modeled through curve fitting techniques that could allow the synthesis of performer gestures given a certain point in the analysis space.

6.2 Future work

The following natural step in this work would be to build a linear segment model of the performer gestures that would be able to synthesize a certain blowing pressure profile given a certain point in the analysis space. For that purpose, the different temporal features defined in 5.1 should be modeled comprehensively. After that, the linear segments of each profile could be the following ones: attack slope, steady state, decrease slope and release state.

These linear segments could be improved through a Bezier curve modeling. First, one would need to find the adequate segments and then parametrize the curves. This

latter model would provide a better approach than the linear segments because the Bezier curves are better adjustable.

Once the model is built, one would need to synthesize the gesture with that model and evaluate it. For the evaluation, there are several options that could be developed. The first one is the straight comparison of the recorded pressure profiles with the synthetic ones. The second one would be the synthesis of sound with both gestures. In this case, a physical model of the contralto recorder should be calibrated in order to feed it with the original and synthetic gestures. In this manner, one could compare the sound synthesis created by the recorder gestures and by the synthetic ones. A user evaluation should test if any difference is found.

Finally, the recorded musical pieces (see 4.2.3) could also be used as an evaluation set for the built model. Although the dynamics and the articulation are not annotated and they change along the piece, one could try different configurations with the analysis dimensions (i.e. annotations) in order to match the original recordings. Furthermore, a continuation in the analysis of the database is highly encouraged in order to obtain a better and more realistic wind synthesizer in the future term.

Appendix A

Technical details of the Prototype II

The modified contralto recorder designed by the luthier Josep Tubau is described in detail in this appendix. First, the diagrams with the design of the mechanical modifications are shown. After that, the details of the sensor connections are explained as well as certain details of the design such as the water tramp.

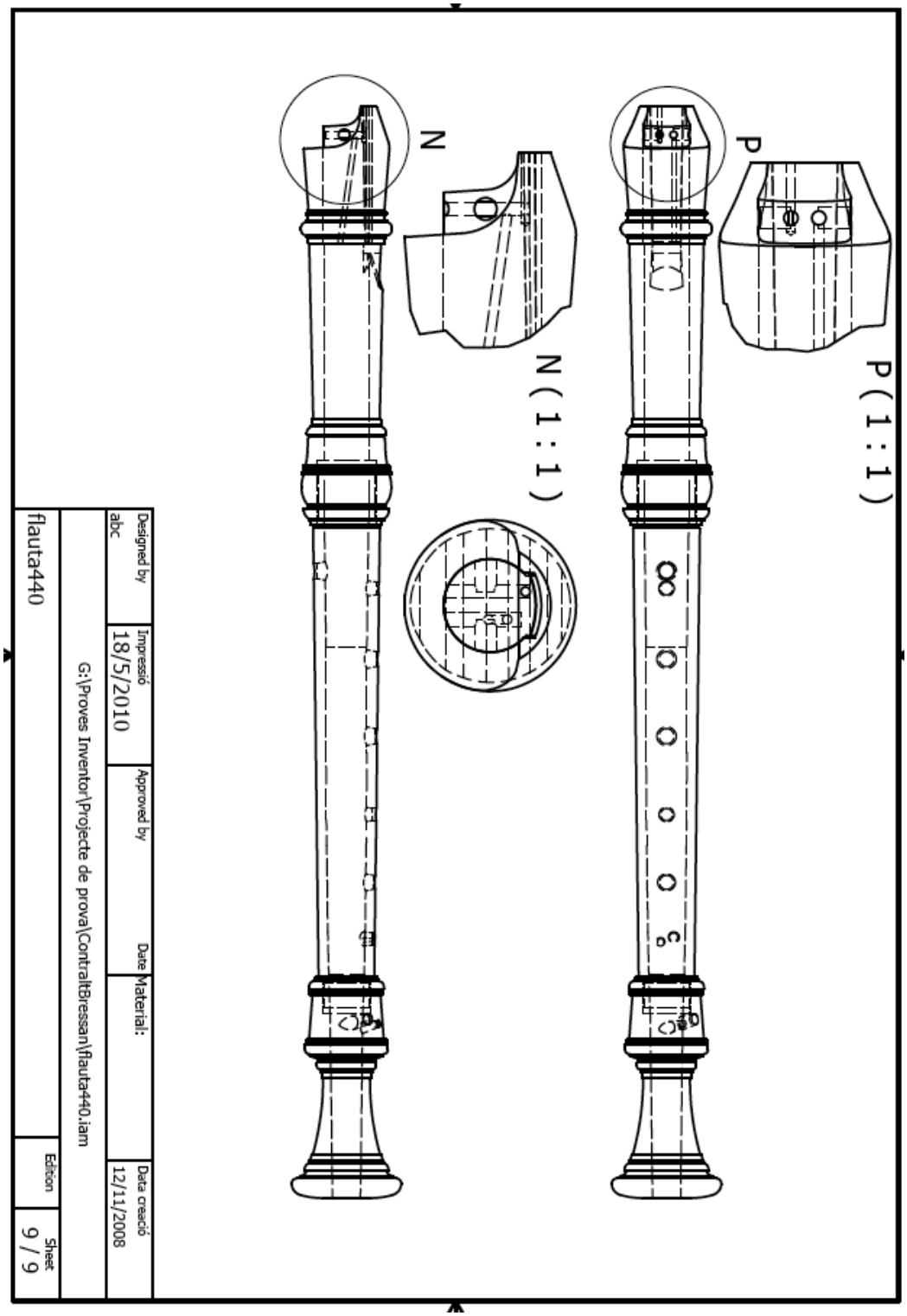

Figure A.1: General scheme of the whole contralto recorder.

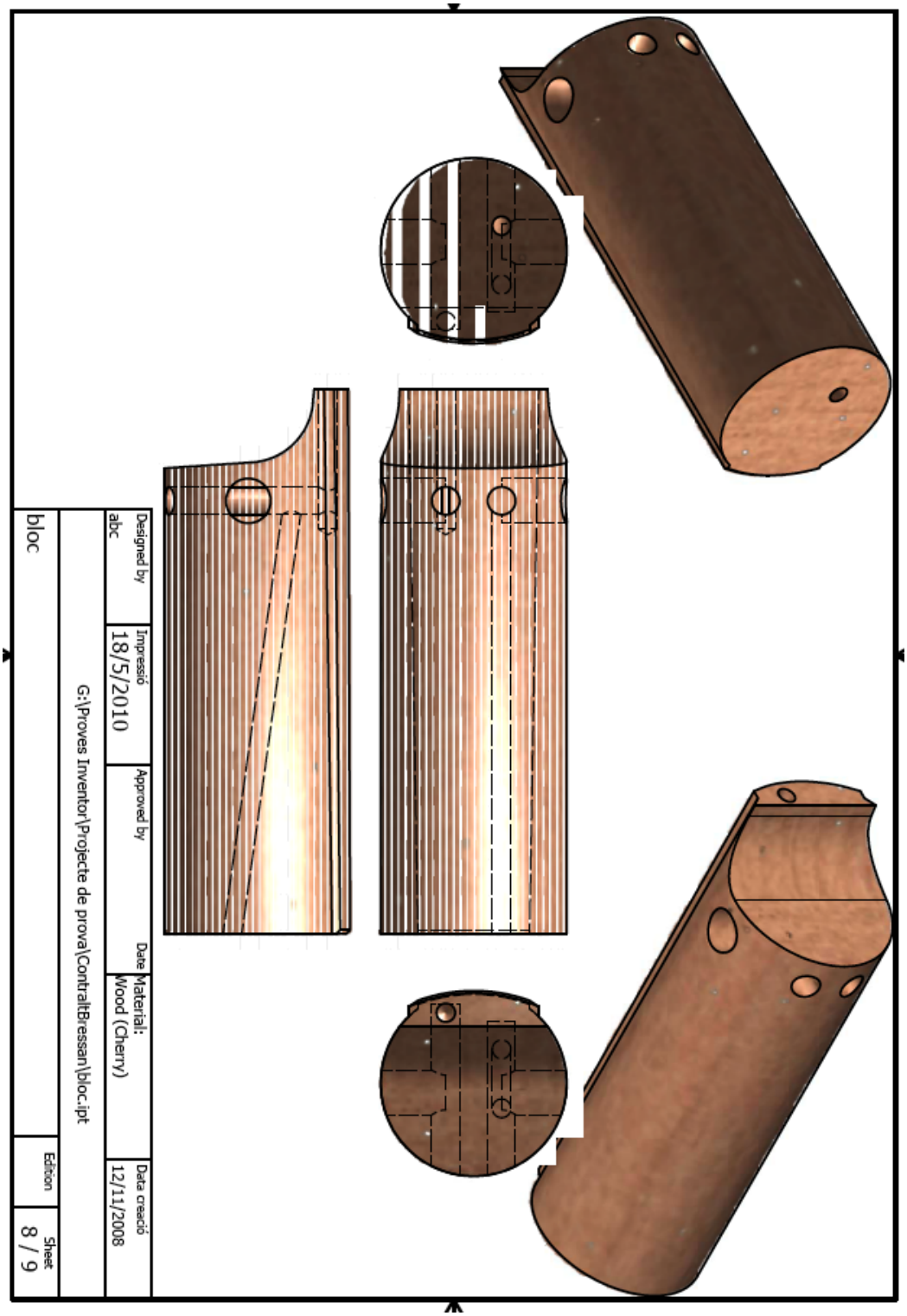

Figure A.2: Detail of the mechanical holes in the mouthpiece.

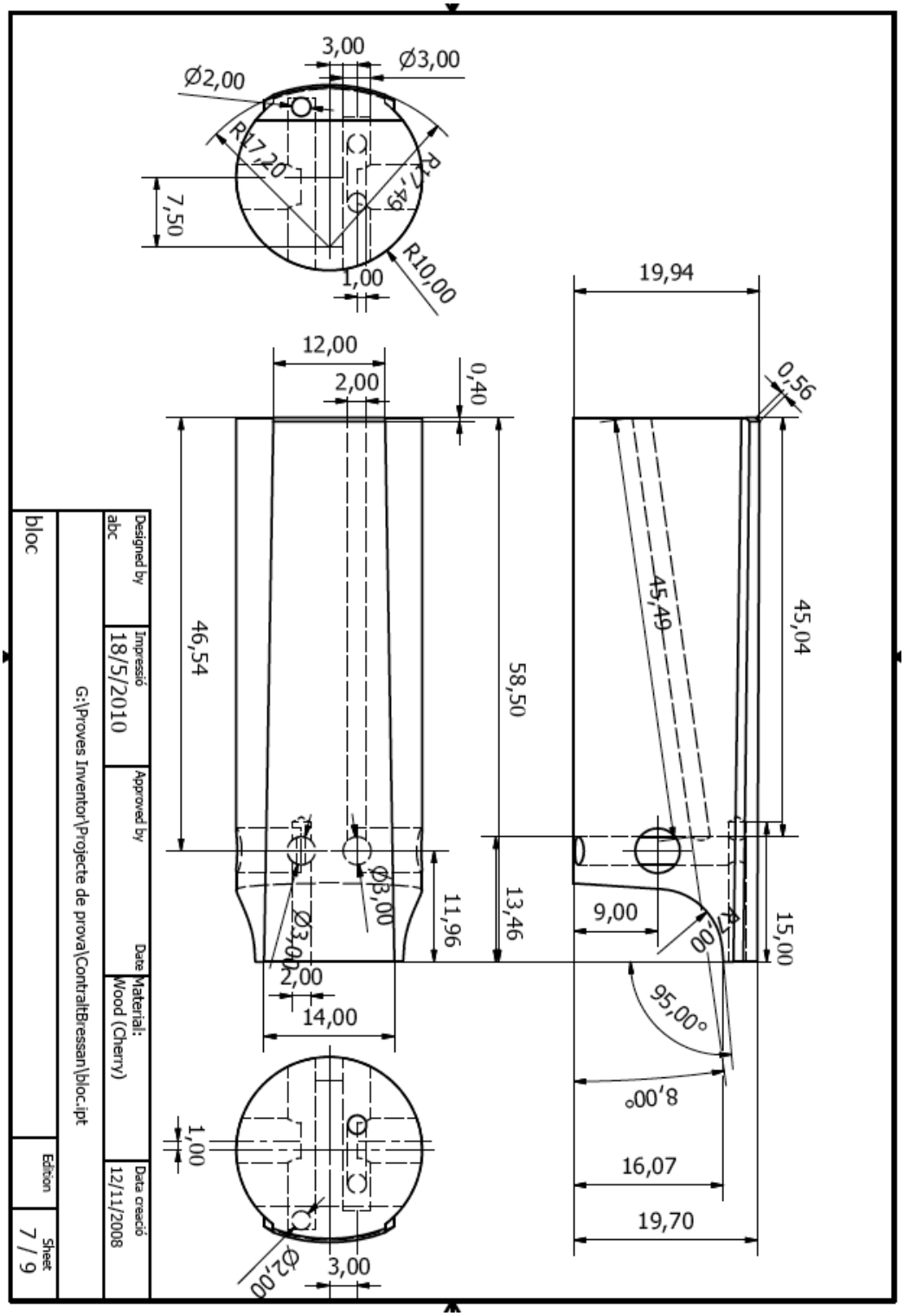

Figure A.3: Dimensions of the mechanical modifications.

Figure A.4: Blowing pressure: sensor connection.

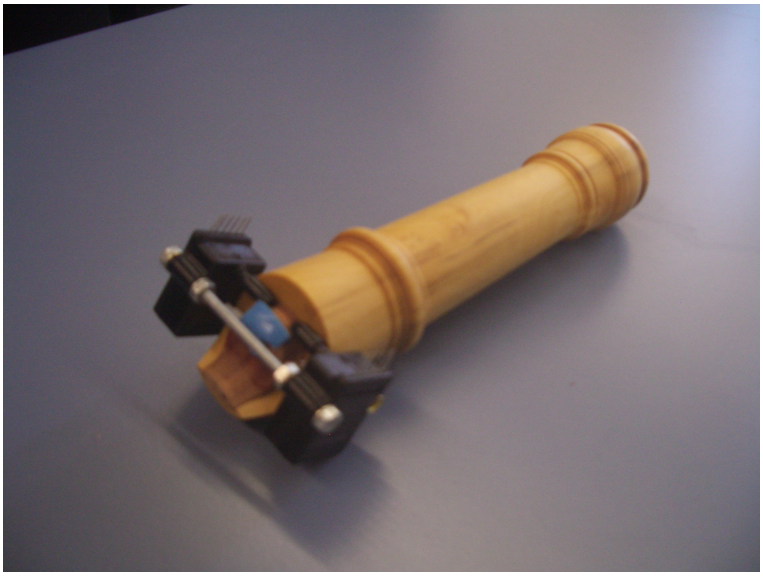

Figure A.5: Connection of the two pressure sensors.

Figure A.6: Detail of the water tramp.

Appendix B

Scores of the musical pieces

The scores of the five musical pieces are shown in this appendix. They were extracted from the book "Preludes and Voluntaries".

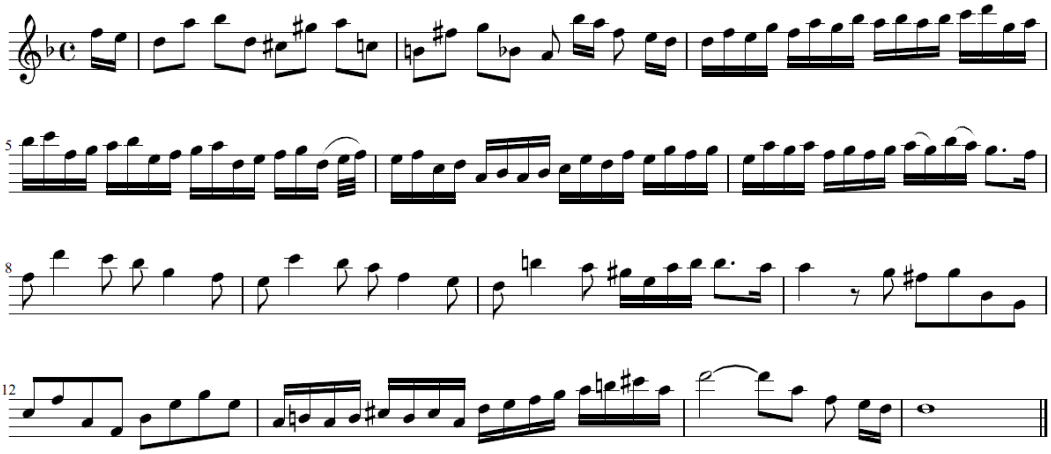

Figure B.1: Score of the first piece. Piece 1 of the book.

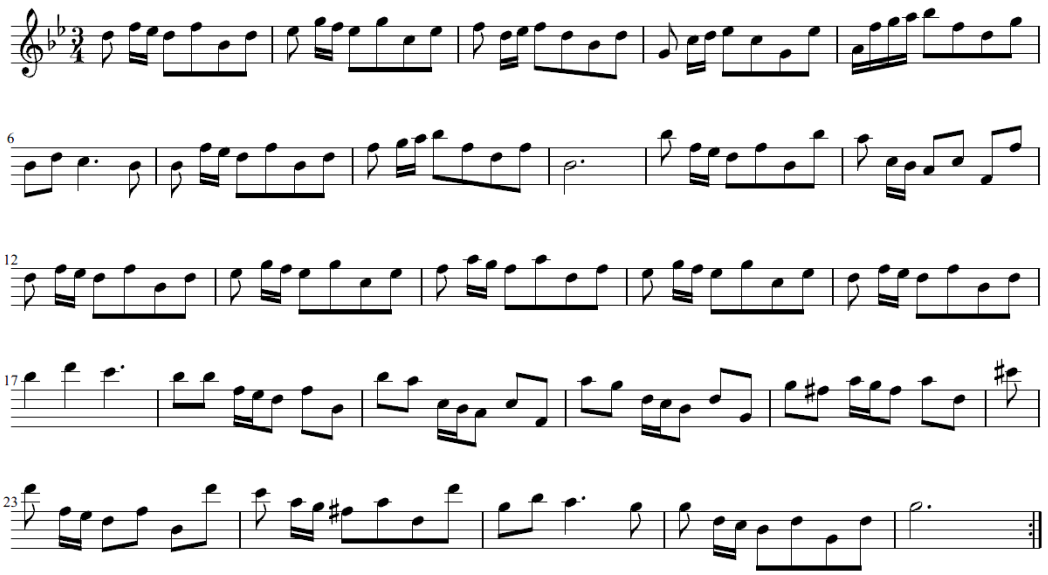

Figure B.2: Score of the second piece. Piece 8 of the book.

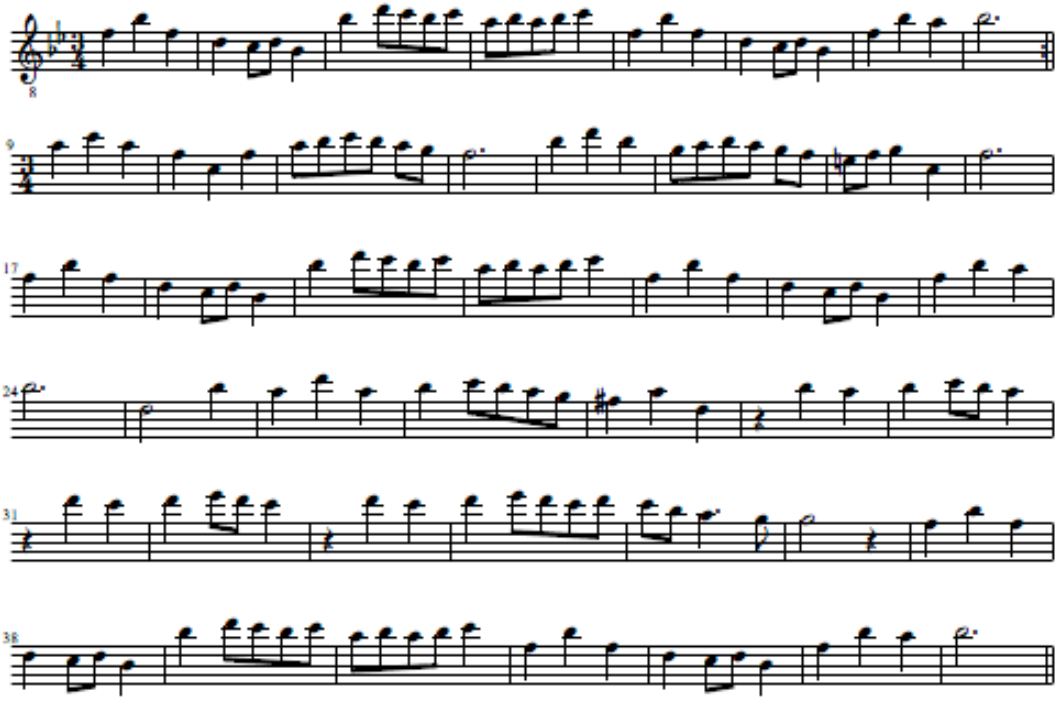

Figure B.3: Score of the third piece. Piece 11 of the book.

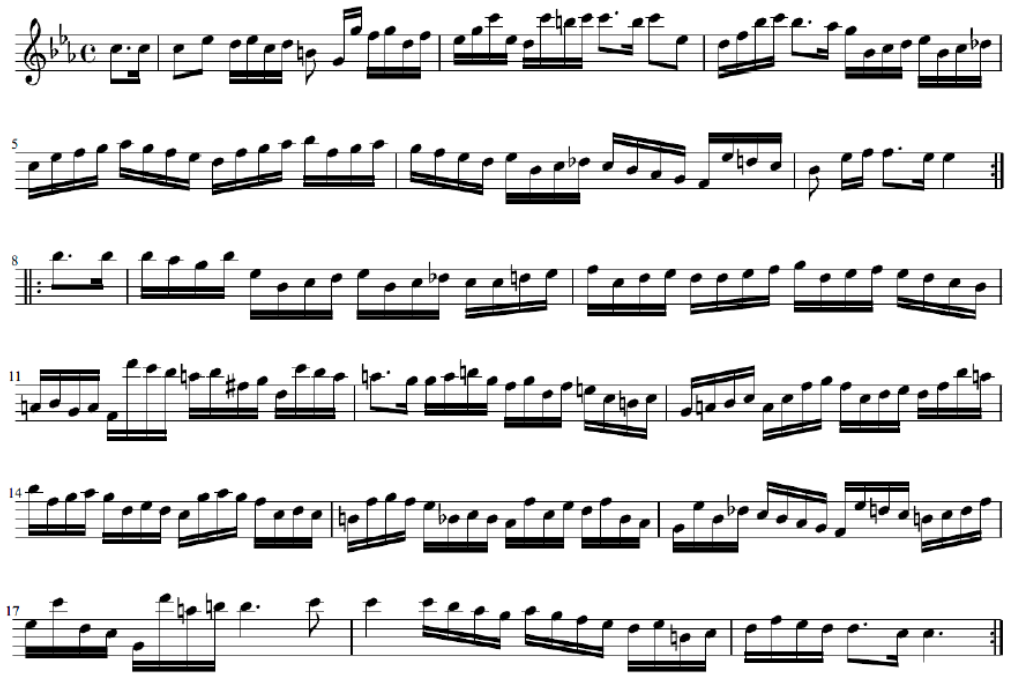

Figure B.4: Score of the fourth piece. Piece 12 of the book.

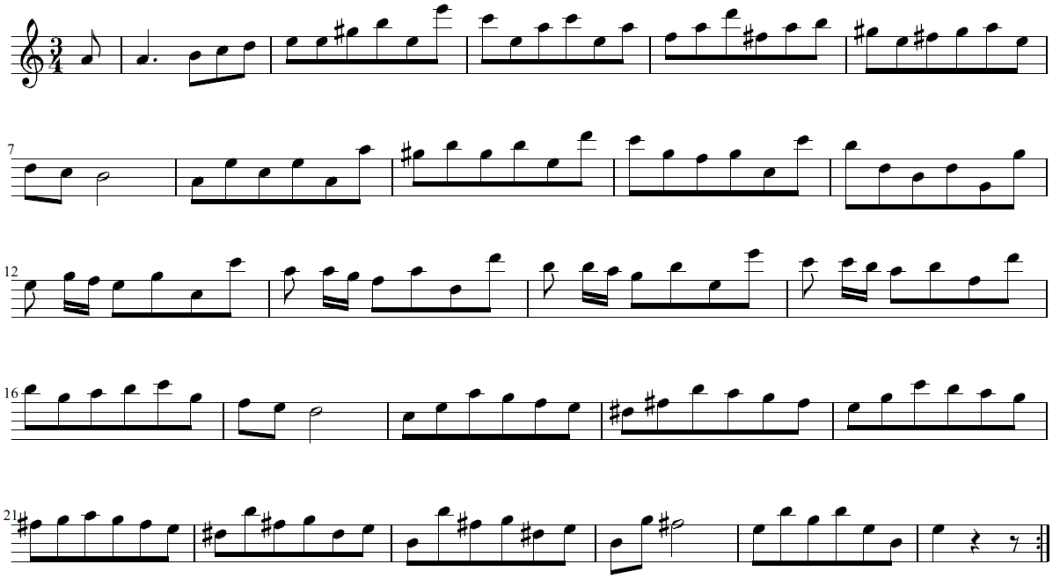

Figure B.5: Score of the fifth piece. Piece 13 of the book.

Bibliography

[1] A. Almeida, R. Chow, J. Smith, and J. Wolfe. The kinetics and acoustics of fingering and note transitions on the flute. *The Journal of the Acoustical Society of America*, 126:1521, 2009.

[2] F. Blanc. Production de son par couplage écoulement/résonateur: Étude des paramètres de facture des flûtes par expérimentations et simulations numériques d'écoulements. 2009.

[3] C. Cadoz. Instrumental gesture and musical composition. In *Proceedings of the International Computer Music Conference*, pages 1–12, 1988.

[4] J.W. Coltman. Sounding mechanism of the flute and organ pipe. *The Journal of the Acoustical Society of America*, 44:983, 1968.

[5] J.W. Coltman. Momentum transfer in jet excitation of flute-like instruments. *The Journal of the Acoustical Society of America*, 69:1164, 1981.

[6] I. Cossette, P. Sliwinski, and P.T. Macklem. Respiratory parameters during professional flute playing. *Respiration physiology*, 121(1):33–44, 2000.

[7] P. de la Cuadra, B. Fabre, N. Montgermont, and L. De Ryck. Analysis of flute control parameters: A comparison between a novice and an experienced flautist. In *Forum Acusticum, Budapest*, 2005.

[8] S. Dequand, JFH Willems, M. Leroux, R. Vullings, M. Van Weert, C. Thieulot, and A. Hirschberg. Simplified models of flue instruments: Influence of mouth geometry on the sound source. *The Journal of the Acoustical Society of America*, 113:1724, 2003.

[9] S.A. Elder. On the mechanism of sound production in organ pipes. *The Journal of the Acoustical Society of America*, 54:1554, 1973.

[10] NH Fletcher. Acoustical correlates of flute performance technique. *J. Acoust. Soc. Am*, 57(1), 1975.

[11] N.H. Fletcher. Acoustic and aerodynamic determinants of the sound quality of flutes. *The Journal of the Acoustical Society of America*, 95:2858, 1994.

[12] N.H. Fletcher and T.D. Rossing. *The physics of musical instruments*. Springer Verlag, 1998.

[13] E. Maestre. *Modeling instrumental gestures: an analysis/synthesis framework for violin bowing*. PhD thesis, Universitat Pompeu Fabra, 2009.

[14] J. Martin. *The acoustics of the recorder*. Moeck, 1994.

[15] N. Montgermont, B. Fabre, and P. de La Cuadra. Flute control parameters: fundamental techniques overview. nicolas montgermont1, benoit fabre1, patricio de la cuadra2. *ISMA*, 2007.

[16] R. Parncutt and G. McPherson. *The science & psychology of music performance: creative strategies for teaching and learning*. Oxford University Press, USA, 2002.

[17] Alfonso Perez. *Enhancing Spectral Synthesis Techniques with Performance Gestures using the Violin as a Case Study*. PhD thesis, Universitat Pompeu Fabra, 2009.

[18] G. Scavone and A. da Silva. Frequency content of breath pressure and implications for use in control. In *Proceedings of the 2005 conference on New interfaces for musical expression*, page 96. National University of Singapore, 2005.

[19] X. Serra. State of the art and future directions in musical sound synthesis. *Chania, Crete*, 2007.

[20] J.O. Smith III. Viewpoints on the history of digital synthesis. 2002.

[21] M.P. Verge, B. Fabre, A. Hirschberg, and APJ Wijnands. Sound production in recorderlike instruments. i. dimensionless amplitude of the internal acoustic field. *The Journal of the Acoustical Society of America*, 101:2914, 1997.

[22] E.M. von Hornbostel and C. Sachs. Systematik der musikinstrumente. *Zeitschrift f*
"ur Ethnologie, 46:553–590, 1914.

[23] M.M. Wanderley and P. Depalle. Gestural control of sound synthesis. *Proceedings of the IEEE*, 92(4):632–644, 2004.

[24] S. Ystad and T. Voinier. A virtually real flute. *Computer Music Journal*, 25(2):13–24, 2001.